



Real-time detection of abnormal driving behavior based on long short-term memory network and regression residuals

Yongfeng Ma^{a,*}, Zhuopeng Xie^a, Shuyan Chen^{a,*}, Fengxiang Qiao^b, Zeyang Li^a

^a Jiangsu Key Laboratory of Urban ITS, School of Transportation, Southeast University, Nanjing 211189, China

^b Innovative Transportation Research Institute, Texas Southern University, Houston, TX 77004, USA



ARTICLE INFO

Keywords:

Real-time abnormal driving behavior detection
Long short-term memory (LSTM) network
Residual algorithm
Smartphone

ABSTRACT

Abnormal driving behavior is one of the main causes of roadway collisions. In most studies of abnormal driving behavior, the abnormal driving status is detected and analyzed using classification algorithms directly or using unsupervised learning algorithms to classify reconstruction or prediction residuals. However, abnormal driving behavior data are difficult to acquire and label. Also, a class imbalance issue is inherent in the algorithm training process due to the relatively sparse data for abnormal driving behavior. Moreover, current studies that include residual analysis tend to focus on individual points and thus fail to capture the continuity characteristic of abnormal driving behavior.

To address these problems, a long short-term memory-residual (LSTM-R) algorithm is proposed to detect abnormal driving behavior in real time. The proposed algorithm (referred to simply as LSTM-R) has two steps. First, an LSTM network is used to fit the current vehicle kinematic data based on historical data to obtain the root mean square residual at each moment. Second, a time window-based residual algorithm is designed and employed to detect abnormal driving behavior according to the magnitude and continuity of the residuals. To verify the effectiveness of LSTM-R, an experimental test was conducted in Nanjing, China. The vehicle kinematic data were collected non-intrusively using a smartphone.

In addition, AdaCost, SMOTEBoost, EasyEnsemble, LightGBM-residual, and linear regression-residual algorithms were employed for comparison with the proposed algorithm to assess its effectiveness. The effects of (1) the degree-of-fit of the LSTM network, (2) the LSTM-R parameters, and (3) the abnormal driving behavior percentage on the detection results were analyzed in detail. First, both the underfitting and overfitting of the LSTM network compromise the detection performance. Second, within a certain range of values, the LSTM-R parameters have little effect on the detection results. Third, the detection results are affected only slightly by the abnormal proportion. The results show that LSTM-R, with a maximum F1-score of 0.866, significantly outperforms the other five algorithms. Furthermore, even if only 10% abnormal driving behavior is in the training set, LSTM-R's F1-score can still be close to 0.86, indicating a significant relaxation of the requirements for labeled data. Also, the required data are easy to collect, which indicates LSTM-R's extensive application possibilities. This paper thus provides an effective method for the real-time detection of abnormal driving behavior and also supports driving risk assessment and driving behavior improvement with the overall goal to enhance roadway safety.

* Corresponding authors.

E-mail addresses: mayf@seu.edu.cn (Y. Ma), 1368334393@qq.com (Z. Xie), chenshuyan@seu.edu.cn (S. Chen), fengxiang.qiao@tsu.edu (F. Qiao), 220183077@seu.edu.cn (Z. Li).

1. Introduction

According to the World Health Organization (WHO), nearly 1.3 million people worldwide die in preventable traffic accidents and 50 million are injured each year, and roadway crashes are the leading killer of children and young people (WHO, 2021). Studies have shown that most collisions are caused by human error and that dangerous driving, referred to also as ‘abnormal driving behavior’, is the dominant reason for such accidents (Bucsuhazy et al., 2020; Skrickij et al., 2020). The real-time detection and analysis of abnormal driving behavior, which are critical components of advanced driver assistance systems (Yuan et al., 2020; Mantouka et al., 2021), are the keys to preventing crashes and improving roadway safety (Ansar et al., 2021; Wang et al., 2021; Zhang et al., 2021). Therefore, effective algorithms for the detection and analysis of abnormal driving behavior need to be developed to help prevent roadway accidents.

Abnormal driving behavior generally refers to the driver’s reckless operation of the vehicle during the driving process and typically occurs within a short period of time. Such driving threatens safety, with observable characteristics such as sharp acceleration, sharp braking, sharp turning, sharp lane changes, and other reckless maneuvers (Tasca et al., 2000; Vindhya et al., 2020; Hong et al., 2014). Numerous studies have been conducted to detect and analyze abnormal driving behavior. Some researchers have identified different types of such behavior by detecting drivers’ states via video monitoring (Huang et al., 2019), drivers’ actions (Martín de Diego et al., 2013; Lu et al., 2022), drivers’ physiological and psychological characteristics (Elassad et al., 2020; Xing et al., 2021), and visual characteristics (Louw and Merat, 2017; Fan et al., 2021). However, such methods are associated with high data collection costs. Also, some of the detection devices used are intrusive, which can interfere with driving. Detection of the driver’s states is not direct and not as efficient as detecting the vehicle’s motions. Moreover, abnormal driving behavior is ultimately reflected in the vehicle’s motion states, not the driver’s states.

Fortunately, with the reduction in cost of in-vehicle equipment and the development of smartphones, increasing numbers of studies have been conducted to detect abnormal driving behavior using vehicle kinematic data. The research methods can be divided into two categories that are based respectively on the threshold rule and machine learning or deep learning algorithms.

For the first category, the research method that is based on the threshold rule to detect abnormal driving behavior identifies and analyzes driving behavior by specifying the thresholds of vehicle kinematic parameters using empirical formulas or analytical derivations. Commonly used parameters include acceleration, yaw rate, jerk, and others (Shi et al., 2015; Brombacher et al., 2017). Although these threshold methods are interpretable and computationally efficient, the rules are too simple to yield satisfactory detection results. Nonetheless, recent studies have used this method to extract abnormal driving behavior data as preliminary information for subsequent research. For example, Ma et al. (2021a) extracted three basic driving behaviors, namely, acceleration, deceleration, and turning, by specifying the thresholds of acceleration and the angular velocity of heading to explore differences in driving styles at different driving stages of online car-hailing. Ma et al. (2020) also analyzed the vehicle’s force in the horizontal and vertical planes of the roadway and developed an accurate physical model to detect aggressive driving behavior in real time based on the limit equilibrium of driving safety and driver/passenger comfort.

The other research method category for abnormal driving behavior detection is based on machine learning or deep learning algorithms to build more complex rules. The most common practice is to label the different types of abnormal driving data manually first and then identify them using classification algorithms. A certain driving behavior usually lasts for a distinct period of time, so the features must be extracted manually from driving behavior segments when using traditional machine learning algorithms. Yu et al. (2017) collected six months of vehicle kinematic data via smartphones and identified six types of abnormal driving behavior using support vector machine (SVM) and artificial neural network (ANN) algorithms, with an average accuracy of 95.36 % and 96.88 %, respectively. The six types of abnormal behavior in the Yu et al. (2017) study are weaving, swerving, sideslipping, fast U-turns, turning with a wide radius, and sudden braking.

However, driving behavior is a dynamic and multi-interactive process, so manually extracted features cannot always capture contextual information. With the development of deep learning techniques, more researchers have started to use different neural networks, especially recurrent neural networks (RNNs), to extract features from temporal segments automatically to improve model generalization. Khodairy and Abosamra (2021) used an optimized stacked long short-term memory (LSTM) model, which is based on signals from smartphone embedded sensors, to generate two classification models: (1) binary, which is used to differentiate between aggressive and non-aggressive driving, and (2) three-class, which is used to distinguish normal, fatigued, and aggressive driving. Jia et al. (2020) constructed a convolutional neural network and LSTM (CNN-LSTM) algorithm to identify-three abnormal driving behaviors, i.e., sharp acceleration, sharp braking, and sharp lane change, based on vehicle kinematic data, with an average accuracy of 95.684 percent.

Although machine learning and deep learning algorithms have a strong learning capability and usually achieve desirable detection results, using classification algorithms directly requires a large amount of manual data labeling, which is highly subjective and time-consuming, while other labeling methods have no uniform standard yet. Also, abnormal driving behavior occurs relatively rarely compared to normal driving behavior, which can lead to a class imbalance problem and weak detection effects (Zhang et al., 2019; Wang et al., 2020).

To address these issues, some researchers have used pretraining followed by fine-tuning in their studies. For example, Hu et al. (2020) constructed a stacked sparse autoencoders (SdsAEs) model, pretrained it using a large amount of unlabeled data, and then fine-tuned it using labeled data to detect three types of abnormal driving behavior: drunk/fatigue, reckless, and phone use. However, the SdsAEs model is essentially a three-layered ANN, so it cannot capture the contextual correlations of the driving data. Moreover, the labeled data in the Hu et al. (2020) study were generated by simulation, which could weaken the authenticity of the study. Further, the

various abnormal driving behaviors are relatively rare in practice compared to normal driving, which may undermine the effects of the fine-tuning process.

Therefore, some researchers have proposed another way to approach the detection of abnormal driving behavior. Because abnormal driving usually is sudden and drastic compared to normal driving, when abnormal driving behavior occurs, the time series-based vehicle kinematic data likewise change dramatically (Jia et al., 2020; Ma et al., 2019). During the driving process, the vehicle is in a normal state most of the time. Therefore, if a regressor is employed to reconstruct or predict the data, it will tend to fit the normal data well to achieve an overall good performance, but will yield large regression residuals when abnormal driving behavior occurs. Based on these considerations, a few studies have used deep learning algorithms to reconstruct or predict vehicle kinematic data to obtain regression residuals and then detect abnormal driving behavior according to the residuals using simple threshold rules (Kieu et al., 2019) or unsupervised learning algorithms, such as the one-class SVM (Fan et al., 2022) and Gaussian mixed model (Ding et al., 2019), etc. For the reconstruction methods, the data are reconstructed mainly by CNN-based or RNN-based autoencoders, and for the prediction methods, the current data are predicted by the historical data using a CNN or RNN directly. For example, Fan et al. (2022) developed a recurrent-convolutional autoencoder to extract the spatial-temporal characteristics of vehicle kinematic data and then used the one-class SVM algorithm and the Pauta criterion to detect abnormal lane changing behaviors based on the extracted characteristics. However, such methods do not require labeled data and focus instead on individual points, thus failing to consider the continuity of abnormal driving behavior. So, these types of methods cannot capture temporal information, which may lead to an increase in the misclassification rate. Moreover, in the case of a large volume of data, these unsupervised learning algorithms are computationally intensive and costly (Alham et al., 2011; Fu et al., 2021; Fan et al., 2022).

To address these problems, this paper proposes a real-time abnormal driving behavior detection method called the LSTM-residual algorithm. The LSTM-R algorithm (referred to simply as LSTM-R in this paper) first uses an LSTM network to fit the vehicle kinematic data at the current moment based on the historical data to obtain regression residuals. Then, LSTM-R employs a residual algorithm that focuses on a specific past time window instead of on individual points to detect abnormal driving behavior according to the magnitude

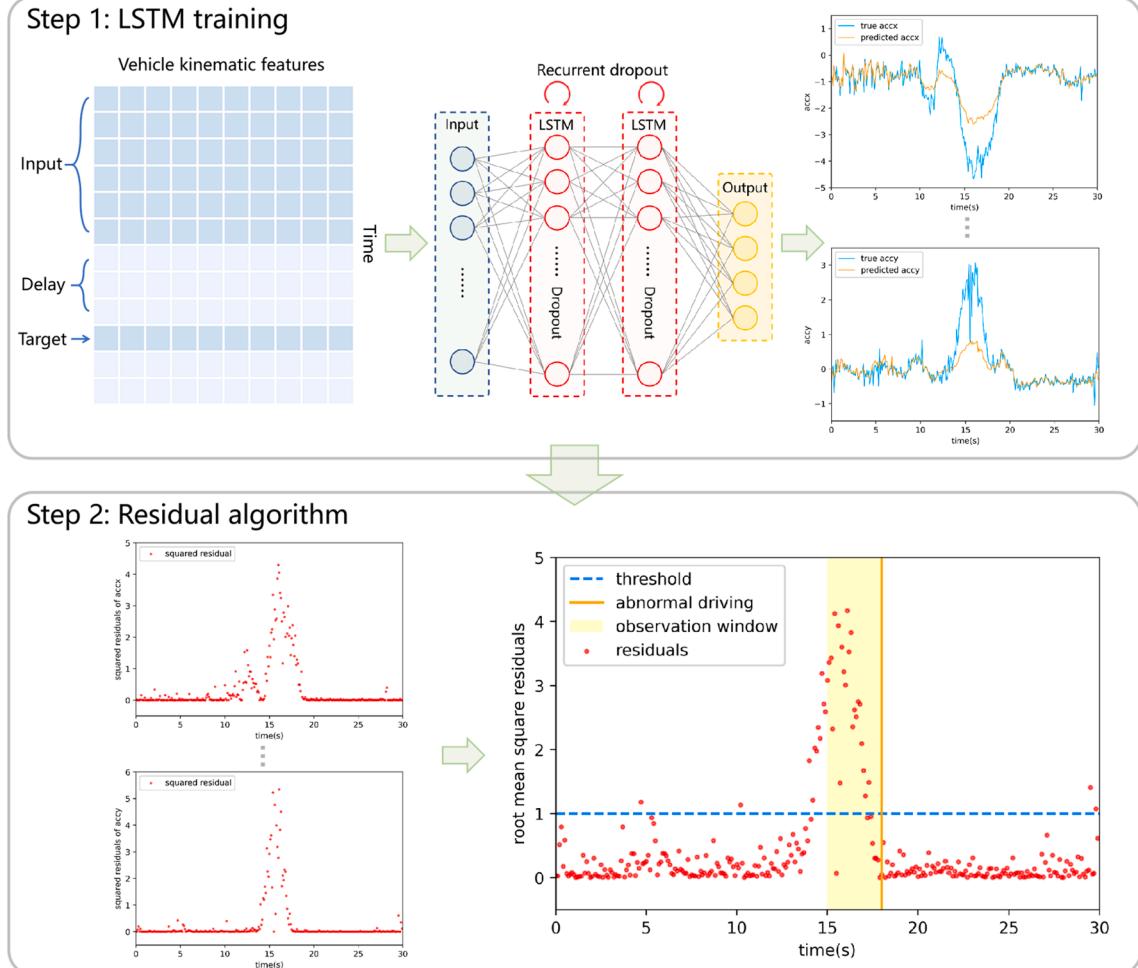


Fig. 1. Framework of two-step LSTM-R algorithm.

and continuity of the residuals. To validate the effectiveness of the proposed LSTM-R, an experimental test was conducted in Nanjing, China. The driver was asked to perform four types of typical abnormal driving maneuvers: sharp acceleration, sharp braking, sharp turning, and sharp lane change. A smartphone was used to collect the vehicle's kinematic data in a non-intrusive manner. Then, LSTM-R was employed to identify the normal and abnormal driving behaviors. Next, for comparative purposes, AdaCost, SMOTEBoost, EasyEnsemble, LightGBM-residual (LightGBM-R), and linear regression-residual (LR-R) algorithms were employed to assess the effectiveness of LSTM-R. In addition, the effects of the degree-of-fit of the LSTM network, the LSTM-R parameters, and the abnormal driving behavior proportion on the detection results were investigated in this study.

The three salient contributions of this paper can be summarized as follows. First, this paper proposes a novel algorithm, LSTM-R, which is used to detect abnormal driving behavior via a regression model. The underlying principle of LSTM-R is to use the LSTM network's tendency to fit most normal samples well, which serves to isolate the relatively rare occurrences of abnormal driving behavior. Compared to using classification algorithms directly, LSTM-R performs the task of abnormal driving behavior detection using the threshold-based method based on the residuals obtained by regression. In this case, data from each moment are used to train the regression model, thus ensuring sufficient training samples. Also, the threshold-based method has a simple structure that avoids the risk of overfitting and requires only a small number of abnormal driving behavior samples. Therefore, the required sample size of LSTM-R is much smaller than that of models that use classification algorithms directly. Second, different from residual processing methods that are focused on individual points, the time window-based residual algorithm proposed in this paper is compatible with the continuity characteristic of abnormal driving behavior. Third, to validate the effectiveness of LSTM-R, an experimental test was conducted to analyze the detection principle, parameter selection, and detection results under different conditions in detail. The data are easy to collect, so the proposed algorithm has extensive application possibilities.

The rest of this paper is organized as follows. [Section 2](#) introduces the underlying principle and processes of LSTM-R and discusses the development methodology. [Section 3](#) describes the experimental test and model training conducted in this study. [Section 4](#) provides discussion and analysis of the results. [Section 5](#) concludes the paper by citing limitations and projecting future research.

2. Methodology

[Fig. 1](#) presents the two steps involved in the LSTM-R algorithm. For Step 1, LSTM training, an LSTM network is used to fit the vehicle kinematic data at the current moment based on the historical data in order to obtain the regression residuals. For Step 2, the root mean square residual (RMSR) at each moment is calculated, and a designed time window-based residual algorithm detects the abnormal driving behavior according to the magnitude and continuity of the residuals.

2.1. Long short-term memory network

An LSTM network ([Hochreiter et al., 1997](#)) is a variant of an RNN and uses an innovative approach to compute the states of hidden layers, including multiple gate units. An LSTM network introduces parameters for controlling the memory states, and each unit can decide whether to pass the upstream states to the downstream states. Thus, the model can recognize patterns at different scales in the time series and achieve a long-distance propagation of information. The gate structure of an LSTM network also has a certain level of anti-noise capability ([Zhuang et al., 2019; Han et al., 2021](#)), and this inherent property of the LSTM network ensures the robustness of the proposed algorithm, LSTM-R. In addition, numerous studies have found that LSTM networks have strong generalization capability and are suitable for multivariate time series-based regression and classification problems with a large amount of data ([Rastgoo et al., 2019; Hu et al., 2020; Hu et al., 2021](#)). The calculation process for each unit of an LSTM network is shown in Equations (1) through (6).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (6)$$

where f_t , i_t , and o_t are the outputs of the forgetting gate, input gate, and output gate at the current moment, respectively; C_{t-1} and C_t are the unit states of the previous moment and the current moment, respectively; \tilde{C}_t is the candidate value of C_t ; h_{t-1} is the output of the hidden layer at the previous moment; x_t is the input at the current moment; W_f , W_i , W_C , and W_o are the weight matrices; b_f , b_i , b_C , and b_o are the bias vectors; and σ is a sigmoidal function.

2.2. Underlying principle of LSTM-R

As mentioned in the Introduction ([Section 1](#)), a regressor that is used to fit vehicle kinematic data tends to fit normal driving data

well but will yield large regression residuals at abnormal points. The regressor must have a strong learning capability and specialize in time-series data to capture temporal characteristics effectively, and an LSTM network was selected in this study for this purpose. After regression modeling, a sequence of residuals was generated. Because vehicle motion states need to be described by multiple features, that is, multiple residuals will be evident at each moment, an appropriate metric must be selected to integrate the residual vector into a scalar. Because both the inputs and outputs are standardized, the RMSR can be used for this purpose, as shown in Equation (7).

$$RMSR_t = \sqrt{\frac{1}{m}[(y_t - \hat{y}_t)(y_t - \hat{y}_t)]^T} \quad (7)$$

where $RMSR_t$ represents the root mean square residual at timet; m represents the number of features; y_t represents the truth vector at timet; and \hat{y}_t represents the fitted value vector at timet.

Given that abnormal driving behavior generally lasts for a specific period of time, and the residual of a single noise point can be too high to affect the detection results, a simple time window-based residual algorithm was designed in this study after obtaining the residual sequence. Specifically, if a certain percentage of residuals in the neighborhood of a point exceeds a specified threshold, then that point will be labeled as an abnormal driving behavior.

2.3. Parameters and procedures of the LSTM-R algorithm

Table 1 and Algorithm 1 respectively show the parameters and procedure of the proposed LSTM-R algorithm.

Algorithm 1 LSTM-R

Input:

Vehicle kinematic data set: $S = \{s_1, s_2, \dots, s_n\}$, where n represents the number of samples; s is a row vector with length m ; and m is the number of features.

Parameters: $l, d, thr, neigh, rate$.

Procedure:

for $t = l+d+1, l+d+2, \dots, n$ do

construct the fitting targets: $y_t = s_t$;

construct the input samples: $X_t = \begin{bmatrix} s_{t-l-d} \\ s_{t-l-d+1} \\ \vdots \\ s_{t-d-1} \end{bmatrix}$;

end for

Construct the training set: $D = \{(X_1, y_1), (X_2, y_2), \dots, (X_p, y_p)\}, p = n - l - d$;

Fit D by the LSTM, obtain the fitted values: $\hat{Y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_p\}$;

for $t = 1, 2, \dots, p$ do

calculate the root mean square residuals: $RMSR_t = \sqrt{\frac{1}{m}[(y_t - \hat{y}_t)(y_t - \hat{y}_t)]^T}$

end for

for $t = neigh, neigh+1, \dots, p$ do

$Rtemp = [RMSR_{t-neigh+1}, RMSR_{t-neigh+2}, \dots, RMSR_t]$;

denote the number of residuals higher than or equal to thr in $Rtemp$ as num ;

if $\frac{num}{neigh} >= rate$ then

label time t as 1;

else

label time t as 0;

end if

end for

Output:

The label of each moment, where 0 and 1 represent normal and abnormal driving behaviors, respectively.

2.4. Comparison of algorithms

LSTM-R aims to address the class imbalance issue that is evident in abnormal driving behavior detection. Therefore, for this study, algorithms that have been designed specifically for imbalanced data were selected for comparison with LSTM-R to assess LSTM-R's effectiveness. These commonly used algorithms can be divided into three main categories, i.e., cost-sensitive learning, oversampling-

Table 1

Parameters of LSTM-R.

Symbol	Description of Parameter	Unit	Value range
l	length of time of input samples	s	1–10
d	time delay of fitting targets	s	1–5
thr	threshold of residuals	–	0.5–1.9
$neigh$	length of the neighborhood time window used to observe residuals	s	1–4
$rate$	ratio threshold of points with large residuals	–	0.1–1.0

based, and undersampling-based. Three typical algorithms were selected respectively from each category: AdaCost (Fan et al., 1999), SMOTEBoost (Chawla et al., 2003), and EasyEnsemble (Liu et al., 2008). In addition, in the first step of the LSTM-R algorithm, two other regressors were considered in place of the LSTM for comparative purposes: LightGBM and linear regression, which are used respectively in the LightGBM-R and LR-R algorithms. Details concerning the underlying principles of these algorithms can be found in the literature and thus are not repeated in this paper.

3. Case study

3.1. Data collection

In order to verify the effectiveness of the proposed LSTM-R algorithm, an experimental test was conducted on July 16, 2019 in Nanjing, China, with good weather conditions that day. A male driver with more than 15 years of driving experience was asked to execute a series of four typical abnormal driving behaviors, i.e., sharp acceleration, sharp braking, sharp turning, and sharp lane changes, evenly distributed throughout the test route. The test route included expressways, arterial roads, collector roads, and local streets, each with free-flowing traffic conditions. Fig. 2 shows the test route and distribution of the different types of abnormal driving behavior.

Before the test, an iPhone 7 was calibrated and affixed horizontally to the dashboard with tape to avoid being dislodged during bumps and sudden movements, as shown in Fig. 3. During the test, Sensorlog, an iOS application, was used to collect the vehicle kinematic data in real time with a sampling frequency of 10 Hz. Two technicians also were in the vehicle with the driver. One technician was responsible for navigating and giving instructions, and the other recorded the start and end times of each abnormal driving behavior, with accuracy to seconds. Note that, for the detection of real-time driving behavior, accuracy measured in seconds is considered to be sufficient.

The test lasted 1 h and 18 min during which time a total of 47,100 pieces of data were collected, including 8 sharp accelerations, 15 sharp brakings, 13 sharp turnings, and 15 sharp lane changes. Because sharp turning and sharp lane changes usually are accompanied by sharp acceleration and sharp braking, different abnormal driving behaviors can overlap or alternate. For this study, these four behaviors were counted together as abnormal behavior uniformly without differentiation. As a result, 3,043 abnormal driving labels were obtained, and the remaining 44,057 pieces of data were labeled as normal driving.

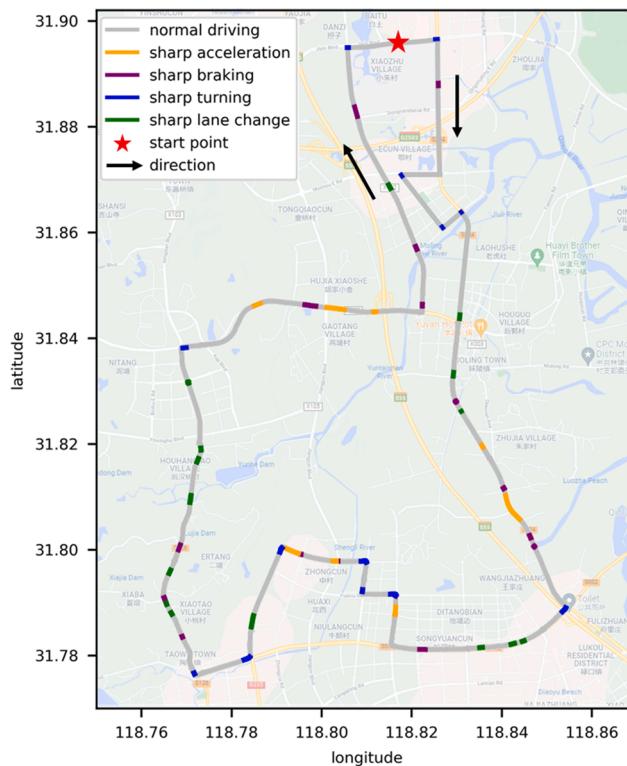


Fig. 2. Test route.

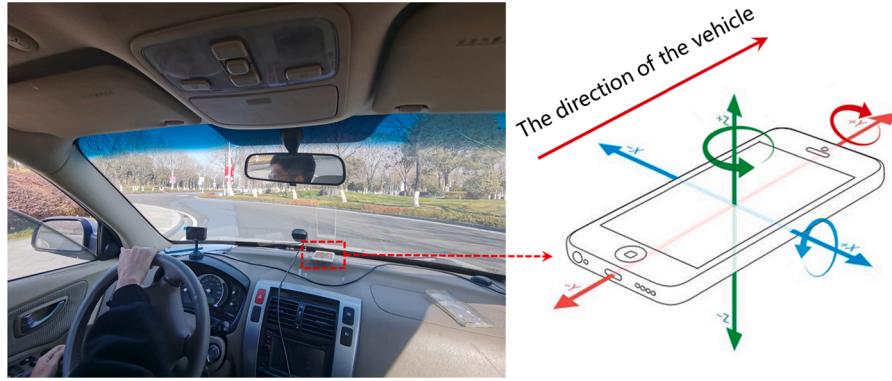


Fig. 3. Test equipment.

3.2. Data preprocessing

Ten typical vehicle kinematic features were selected for this study and include velocity, tri-axis acceleration, tri-axis angle, and tri-axis angular velocity (Xie et al., 2018; Ma et al., 2021b; Xie et al., 2021), which can determine the vehicle's motion states. [Table 2](#) presents descriptions of these features.

The next procedure is data preprocessing. First, missing values were processed. These data were collected at a high frequency with few missing values. Thus, linear interpolation was used. Then, the yaw, roll, and pitch angles were converted to cumulative angles to accommodate the sudden changes of each. Next, considering that different features have different scales, z-score normalization, as expressed in Equation (8), was used to transform the data into a uniform scale in order to meet the requirements for the calculation of the residuals. Finally, the times of the manually labeled abnormal driving behaviors and the times of the vehicle kinematic data were correlated.

$$x^* = \frac{x - \bar{x}}{\sigma} \quad (8)$$

where x and x^* are the raw and normalized values, respectively; \bar{x} is the mean value; and σ is the standard deviation.

Significantly, LSTM-R assumes that the values of the residuals will be higher when abnormal driving behavior occurs. Data smoothing will cause the loss of abnormal driving information and thus is incompatible with the real-time requirements of the proposed algorithm. A similar phenomenon of data smoothing that causes data loss was encountered in a previous study by the authors (Ma et al., 2019). In addition, as mentioned in the Methodology section ([Section 2](#)), the LSTM network itself has a certain level of anti-noise capability. Therefore, data smoothing was not necessary for this study.

3.3. Model training

The data set was divided into a training set, validation set, and test set with the ratio of 6:2:2, respectively, and in chronological order. The structure and parameters of the LSTM network were determined using the validation set. Through repeated testing and comparison, the constructed LSTM network contained two LSTM layers and one output layer. Each LSTM layer had 64 units with the activation function of a rectified linear unit (ReLU), whereas the output layer used linear activation. During the training process, both dropout and recurrent dropout regularization processes were used to prevent overfitting, with the dropout ratio set to 0.2. The batch size was 128, and Adam was chosen as the optimizer, with a learning rate of 0.001. The upper limit of the number of epochs was set to 500 with early stopping, whereby the results were saved every 10 epochs. With regard to the five algorithms used for comparison with

Table 2

Descriptions of Vehicle Kinematic Data.

Field Name	Description	Units	Min	Max	Mean	Standard deviation
Velocity	velocity	m/s	0.000	32.240	10.019	8.231
Accx	x-axis acceleration	g	-0.502	0.732	0.080	0.066
Accy	y-axis acceleration	g	-0.576	1.040	0.006	0.093
Accz	z-axis acceleration	g	-0.969	0.773	0.009	0.053
Yaw	yaw angle	rad	-3.141	3.141	-0.225	1.945
Roll	roll angle	rad	-0.020	0.140	0.074	0.016
Pitch	pitch angle	rad	-0.099	0.085	-0.006	0.024
Yawrate	yaw rate	rad/s	-0.260	0.302	0.000	0.024
Rollrate	roll rate	rad/s	-0.297	0.226	0.000	0.018
Pitchrate	pitch rate	rad/s	-0.568	0.512	-0.001	0.057

the LSTM-R, AdaCost, SMOTEBoost, and EasyEnsemble take each data frame (0.1 s) as input, whereas the time series-based input data of LightGBM and linear regression, as regressors, are flattened. The hyper-parameters of these algorithms, except for linear regression, were selected using fivefold cross-validation. **Table 3** shows the parameters for each of the five algorithms compared in this study.

3.4. Model evaluation

In this study, the mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2) were selected to evaluate the fitting effects of the LSTM network, as expressed in Equations, (10), and, respectively.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (9)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (10)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (11)$$

where n is the number of samples; y_i and \hat{y}_i are the true and fitted values of the i th sample, respectively; and \bar{y} is the mean value of all samples.

The detection of abnormal driving behavior is essentially a classification problem, and the commonly used evaluation metrics for classification are precision, recall, and F1-score. Equations (12), (13), and (14) are the calculation equations for these three metrics, respectively.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (12)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (13)$$

$$\text{F1 - score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (14)$$

where TP, FP, FN, and TN are true positive, false positive, false negative, and true negative, respectively, and indicate the number of each metric.

To compare the performance of the six algorithms more intuitively, a receiver operating characteristic (ROC) curve was used. The horizontal coordinate of the ROC curve is the false positive rate (FPR), defined in Equation (15), whereas the vertical coordinate is the true positive rate, namely, the recall, defined in Equation (13). The measurement variable, ‘area under the curve’ (AUC), is the total area under the ROC curve, with a larger area indicating better model performance.

$$\text{FPR} = \frac{FP}{FP + TN} \quad (15)$$

Table 3
Parameter Selection for Each Compared Algorithm.

Algorithms	Parameters	Parameter Range	Optimal Parameters
LSTM	number of units N_{unit}	$N_{unit} \in \{16, 32, 64, 128\}$	$N_{unit} = 64, m = 128$
	batch size m	$m \in \{32, 64, 128, 256\}$	$D_r = 0.2, R_r = 0.2,$
	dropout rate D_r	$D_r \in \{0.1, 0.2, 0.5\}$	$l_r = 10^{-3}$
	recurrent dropout rate R_r	$R_r \in \{0.1, 0.2, 0.5\}$	
	learning rate l_r	$l_r \in \{10^{-3}, 10^{-2}\}$	
AdaCost	number of decision trees N_d	$N_d \in \{100, 200, 300, 500\}$	$N_d = 300, n_d = 10,$
	maximum depth of each tree n_d	$n_d \in \{5, 10, 20, 50\}$	$r = 10$
	ratio of the false negative cost and false positive cost r	$r \in \{2, 5, 10, 15\}$	
SMOTEBoost	number of decision trees N_d	$N_d \in \{100, 200, 300, 500\}$	$N_d = 200, n_d = 20,$
	maximum depth of each tree n_d	$n_d \in \{5, 10, 20, 50\}$	$k = 5$
	number of nearest neighbors in SMOTE k	$k \in \{1, 2, 5, 10\}$	
EasyEnsemble	number of AdaBoost learners N_A	$N_A \in \{100, 200, 300, 500\}$	$N_A = 300,$
	number of base learners in each AdaBoost learner n_b	$n_b \in \{10, 20, 50, 100\}$	$n_b = 10$
LightGBM	learning rate l_r	$l_r \in \{0.01, 0.05, 0.1, 0.2\}$	$l_r = 0.05, N_e = 1000,$
	number of estimators N_e	$N_e \in \{300, 500, 1000, 2000\}$	$d = 20, ml = 100$
	maximum depth of each tree d	$d \in \{10, 20, 30, \infty\}$	
	maximum tree leaves of each tree ml	$ml \in \{10, 31, 100, 200\}$	

4. Results and discussion

4.1. Comparison between multi-output and single-output regression

For the regression modeling undertaken in this study, two approaches were compared: multi-output regression, where the fitting targets are all the features (Nguyen et al., 2021), and single-output regression, where only one feature at a time is modeled. Table 4 presents a comparison of the two sets of regression results and Fig. 4 and Fig. 5 respectively present the results for multi-output regression and single-output regression using graphs for each feature. In the case of multi-output regression, the velocity, yaw, roll, pitch, and pitch rate are well fitted, whereas the fit of the other features is poor, which may be due to the fact that most of the features (i.e., velocity, yaw, roll, pitch, and pitch rate) are continuously changing variables with little fluctuation, whereas the other variables are more volatile and thus not as easy to fit. Compared with multi-output regression, single-output regression yielded better results for all the features because, when the output is a single feature, the fitting target is specific and the features do not interfere with each other. Single-output regression requires more computations, but the training is performed offline, so it is recommended for practical applications.

4.2. Residuals of abnormal driving behavior

After fitting the data using the LSTM network, the RMSR series was used to detect abnormal driving behavior. Two fitting segments of two typical features, lateral and longitudinal acceleration, were selected to analyze the algorithmic principle, as shown in Fig. 6. The light green and pink areas represent the manually recorded normal and abnormal driving behaviors, respectively. The LSTM network fitted the data well for normal driving behavior, but yielded much higher residual values for abnormal driving behavior. Consequently, the start and end times of the abnormal driving behavior are clearly distinguishable in the figure, indicating that LSTM-R can successfully exploit the LSTM network's tendency to fit normal samples well and thus fulfills its aim to isolate the residuals of abnormal driving behavior.

4.3. Comparison of different parameters

Fig. 7 shows the maximum F1-scores for abnormal driving behavior detection for different l (length of time of input samples) and d (time delay of fitting targets) values. With an increase in the l or d parameter, the F1-score basically shows a trend of first increasing and then decreasing. Too short an observation will lead to the capture of insufficient information, whereas too long will introduce unnecessary information, which can undermine the fitting effects. The delay of fitting controls the temporal correlation between the fitting targets and the training samples to some extent. As shown in Fig. 7, the best scheme is to use 6 s of historical data to fit the data after 3 s. The detection results are not particularly sensitive to l and d within a certain range, but a longer observation will increase the calculation burden. Therefore, for practical applications, the recommended ranges of values for l and d are 3 s to 6 s and 2 s to 4 s, respectively.

Fig. 8 shows the F1-scores for different times of $neigh$, thr , and $rate$ when $l = 6$ s and $d = 3$ s. The best detection results are achieved by a scheme whereby, if 40 % of the RMSRs in the past 1 s are greater than or equal to 1.2, then the current moment should be labeled as an abnormal driving instance. As $neigh$ increases, the requirement for the continuity of abnormal points becomes increasingly greater, leading to a decrease in the accuracy of the detection results. Common sense suggests that the detection results near the main diagonal of the heat maps are more accurate than the outer ones. Specifically, as thr increases, $rate$ needs to decrease continuously to maintain an equivalent result. As shown, the F1-score can reach about 0.8 for a wide range of values for the different parameters.

Table 4
Comparison between Multi-Output and Single-Output Regression Outcomes.

Field Name	multi-output regression			single-output regression		
	MAE	RMSE	R^2	MAE	RMSE	R^2
Velocity	0.728	1.138	0.998	0.356	0.586	0.999
Accx	0.179	0.323	0.754	0.112	0.231	0.874
Accy	0.192	0.337	0.862	0.122	0.246	0.926
Accz	0.292	0.504	0.070	0.092	0.207	0.843
Yaw	1.657	3.277	0.998	0.651	2.654	0.999
Roll	0.055	0.089	0.991	0.028	0.046	0.998
Pitch	0.078	0.136	0.991	0.036	0.065	0.998
Yawrate	0.746	1.270	0.142	0.237	0.442	0.896
Rollrate	0.480	0.884	0.243	0.195	0.433	0.819
Pitchrate	0.221	0.399	0.985	0.128	0.205	0.996

Note: MAE is mean absolute error; RMSE is root mean square error, and R^2 is the coefficient of determination.

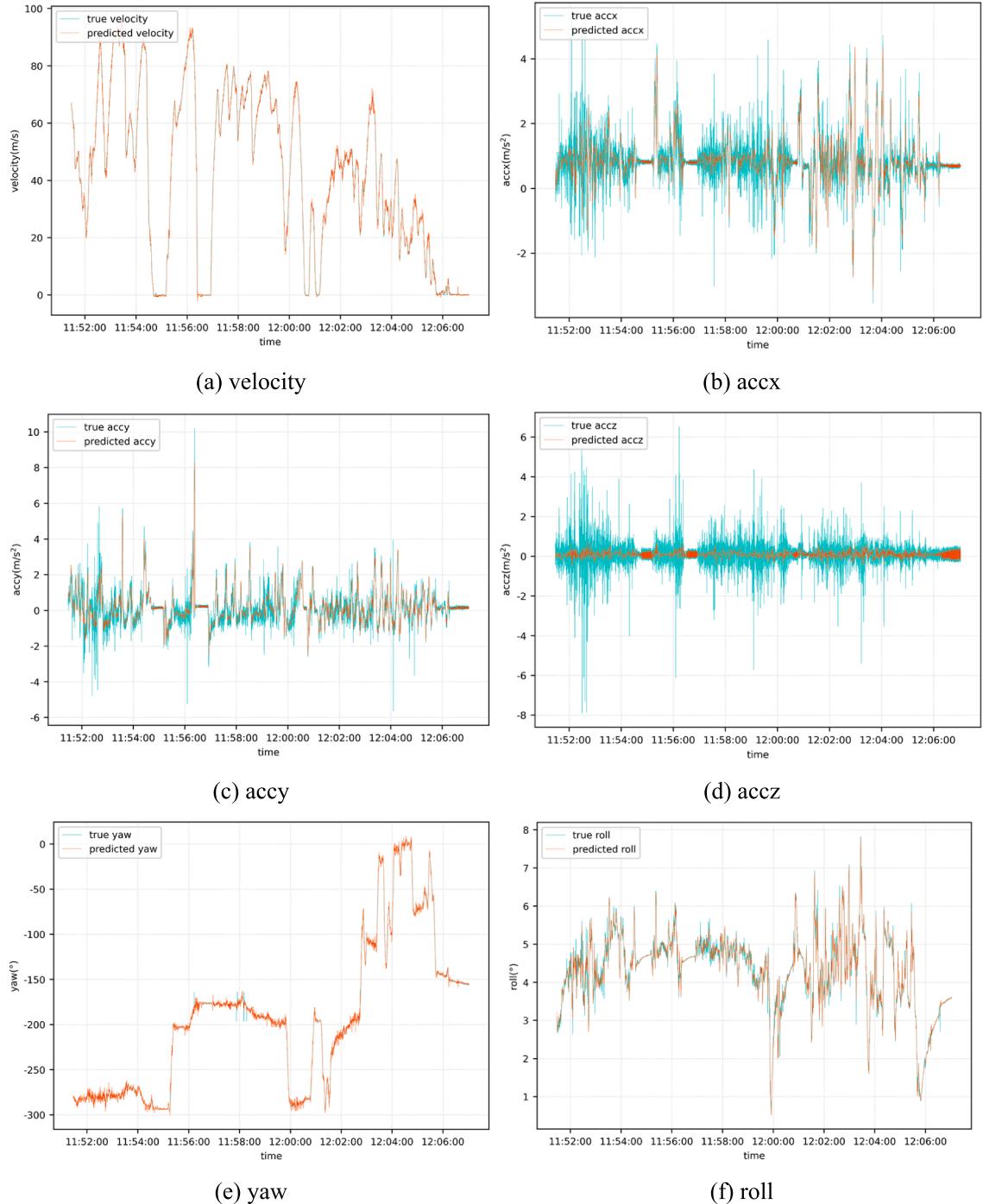


Fig. 4. Multi-output regression results of LSTM network.

4.4. Effects of underfitting and overfitting of the LSTM network on the detection results

The underlying principle of LSTM-R is to use the LSTM network to fit normal samples well but to yield large residuals in abnormal samples. Both the overfitting and underfitting of the LSTM network will increase the residuals of the test set, which is not conducive to anomaly detection. To prove this point, the effects of the degree-of-fit of the LSTM network on the detection results were investigated through a series of comparison tests. Taking the LSTM network's unit number N_{unit} equals 64 as an example, the degree-of-fit is

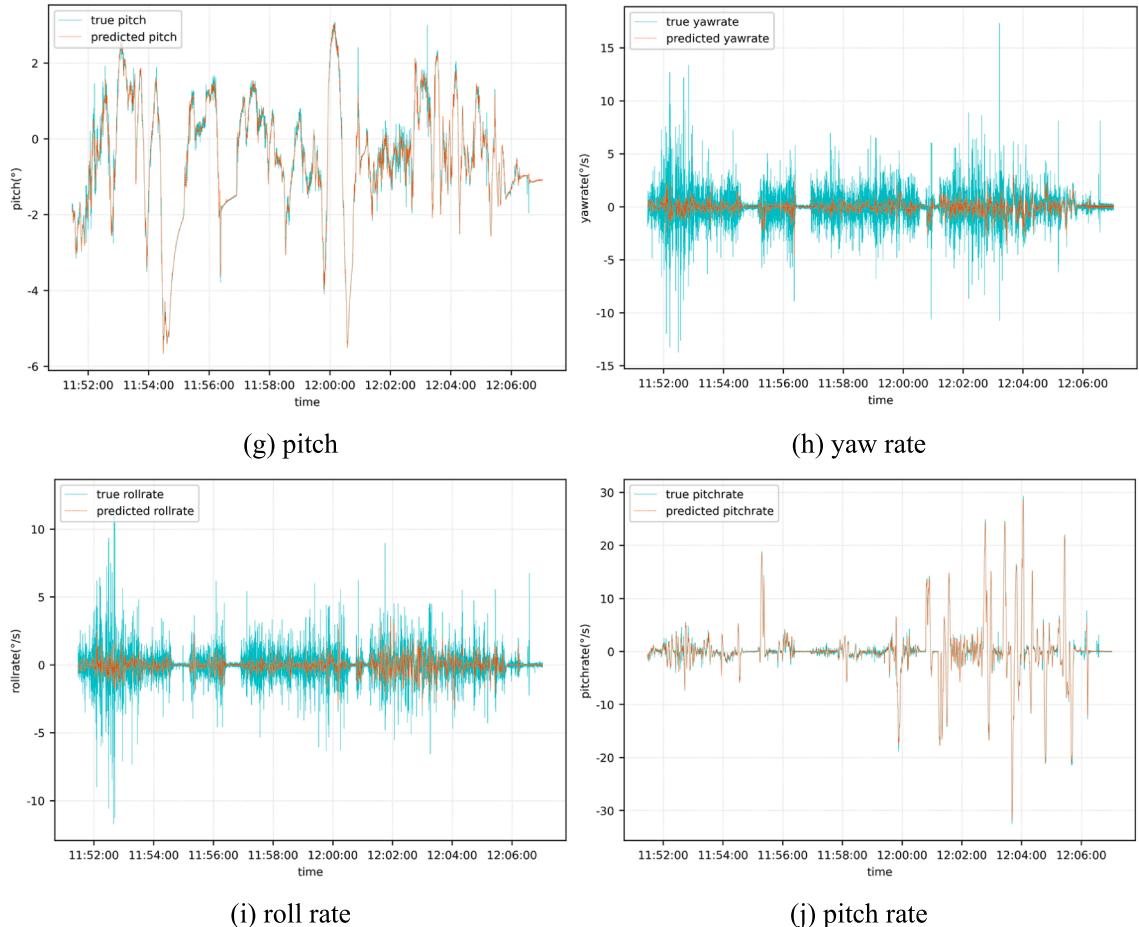


Fig. 4. (continued).

controlled by the number of epochs. Fig. 9 shows the F1-score and average RMSR in different epochs. As the number of epochs increases, the LSTM network changes from underfitting to overfitting, where the average RMSR of the test set first decreases and then increases. The F1-score shows the opposite trend and reaches the maximum at the epoch number of 200. Fig. 10 shows the regression residuals of the abnormal driving behavior segments in the training set and test set in the case of underfitting, well-fitting, and overfitting, respectively. As expected, both underfitting and overfitting are shown to lead to higher residual values and poorer detection performance compared to the well-fitting case. Therefore, to optimize the detection results, the LSTM network parameters must be selected using the validation set in the regression stage to ensure the well-fitting case for the LSTM network.

4.5. Effects of abnormal proportion on detection results

In practice, abnormal driving behavior is relatively rare. To explore the effects of its frequency on the regression and detection results, a stratified sampling method was used to retain different proportions of the four types of abnormal driving behavior investigated in this study. The proportions were set as 0 to 1 in intervals of 0.1. For example, when the proportion is 0.5, half of each type of abnormal driving behavior in the training set is retained randomly, whereas the other abnormal driving behaviors are deleted. Fig. 11 shows the LSTM network's average RMSR values and F1-scores for different abnormal proportions. As the abnormal proportion decreases, the LSTM network's ability to predict the vehicle kinematic data with severe fluctuations diminishes accordingly, and thus, the regression residuals increase. However, the detection results are affected only slightly by this phenomenon. Even if only 10 % abnormal driving behavior is in the training set, the F1-score can still be close to 0.86, indicating that LSTM-R is applicable for scenarios with the rare occurrence of abnormal driving behavior.

4.6. Comparison of different algorithms

Fig. 12 and Table 5 present comparative results for LSTM-R, AdaCost, SMOTEBoost, EasyEnsemble, LightGBM-R, and LR-R. As shown, the ROC curve of LSTM-R is closer to the upper left, and LSTM-R's precision, recall, and F1-score values are significantly higher

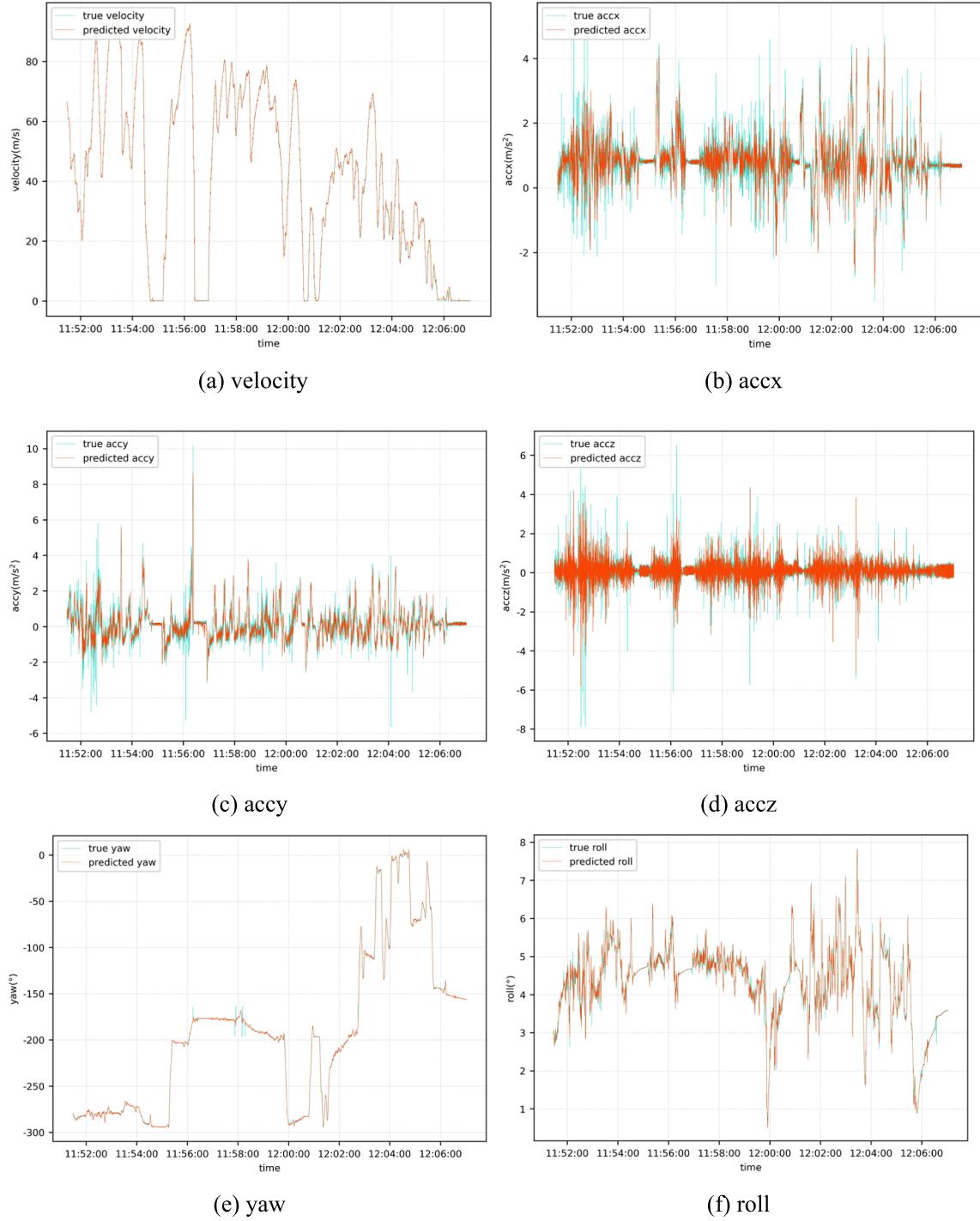


Fig. 5. Single-output regression results of LSTM network.

than those of the other five algorithms, which can be attributed to the following two factors. First, LSTM-R uses historical data to fit the current data based on time windows whereby the temporal information for abnormal driving behavior can be learned adequately. By contrast, AdaCost, SMOTEBoost, and EasyEnsemble take each frame as an input, which can result in easily misclassifying some noise points as abnormal driving. Also, LightGBM-R and LR-R cannot capture the continuity of driving behavior and thus fail to fit the vehicle kinematic data effectively. Second, the multi-dimensional spatial distributions of vehicle kinematic features for normal and abnormal driving behaviors are complex and highly distinct, which would undermine the effects of the techniques inherent in AdaCost,

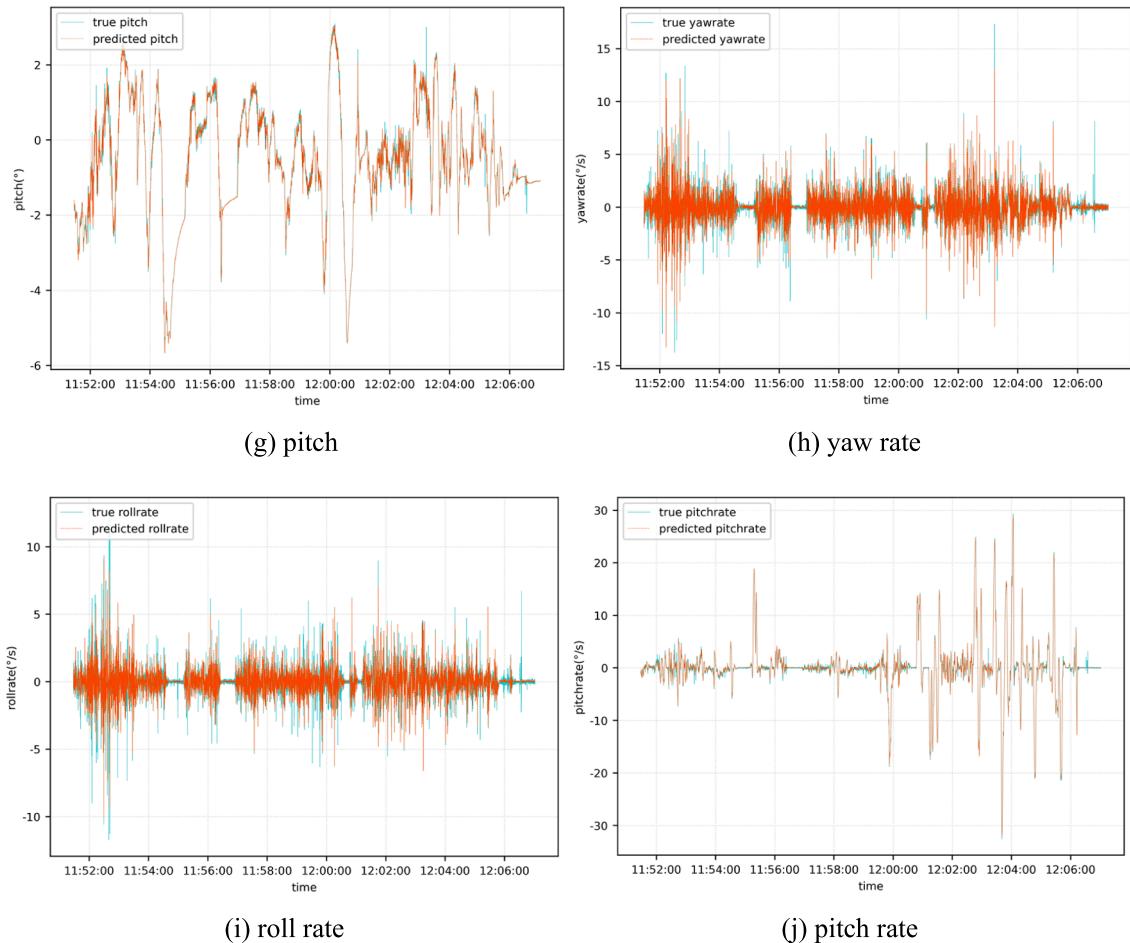


Fig. 5. (continued).

SMOTEBoost, and EasyEnsemble, especially with regard to the undersampling and synthetic oversampling techniques inherent in EasyEnsemble and SMOTEBoost, respectively.

4.7. Detection effects on test route map

[Fig. 13](#) shows the results for the abnormal driving behavior detection plotted on the test route map. As shown, the detection rate for abnormal driving is high, and only a small number of normal driving behavior instances have been misidentified as abnormal driving behavior.

5. Conclusion

In this paper, an LSTM-R algorithm is proposed to detect different types of abnormal driving behavior in real time. First, using historical data as input, an LSTM network was used to fit the vehicle kinematic data at the current moment to obtain residuals. Next, a time window-based residual algorithm was designed and implemented to detect abnormal driving behavior according to the magnitude and continuity of the residuals. In order to verify the effectiveness of the proposed LSTM-R algorithm, an experimental test was conducted in Nanjing, China. A smartphone application was used to collect vehicle kinematic data while instances of different types of abnormal driving behavior were recorded manually. Also, AdaCost, SMOTEBoost, EasyEnsemble, LightGBM-R, and LR-R were employed for comparison with the proposed algorithm.

The results show that: (1) single-output regression outperforms multi-output regression; (2) LSTM-R is not particularly sensitive to its parameters within a certain range of values; (3) optimizing the structure and parameters of the LSTM network using the validation set in the regression stage is necessary to control the degree-of-fit to optimize the detection performance; (4) the proportion of abnormal driving behavior in the training set does not significantly affect the detection results, indicating that LSTM-R can be applied to cases where abnormal driving data are sparse (which is typically the case); and (5) LSTM-R significantly outperforms the other five compared algorithms in terms of precision, recall, F1-score, and AUC, indicating LSTM-R's superiority and effectiveness.

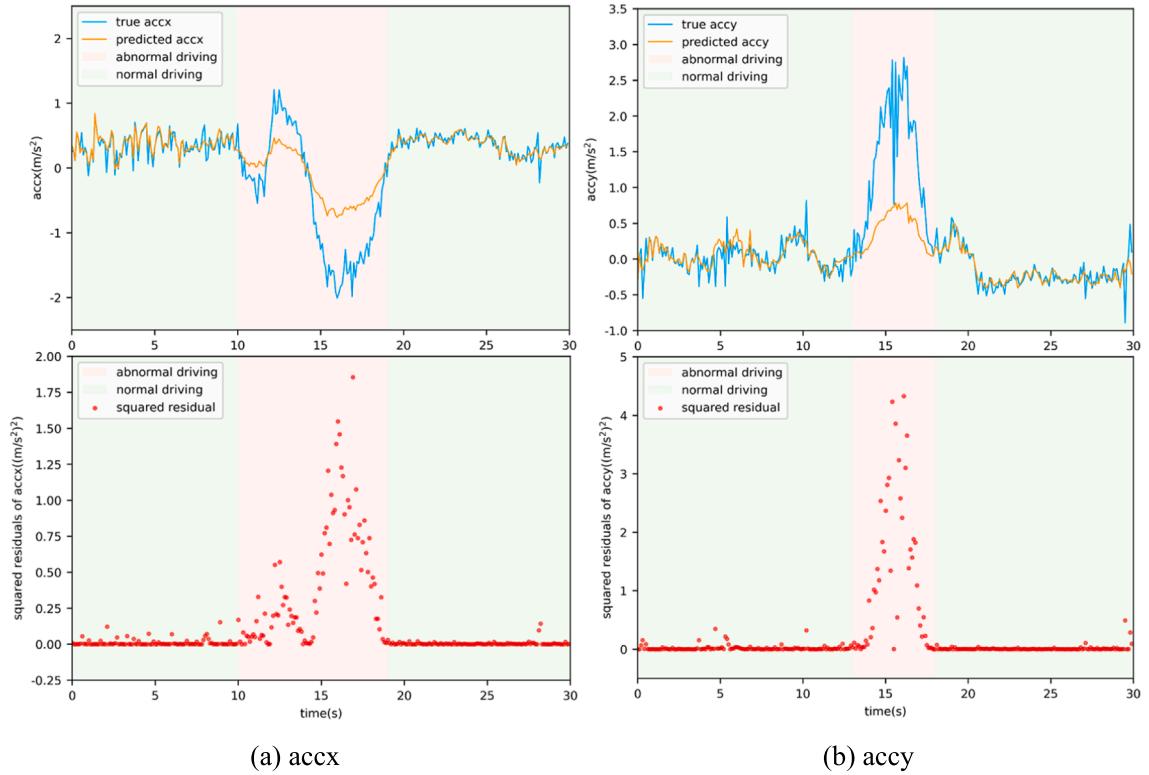


Fig. 6. Residuals of abnormal driving behavior.

Fig. 7. Impacts of l and d parameters on driving behavior detection results.

For practical applications, the proposed framework can be embedded in advanced driver assistance systems so that real-time detection and feedback for abnormal driving behavior can be utilized to improve drivers' driving habits and roadway safety. The detected abnormal driving segments can be used to understand the essential characteristics of vehicular motions during reckless driving. They also can be employed to quantify the risk that is sometimes associated with autonomous vehicles by mining their deep patterns through various deep learning methods (Ryan et al., 2021). In a connected vehicle environment, the abnormal driving

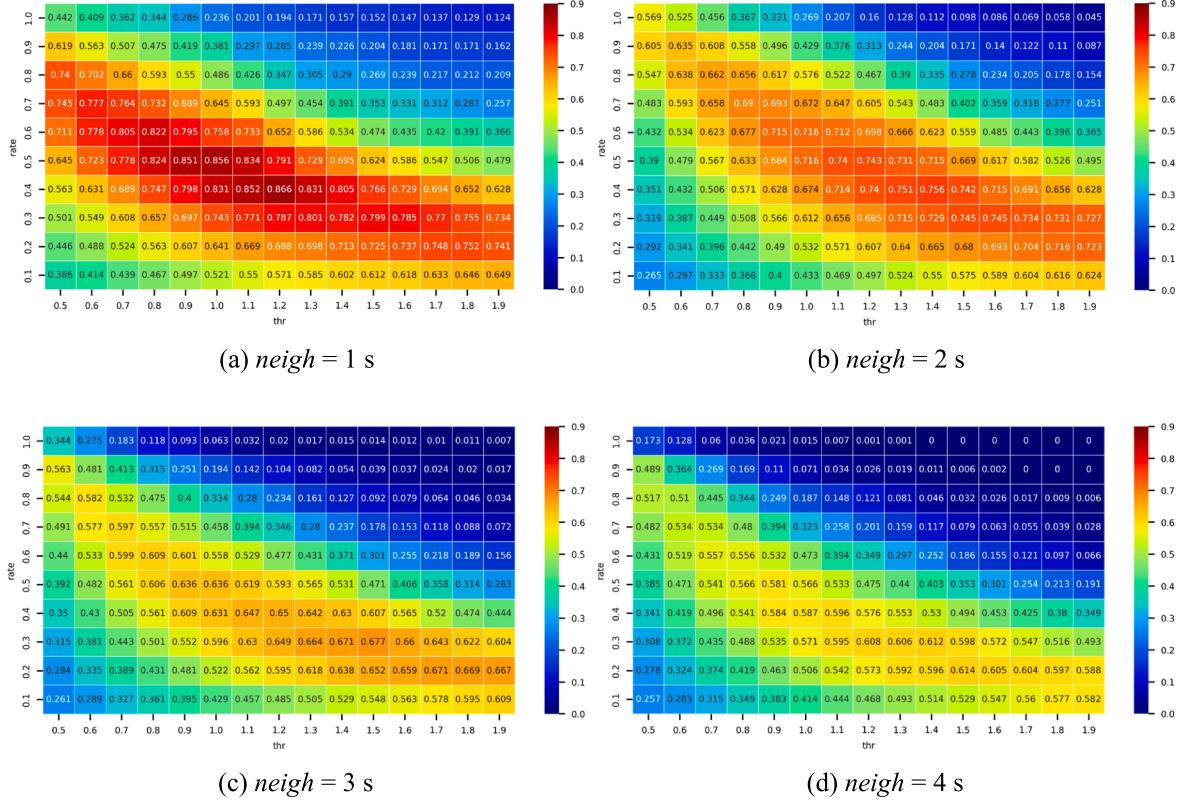
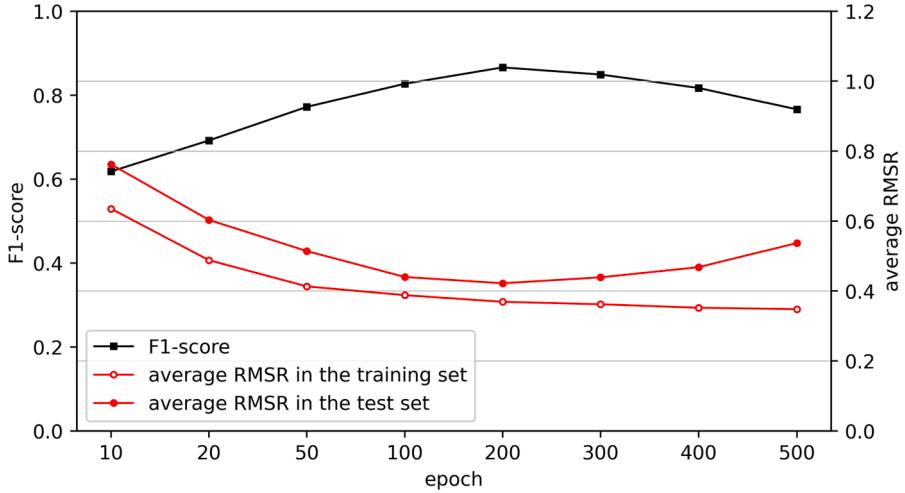
Fig. 8. Impacts of $neigh$, thr , and $rate$ on detection results.

Fig. 9. Effects of degree-of-fit of LSTM network on F1-score and average root mean square residual (RMSR) values.

information or warning information of the subject vehicle can be sent to the surrounding vehicles or pedestrians in real time through the Internet of Vehicles or smartphones so that road users can be aware of surrounding danger and thus avoid collisions (Chen and Chen, 2021). By analyzing the characteristics and frequency of drivers' abnormal driving behaviors and combining them with driver demographic characteristics, insurance companies can formulate personalized insurance services, i.e., pay-as-you-drive provisions (Arumugam and Bhargavi, 2019). Fleet-operating companies can evaluate drivers' performance based on abnormal driving conditions to improve drivers' driving skills in efforts to decrease roadway crash risks as well as reduce fuel consumption (Zhang et al., 2017; Zhang et al., 2020). Moreover, the proposed abnormal detection framework can be extended to other application fields, such as fault detection in manufacturing, leak detection in gas-chemical processes, cyber intrusion detection, and structural health monitoring of

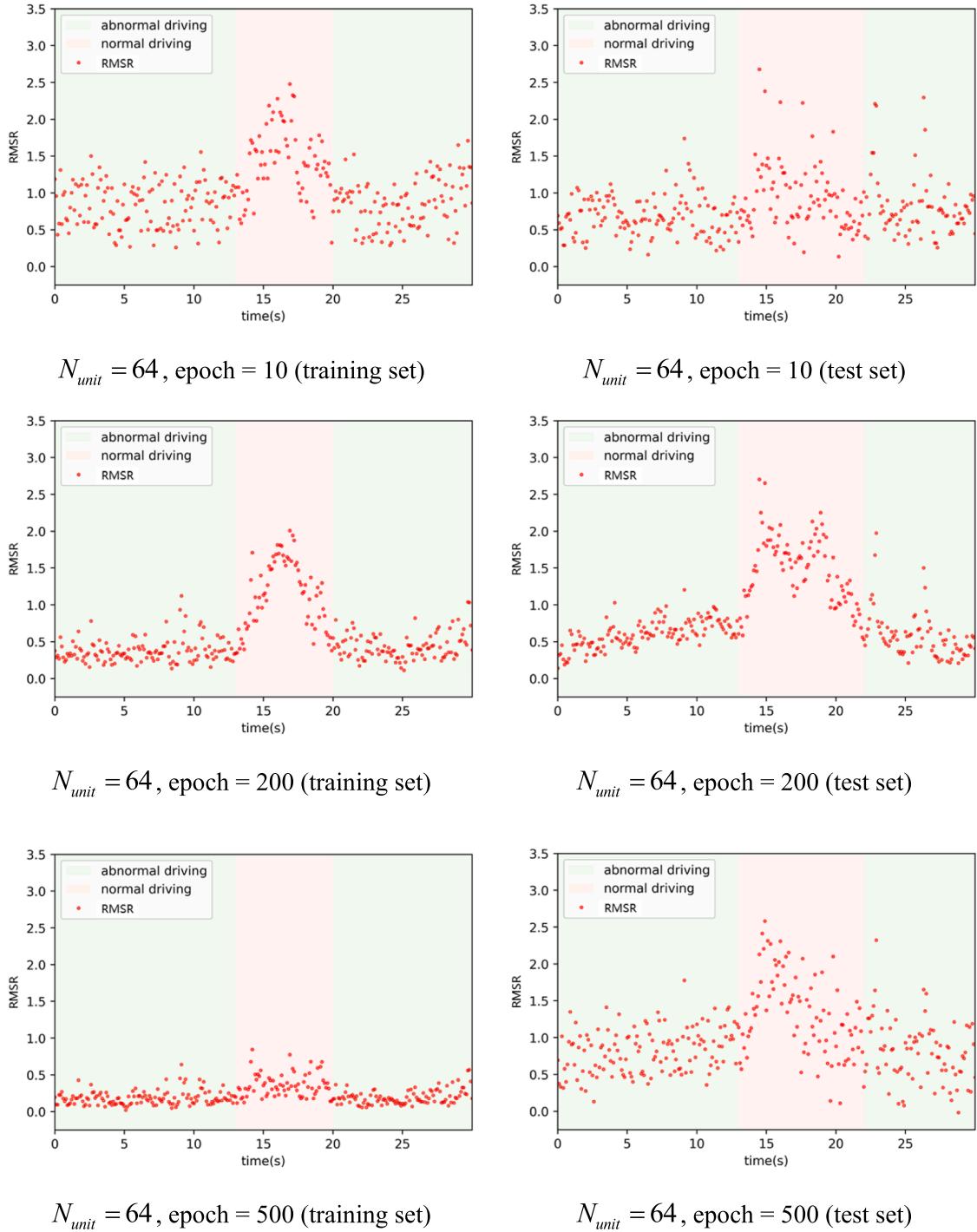


Fig. 10. Effects of degree-of-fit of LSTM network on root mean square residual (RMSR) values for abnormal driving behavior.

infrastructure (Choi et al., 2021).

In planned future research by the authors, more advanced devices and systems, such as VBOX GPS data loggers and on-board diagnostics, will be used to collect high-quality data. The driving process is a dynamic interaction among humans, vehicles, roadways, and the environment, so multisource data, including the driver's demographic, physiological, and psychological characteristics, weather and road conditions, and surrounding vehicles, can be integrated to improve abnormal driving behavior detection. In addition, different drivers have different driving styles and habits, so the number of subjects should be increased to extend the

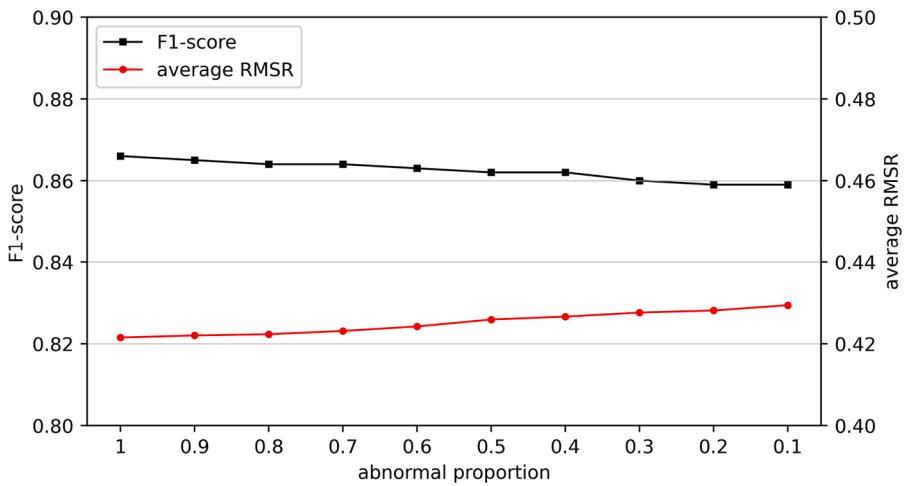


Fig. 11. F1-score and average root mean square residual (RMSR) values for different abnormal proportions.

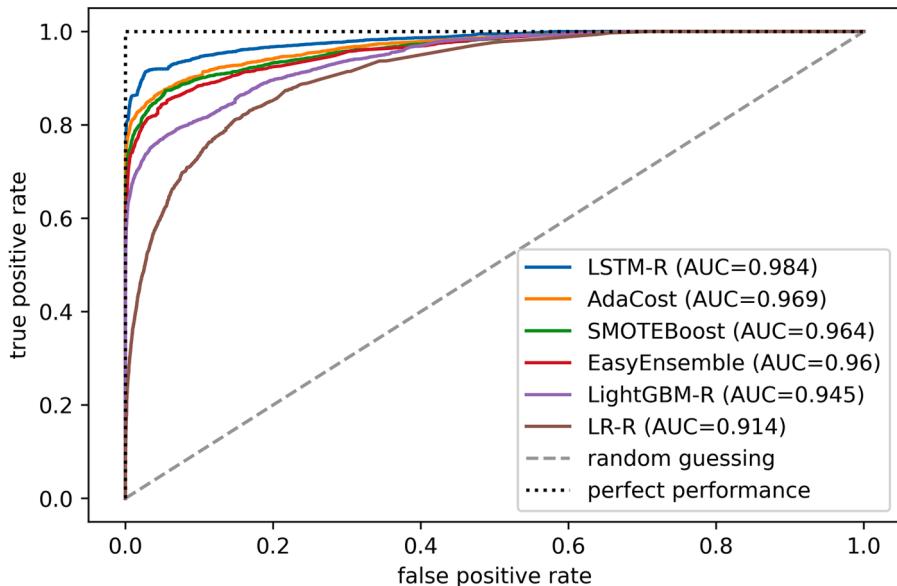


Fig. 12. Receiver operating characteristic curve of each algorithm. Note: AUC is ‘area under the curve’.

Table 5
Metrics of Each Compared Algorithm.

Algorithm	Precision	Recall	F1-score
LSTM-R	0.772	0.987	0.866
AdaCost	0.663	0.965	0.786
SMOTEBoost	0.611	0.951	0.744
EasyEnsemble	0.598	0.946	0.733
LightGBM-R	0.574	0.903	0.702
LR-R	0.525	0.742	0.615

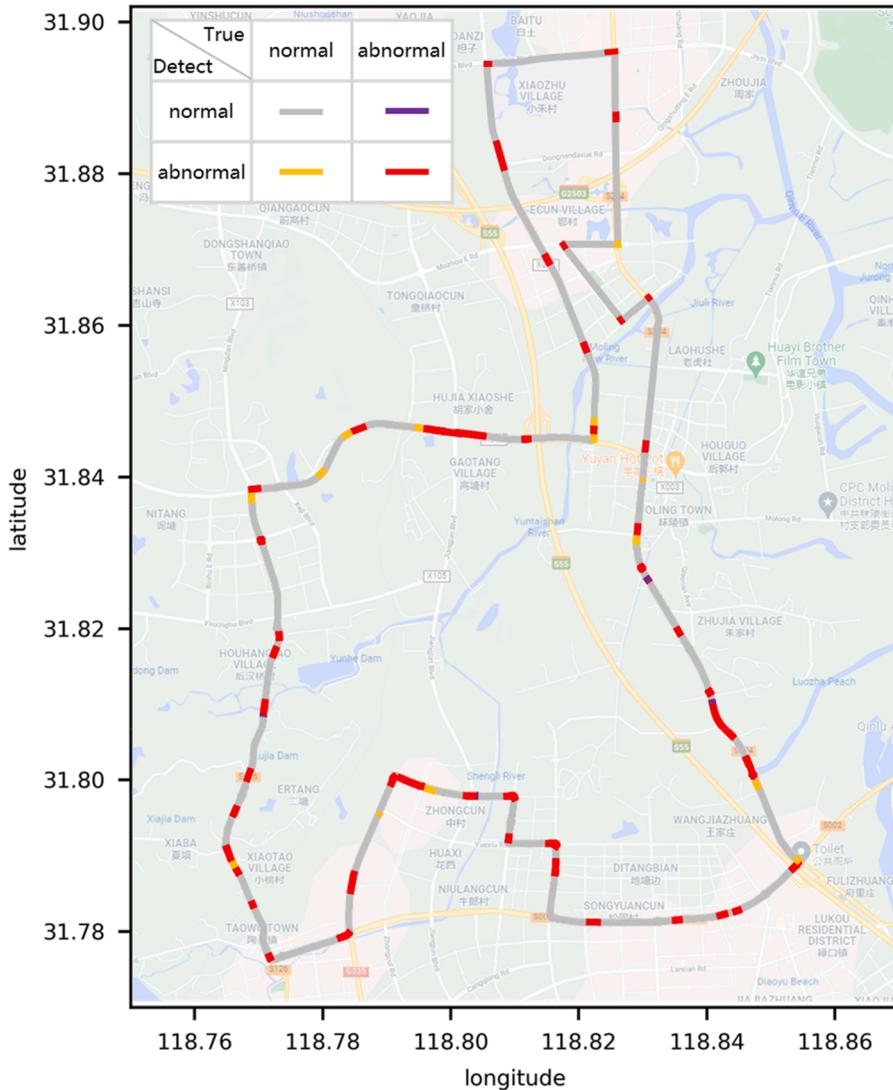


Fig. 13. Detection results for abnormal driving behavior plotted on test route map.

universality of the research. Moreover, residuals of different vehicle kinematic features will be considered for the fine-grained detection of abnormal driving behavior in future research.

CRediT authorship contribution statement

Yongfeng Ma: Project administration, Conceptualization, Funding acquisition, Supervision. **Zhuopeng Xie:** Formal analysis, Investigation, Methodology, Visualization, Writing - original draft. **Shuyan Chen:** Formal analysis, Supervision, Validation, Writing - review & editing. **Fengxiang Qiao:** Editing. **Zeyang Li:** Data collection.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (52172342).

References

- Alham, N.K., Li, M., Liu, Y., Hammoud, S., 2011. A MapReduce-based distributed SVM algorithm for automatic image annotation. *Comput. Math. Appl.* 62 (7), 2801–2811.
- Ansar, M.S., Ma, Y., Chen, S., Tang, K., Zhang, Z., 2021. Investigating the trip configured causal effect of distracted driving on aggressive driving behavior for e-hailing taxi drivers. *J. Traffic Transport. Eng. (English Edition)* 8 (5), 725–734.
- Arumugam, S., Bhargavi, R., 2019. A survey on driving behavior analysis in usage based insurance using big data. *J. Big Data* 6, 1–21.
- Brombacher, P., Masino, J., Frey, M., Gauterin, F., 2017. Driving event detection and driving style classification using artificial neural networks. In: Proceedings of 2017 IEEE International Conference on Industrial Technology (ICIT), Toronto, ON. 2017/3/22 - 2017/3/25. IEEE, 997–1002.
- Bucsuházy, K., Matuchová, E., Zúvala, R., Moravcová, P., Kostková, M., Mikulec, R., 2020. Human factors contributing to the road traffic accident occurrence. *Transp. Res. Procedia* 45, 555–561.
- Chawla, N. V., Lazarevic, A., Hall, L. O., Bowyer, K. W., 2003. SMOTEBoost: Improving prediction of the minority class in boosting. In: Proceedings of European Conference on Principles of Data Mining and Knowledge Discovery. Springer, Berlin, Heidelberg, 107–119.
- Chen, L.-W., Chen, H.-M., 2021. Driver behavior monitoring and warning with dangerous driving detection based on the Internet of Vehicles. *IEEE Trans. Intell. Transp. Syst.* 22 (11), 7232–7241.
- Choi, K., Yi, J., Park, C., Yoon, S., 2021. Deep learning for anomaly detection in time-series data: Review, analysis, and guidelines. *IEEE Access* 9, 120043–120065.
- Ding, N., Ma, HaoXuan, Gao, H., Ma, YanHua, Tan, GuoZhen, 2019. Real-time anomaly detection based on long short-term memory and Gaussian Mixture Model. *Comput. Electr. Eng.* 79, 106458.
- Elamrani Abou Ellassad, Z., Mousannif, H., Al Moatassime, H., 2020. A real-time crash prediction fusion framework: An imbalance-aware strategy for collision avoidance systems. *Transport. Res. Part C: Emerg. Technol.* 118, 102708.
- Fan, P., Guo, J., Wang, Y., Wijnands, J.S., 2022. A hybrid deep learning approach for driver anomalous lane changing identification. *Accid. Anal. Prev.* 171, 106661.
- Fan, W., Stolfo, S.J., Zhang, J., Chan, P.K., 1999. AdaCost: Misclassification cost-sensitive boosting. *Proc. Int. Conf. Mach. Learn. (ICML)* 97–105.
- Fan, X., Wang, F., Song, D., Lu, Y., Liu, J., 2021. GazMon: Eye gazing enabled driving behavior monitoring and prediction. *IEEE Trans. Mob. Comput.* 20, 1420–1433.
- Fu, Y., Liu, X., Sarkar, S., Wu, T., 2021. Gaussian mixture model with feature selection: An embedded approach. *Comput. Ind. Eng.* 152, 107000.
- Han, S., Zhong, X., Shao, H., Xu, T., Zhao, R., Cheng, J., 2021. Novel multi-scale dilated CNN-LSTM for fault diagnosis of planetary gearbox with unbalanced samples under noisy environment. *Meas. Sci. Technol.* 32, 124002.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9, 1735–1780.
- Hong, J.-H., Margines, B., Dey, A.K., 2014. A smartphone-based sensing platform to model aggressive driving behaviors. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '14: CHI Conference on Human Factors in Computing Systems, Toronto, Ontario, Canada. 26 04 2014 01 05 2014. ACM, New York, NY, USA, pp. 4047–4056.
- Hu, H., Wang, Q., Cheng, M., Gao, Z., 2021. Cost-sensitive semi-supervised deep learning to assess driving risk by application of naturalistic vehicle trajectories. *Expert Syst. Appl.* 178, 115041.
- Hu, J., Zhang, X., Maybank, S., 2020. Abnormal driving detection with normalized driving behavior data: A deep learning approach. *IEEE Trans. Veh. Technol.* 69, 6943–6951.
- Huang, W., Liu, X., Luo, M., Zhang, P., Wang, W., Wang, J., 2019. Video-based abnormal driving behavior detection via deep learning fusions. *IEEE Access* 7, 64571–64582.
- Jia, S., Hui, F., Li, S., Zhao, X., Khattak, A.J., 2020. Long short-term memory and convolutional neural network for abnormal driving behaviour recognition. *IET Intel. Transport Syst.* 14, 306–312.
- Khodairy, M.A., Abosamra, G., 2021. Driving behavior classification based on oversampled signals of smartphone embedded sensors using an optimized Stacked-LSTM neural network. *IEEE Access* 9, 4957–4972.
- Kieu, T., Yang, B., Guo, C., Jensen, C.S., 2019. Outlier detection for time series with recurrent autoencoder ensembles. In: Proceedings of Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19), Macao, China. 2019/8/10 - 2019/8/16. International Joint Conferences on Artificial Intelligence Organization, California, 2725–2732.
- Liu, X.Y., Wu, J., Zhou, Z.H., 2008. Exploratory undersampling for class-imbalance learning. *IEEE Trans. Syst., Man, Cybernetics Part B (Cybernetics)* 39, 539–550.
- Louw, T., Merat, N., 2017. Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation. *Transport. Res. Part C: Emerg. Technol.* 76, 35–50.
- Lu, M., Hu, Y., Lu, X., 2022. Pose-guided model for driving behavior recognition using keypoint action learning. *Signal Process. Image Commun.* 100, 116513.
- Ma, Y., Zhang, Z., Chen, S., Yu, Y., Tang, K., 2019. A comparative study of aggressive driving behavior recognition algorithms based on vehicle motion data. *IEEE Access* 7, 8028–8038.
- Ma, Y., Tang, K., Chen, S., Khattak, A.J., Pan, Y., 2020. On-line aggressive driving identification based on in-vehicle kinematic parameters under naturalistic driving conditions. *Transport. Res. Part C: Emerg. Technol.* 114, 554–571.
- Ma, Y., Li, W., Tang, K., Zhang, Z., Chen, S., 2021a. Driving style recognition and comparisons among driving tasks based on driver behavior in the online car-hailing industry. *Accid. Anal. Prev.* 154, 106096.
- Ma, Y., Xie, Z., Chen, S., Wu, Y., Qiao, F., 2021b. Real-time driving behavior identification based on multi-source data fusion. *Int. J. Environ. Res. Public Health* 19, 348.
- Mantouka, E., Barmpounakis, E., Vlahogianni, E., Golias, J., 2021. Smartphone sensing for understanding driving behavior: Current practice and challenges. *Int. J. Transp. Sci. Technol.* 10, 266–282.
- Martín de Diego, I.S., Siordia, O., Crespo, R., Conde, C., Cabello, E., 2013. Analysis of hands activity for automatic driving risk detection. *Transport. Res. Part C: Emerg. Technol.* 26, 380–395.
- Nguyen, H.-P., Baraldi, P., Zio, E., 2021. Ensemble empirical mode decomposition and long short-term memory neural network for multi-step predictions of time series signals in nuclear power plants. *Appl. Energy* 283, 116346.
- Rastgooh, M.N., Nakisa, B., Maire, F., Rakotonirainy, A., Chandran, V., 2019. Automatic driver stress level classification using multimodal deep learning. *Expert Syst. Appl.* 138, 112793.
- Ryan, C., Murphy, F., Mullins, M., 2021. End-to-end autonomous driving risk analysis: A behavioural anomaly detection approach. *IEEE Trans. Intell. Transp. Syst.* 22, 1650–1662.
- Shi, B., Xu, L., Hu, J., Tang, Y., Jiang, H., Meng, W., Liu, H., 2015. Evaluating driving styles by normalizing driving behavior based on personalized driver modeling. *IEEE Trans. Syst., Man, Cybernet.: Syst.* 45, 1502–1508.
- Skrickij, V., Šabanović, E., Žuraulis, V., 2020. Autonomous road vehicles: Recent issues and expectations. *IET Intel. Transport Syst.* 14, 471–479.
- Tasca, L., 2000. A review of the literature on aggressive driving research. *Citeseer* 37–46.
- Vindhya, Venkatraman, Richard, Christian M., Magee, Kelly, Battelle Memorial Institute, Kristie Johnson, N., 2020. Countermeasures That Work: A Highway Safety Countermeasure Guide for State Highway Safety Offices, 10th Edition.
- Wang, C., Xie, Y., Huang, H., Liu, P., 2021. A review of surrogate safety measures and their applications in connected and automated vehicles safety modeling. *Accid. Anal. Prev.* 157, 106157.
- Wang, K., Xue, Q., Xing, Y., Li, C., 2020. Improve aggressive driver recognition using collision surrogate measurement and imbalanced class boosting. *Int. J. Environ. Res. Public Health* 17.
- World Health Organization, 2021. WHO kicks off a decade of action for road safety. Retrieved from <https://www.who.int/news/item/28-10-2021-who-kicks-off-a-decade-of-action-for-road-safety>. Accessed January 7, 2022.
- Xie, J., Hilal, A.R., Kulic, D., 2018. Driving maneuver classification: A comparison of feature extraction methods. *IEEE Sensors* 18, 4777–4784.
- Xie, J., Hu, K., Li, G., Guo, Y., 2021. CNN-based driving maneuver classification using multi-sliding window fusion. *Expert Syst. Appl.* 169, 114442.

- Xing, Y., Lv, C., Cao, D., Velenis, E., 2021. Multi-scale driver behavior modeling based on deep spatial-temporal representation for intelligent vehicles. *Transport. Res. Part C: Emerg. Technol.* 130, 103288.
- Yu, J., Chen, Z., Zhu, Y., Chen, Y., Kong, L., Li, M., 2017. Fine-grained abnormal driving behaviors detection and identification with smartphones. *IEEE Trans. Mob. Comput.* 16, 2198–2212.
- Yuan, Y., Lu, Y., Wang, Q., 2020. Adaptive forward vehicle collision warning based on driving behavior. *Neurocomputing* 408, 64–71.
- Zhang, M., Chen, C., Wo, T., Xie, T., Bhuiyan, M.Z.A., Lin, X., 2017. SafeDrive: Online driving anomaly detection from large-scale vehicle data. *IEEE Trans. Ind. Inf.* 13, 2087–2096.
- Zhang, Y., Li, H., Sze, N.N., Ren, G., 2021. Propensity score methods for road safety evaluation: Practical suggestions from a simulation study. *Accid. Anal. Prev.* 158, 106200.
- Zhang, H., Sun, J., Tian, Y., 2020. The impact of socio-demographic characteristics and driving behaviors on fuel efficiency. *Transp. Res. Part D: Transp. Environ.* 88, 102565.
- Zhang, J., Wu, Z., Li, F., Luo, J., Ren, T., Hu, S., Li, W., Li, W., 2019. Attention-based convolutional and recurrent neural networks for driving behavior recognition using smartphone sensor data. *IEEE Access* 7, 148031–148046.
- Zhuang, Z., Lv, H., Xu, J., Huang, Z., Qin, W., 2019. A deep learning method for bearing fault diagnosis through stacked residual dilated convolutions. *Appl. Sci.* 9, 1823.