Transportation Research Part C 136 (2022) 103561

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Driver distraction detection based on vehicle dynamics using naturalistic driving data

Xuesong Wang [a,b,c,*], Rongjiao Xu [a,b], Siyang Zhang [a], Yifan Zhuang [d], Yinhai Wang [d]

[a] College of Transportation Engineering, Tongji University, Shanghai 201804, China
[b] The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Shanghai, 201804, China
[c] National Engineering Laboratory for Integrated Optimization of Road Traffic and Safety Analysis Technologies, 88 Qianrong Rd, Wuxi 214151, China
[d] Department of Civil and Environmental Engineering, University of Washington, Seattle, WA 98195, USA

A R T I C L E I N F O

Keywords:
Phone use detection
Naturalistic driving
Bidirectional long short-term memory network
Attention mechanism
Vehicle dynamics

A B S T R A C T

Distracted driving such as phone use during driving is risky, as it increases the probability of severe crashes. Detecting distraction using Naturalistic Driving Studies was attempted in existing studies, and most of them used facial motions, which would be highly influenced by light conditions and algorithm effectiveness, still could not fully indicate auditory and physical distractions. This study aims to optimize Long Short-Term Memory (LSTM) model for phone usage detection based on vehicle dynamics sensor data from Shanghai Naturalistic Driving Study (SH-NDS), China. A total of 1244 phone use events were extracted from videos of SH-NDS, and analyzed against focus driving baseline. Performance attributes included speed, longitudinal acceleration, lateral acceleration, lane offset, and steering wheel rate. Their mean, standard deviation, and predicted error (PE) were calculated, and derived 15 indicators. A Bidirectional layer and attention mechanism were added to the LSTM model for higher accuracy. Results showed that besides the mean and standard deviation of steering wheel rate, all the other 13 indicators were significant and effective in the model. The Bidirectional Long Short-Term Memory (Bi-LSTM) model reached a promising result of approximately 91.2% accuracy using 5-fold cross validation, which was better than other machine learning methods such as recurrent neural network, support vector machine, k-nearest neighbor, and adaptive boosting. This Bi-LSTM model with attention mechanism could potentially be applied in advanced driving assistant systems to warn driver and reduce phone involved distracted driving.

1. Introduction

The exploding growth of vehicles brings an unprecedented concern to traffic safety. Among all traffic crashes, especially fatal crashes, distracted driving is a major cause based on the report from National Highway Traffic Safety Administration (NHTSA) (NHTSA, 2015). In 2018, 2841 people were killed and over 400,000 people got injured due to distracted driving in the United States (NCSA, 2017). NHTSA defines distracted driving as an activity that diverts attention from driving, e.g., in-vehicle or phone calls, texting, eating and drinking, and fiddling with in-vehicle systems (NCSA, 2017). From 2010 to 2016, 9.5% of fatal crashes involved a distracted driver, and cellphone usage was one of the main distracting factors (Oviedo-Trespalacios et al., 2016).

* Corresponding author at: College of Transportation Engineering, Tongji University, Shanghai 201804, China.
E-mail address: wangxs@tongji.edu.cn (X. Wang).

https://doi.org/10.1016/j.trc.2022.103561
Received 21 December 2020; Received in revised form 27 December 2021; Accepted 7 January 2022
Available online 24 January 2022
0968-090X/

Currently, there are two major methods for driver status detection. One is based on the onboard video cameras toward the driver's face. And the driver status is recognized by computer vision algorithms that extract drivers' facial or physical features (Liang et al., 2007; Naqvi et al., 2018). The representative of this method is the Driver Monitoring System (DMS), which is widely installed on commercial vehicles (Wang et al., 2018). The system will send an alarm when the driver is detected as distracted status. However, the false alarm rate is unsatisfying due to various reasons, e.g., severe weather, light condition, and driving instability. Another method is analyzing vehicle dynamics captured by sensors mounted on the vehicle (Ye et al., 2017; Atiquzzaman et al., 2018). Some studies utilize vehicle dynamic data obtained from controller area network (CAN), GPS receivers, and/or inertial sensors for distraction detection (Kircher and Ahlstrom, 2010; Wollmer et al., 2011; Ye et al., 2017). The advantages of CAN data include easy access and good stability. The data sources mainly come from simulator studies and naturalistic driving studies (NDS) (Kircher and Ahlstrom, 2010; Wollmer et al., 2011; Tango and Botta, 2013; Wang and Xu, 2016; Ye et al., 2017; Atiquzzaman et al., 2018; Kouchak and Gaffar, 2020; Zhang et al., 2020). Compared with simulator studies, NDS is more realistic and can reflect natural driving status more accurately. Therefore, NDS is selected as the data source in this study to analyze the impacts of distracted driving on vehicle dynamics.

Previous studies indicate that driver distraction has negative impacts on vehicle dynamics significantly and increases the risk of crashes and safety–critical events. Typical phenomena of vehicle dynamics on distracted status include reduced speed (Fitch et al., 2013; Yannis et al., 2014; Choudhary and Velaga, 2017a,b), increased speed fluctuation (Yan et al., 2018), and worse lane-keeping performance (Choudhary and Velaga, 2017a,b; Jeong and Liu, 2019). This indicates a strong correlation, and it is possible to derive the distraction status from the vehicle dynamics.

Considering complicated variables and nonlinear relations, machine learning models perform significantly better than traditional statistical approaches in distraction detection (Zhang et al., 2004; Kircher and Ahlstrom, 2010). Support vector machine (SVM) and adaptive boosting (AdaBoost) (Atiquzzaman et al., 2018) are used for distraction detection and reach around 80% accuracy. However, these machine learning methods still have drawbacks like no memory and are not able to take past events into consideration. For example, phone use is a continuous event and requires capability of analyzing time-series data. Long Short-Term Memory (LSTM) is one kind of Recurrent Neural Network (RNN) that is able to selectively memorize patterns for a long duration and detect the driver distraction (Hochreiter and Schmidhuber, 1997).

This paper proposes a LSTM-based model to detect driver's distraction status and evaluates this model using vehicle dynamic data from Shanghai Naturalistic Driving Study (SH-NDS). This work concentrates on utilizing the vehicle dynamic data to detect the driver's phone use during driving. There are three major original contributions of this study: (1) used massive naturalistic driving data, which is much more realistic than driving simulator data, (2) used vehicle dynamics data which is easily accessible, (3) improved model by adding attention mechanism and bidirectional layer to Long Short-Term Memory model. This combination is significant.

2. Literature review

Distraction detection is an important topic in both academic and industrial fields as it is highly related to traffic safety. Previous studies attempted to detect different types of driver distraction using statistical models and machine learning models. In statistical models, typical driving performance variables include speed, lane position, acceleration, and steering wheel angle, along with facial and eye movements. Statistical distraction detection models included decision tree (Zhang et al., 2004) and logistic regression (Kircher and Ahlstrom, 2010), however accuracy reached only 76.2%. Machine learning models such as SVM (Ersal et al., 2010; Liao et al., 2016; Li et al., 2018), Hybrid Bayesian Network (HBN) (Liang and Lee, 2014; Ryu et al., 2015), were introduced to improve accuracy. Several evaluation metrics, e.g., mean value, standard deviation, and predicted error (PE), were computed within each observation window. Among these studies, models based on NDS were found to be more accurate than ones based on driving simulator data. Recently, researchers developed a state-space model to detect visual-manual distraction tasks like texting (Sun, et al., 2021), which may not be the most appropriate as state-space model handles stable data well, but driving data is not always stable, especially when the driver is distracted.

In addition to traditional machine learning methods, neural networks, as a powerful tool to solve nonlinear problems, were also used for driver distraction detection (Tango and Botta, 2013; Im et al., 2014; Yang et al., 2015; Ye et al., 2017). Extreme learning machine (ELM) is one type of feedforward neural network that achieved higher accuracy (87.0%) than SVM (82.9%) (Yang et al., 2015). Different neural networks, e.g., Orthogonal Minimum Square (OMS) - Radial Basis Function (RBF), studied phone call distraction and achieved great success (Luo et al., 2019). Kouchak and Gaffar (2019) used deep neural networks to detect visual-manual distraction in simulator studies and the accuracy increased by 67% compared to other neural networks. Considering the distraction detection is studying sequence data, Wollmer et al. (2010) applied a unidirectional LSTM model to detect distraction caused by the in-vehicle information system (IVIS), using six vehicle dynamic signals as steering wheel angle, throttle position, speed, heading angle, lateral deviation, and head rotation. Kouchak and Gaffar (2020) studied the performance of stacked LSTM with attention using ten vehicle dynamics variables based on a simulator study, turning out that LSTM network with a combination of attention and bidirectional layer overcame overfitting problem and reduced model error.

Previous literature presented various methods to detect distraction, but either data used facial recognition only without vehicle status, or the accuracy was not ideal. Using vehicle dynamics data has the advantages of easy accessibility and sufficiency in detection. LSTM has the ability of handling time-series data, and attention mechanism and bidirectional layer greatly improve detection accuracy (Kouchak and Gaffar, 2020). However, this conclusion was from simulator studies, which was not as realistic as using NDS data. This paper adopted LSTM network to handle massive time-series vehicle dynamics data from Shanghai Naturalistic Driving Study and added bidirectional layer and attention mechanism to improve phone use distraction detection accuracy.

3. Methodology

This study utilized vehicle dynamics data to detect driver distraction involving phone use during driving. A bidirectional layer and attention mechanism were added to a LSTM network to recognize driving behavior patterns highly correlated to phone use. The procedure undertook three steps: data collection from Shanghai Naturalistic Driving Study, data processing, and LSTM model development.

3.1. Data collection

The data utilized in this study came from the Shanghai Naturalistic Driving Study (SH-NDS) (Zhu et al., 2018), a joint study conducted by Tongji University, General Motors (GM), and the Virginia Tech Transportation Institute (VTTI). Data collection period was three years from 2012 to 2015. Five GM light vehicles equipped with the Strategic Highway Research Program 2 (SHRP2) data acquisition systems were used to collect naturalistic driving data. The data acquisition systems (DAS) integrated an interface box collecting vehicle Controller Area Network (CAN) data, an accelerometer for longitudinal and lateral acceleration, a radar system that measured range and range rate to the lead vehicle and vehicles in adjacent lanes, a light meter, a temperature/humidity sensor, a Global Positioning System (GPS) sensor, and four synchronized cameras to validate the sensor-based findings (Yang et al., 2019). The sampling frequency of data collection was 10 Hz.

Sixty licensed Shanghai drivers were recruited from the public for the SH-NDS, with distributions of genders, ages, and driving experience matching with the general Chinese driver population. Five experimental vehicles were used, with each participant driving one of the instrumented vehicles for two months. Due to data acquisition system malfunction, only 52 drivers' data were valid, of which 12 drivers were female and 40 drivers were male, aging from 25 to 59 (mean = 38.7, standard deviation = 9.0).

For each driver, two trips were randomly selected from each of the five trip duration categories of 0–20 min, 20–40 min, 40–60 min, 60–80 min, and over 80 min, which obtained 10 trips from each driver. Distraction events were extracted from the videos recorded by four synchronized cameras and categorized into different phone use types. The phone use distraction includes browsing, changing phone position, cleaning the phone, phone calls, listening to music, holding in hand, texting, voice messaging, playing games, and combination. The frequency and duration of each kind of phone use tasks were described in Fig. 1.

The phone use distraction period started from first sight when driver glanced at the phone or taking a hand off the steering wheel to reach the phone, and ended when the driver refocused on the roadway or put both hands back on the steering wheel. All phone use events were labeled as distracted status. Focused driving periods, extracted as 20 s before and 10 s after the phone use event, were used as the baselines according to the research method raised before (Wang et al., 2020). For example, t_0 to t_1 is the 20 s baseline before distraction event, named as Baseline$_{i1}$, t_1 to t_2 is the distracted event, t_2 to t_3 is the 10 s baseline after the distraction event, named as Baseline$_{i2}$ (Fig. 2). A total of 1244 driving periods were extracted, and further divided into 1244 distracted driving periods and 2488 focused driving periods.

3.2. Data processing

From the SH-NDS time-series data, five vehicle dynamic variables, including speed, longitudinal acceleration, lateral acceleration, lane offset, and steering wheel rate were extracted, as listed in Table 1. According to previous studies, the feature of these variables significantly differed between focused driving status and distracted driving status, especially when engaged in phone use tasks (Yan et al., 2018; Jeong and Liu, 2019; Choudhary and Velaga, 2017a,b). The time-series data of five driving performance indicators were

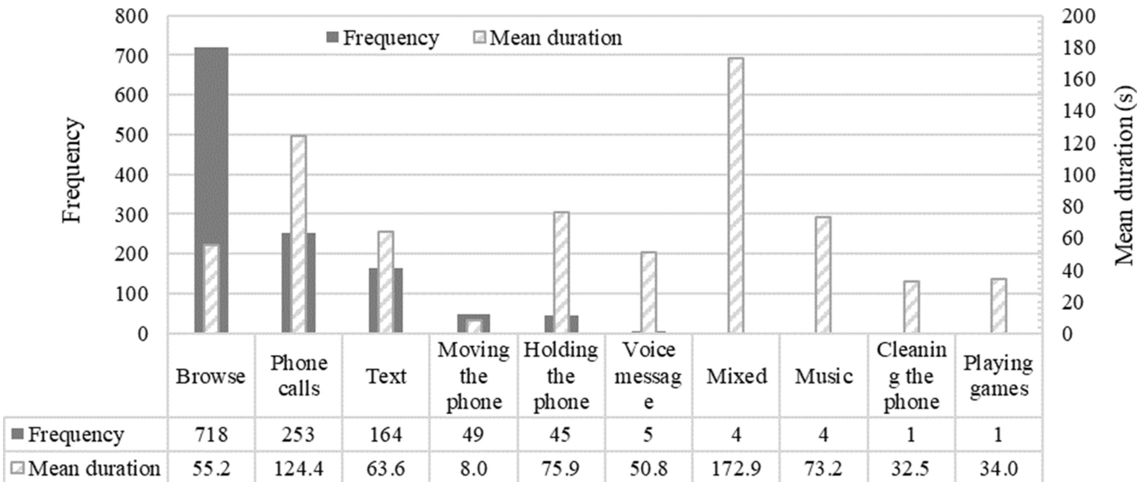


Fig. 1. Frequency and mean duration of phone use tasks.

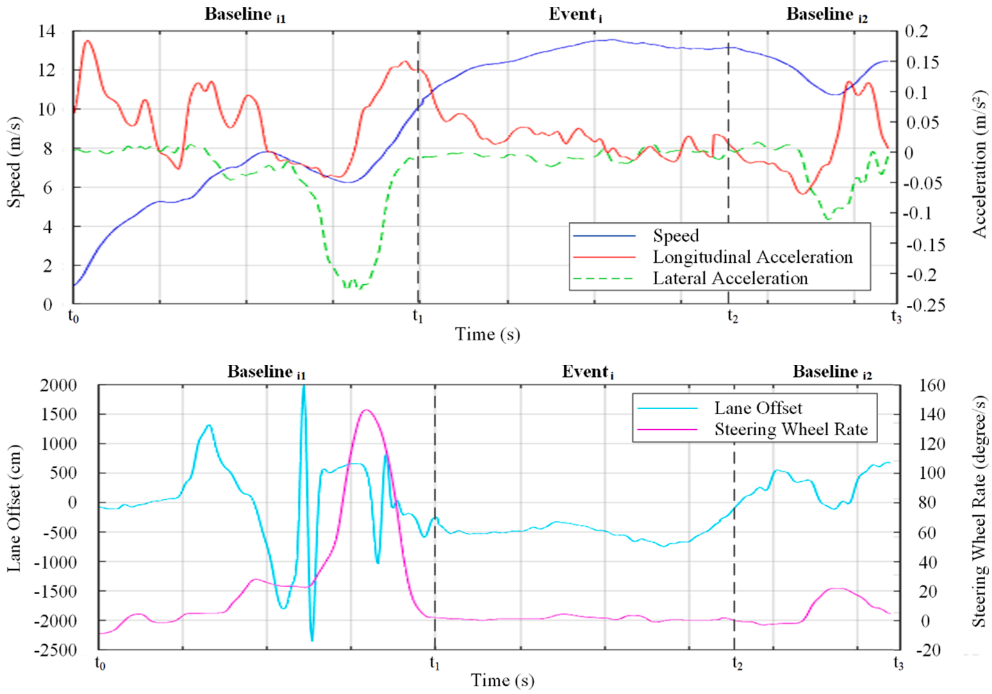


Fig. 2. Time series observation of a cell phone event.

shown in Fig. 2.

Considering the occasional signal loss, missing data was estimated by linear interpolation. When missing data took up over 85%, then that period was regarded as invalid and data would be discarded. To make sure the vehicle was not parked or idling, periods with 85% of the time having speed less than 8 km/h (about 2.22 m/s) were eliminated. The threshold of 8 km/h was based on the distraction periods extraction method used in the 100-Car Naturalistic Driving Study (Klauer et al., 2014).

After removing erroneous data, noise data were filtered by 10 data points (one second) moving window. Noise elimination and data averaging were conducted in a one-second time window by Gaussian smoothing method (Haddad and Akansu, 1991), the value of each data point is the weighted average of the surrounding datapoint values. The farther from the original data point, the smaller weights become. In each time window, the standard deviation was computed. Then the window moved forward one step of 0.1 s to compute the standard deviation of the next time window (Ye et al., 2017). The PE was computed by using the second-order Taylor approximation (Atiquzzaman et al., 2018). The predicted value of one data point was obtained by three data points before it uses second-order Taylor expansion, and PE was calculated by the absolute value of the difference between the predicted value and the actual value.

Fifteen vehicle dynamic data features, including smoothened data, standard deviation, and predicted error of the five performance variables, were obtained. The standard deviation of speed and longitudinal acceleration can reflect the stability of drivers' vehicle control in the longitudinal direction (Fitch et al., 2013), while standard deviation of lateral acceleration, lane offset, and steering wheel rate can reflect the stability of drivers' vehicle control in the lateral direction (Fitch et al., 2013). The PEs of these vehicle dynamic variables measured the smoothness of vehicle control (Atiquzzaman et al., 2018). For example, the PE of steering wheel rate measured the smoothness of steering wheel movements.

According to the distraction status, driving distraction status was labeled as two classes focused and distracted. However, partial indicators have significant impacts on driving status, so finding out indicators that were significantly different between focused and distracted was necessary. All the indicators were tested by Kolmogorov-Smirnov (K-S) tests and were indicated to be normally distributed. Feature reduction was performed using Student's *t*-test and correlation coefficients were calculated among fifteen

Table 1
Input and Output Variables Description.

Input/Output	Variables	Description	Units
Input	Speed	Vehicle speed sampled from the vehicle network	m/s
	Longitudinal Acceleration	Vehicle acceleration in the longitudinal direction versus time	m/s^2
	Lateral Acceleration	Vehicle acceleration in the lateral direction versus time	m/s^2
	Lane Offset	Distance to the left or right of the center of the lane based on machine vision	cm
	Steering Wheel Rate	Angular velocity of steering wheel sampled from the vehicle network	degree/s
Output	1/0	Distraction/ No Distraction	N/A

indicators, as shown in Section 4.1 Table 2, turning out that only the smoothened steering wheel rate and its standard deviation were not significant, and were removed from further analysis. The rest of the 13 indicators were not significantly correlated with each other based on Pearson correlation results and were kept as the inputs of the LSTM model. The data was divided into small segment of 10-second length with 100 data points. For those segments less than 100 data points, padding methods were adopted to make all the cutting periods to be the same length by filling in the blank at the end of the sequence (Chrysostomou et al., 2011). The final dataset had 3077 10-second periods, with 1939 periods being distracted samples and 1138 periods being focused driving samples. And all data was normalized to obtain the same order of magnitude.

3.3. LSTM model development

LSTM networks are widely used in pattern recognition applications as a supervised algorithm (Yang et al., 2016; Chen et al., 2019). It is good at time-series data processing because of its capability for memory, and the output of the current cell depends on the current input and the previous cell's output (Hochreiter and Schmidhuber, 1997). As shown in Fig. 3, c <t − 1> is the previous state, a <t − 1> is the previous activity, x <t> is the input of the current cell, y<t> is the output of the current cell, c <t> is the current state, and a <t> is the current activity. Data goes through gates to be filtered and integrated, determining the information in the memory cells. LSTM selectively discards redundant information and only memorizes the useful information, making it highly efficient.

A bidirectional LSTM layer was introduced to our distraction detection algorithm, which obtained the context information in both past and future directions. A bidirectional LSTM layer consists of two separate recurrent hidden layers that can scan the input information in opposite directions and connect to the same output layer (Schuster and Paliwal, 1997). According to previous studies, the bidirectional LSTM network achieved higher accuracy in sequence labeling tasks because of its unique network structure (Graves and Schmidhuber, 2005; Wollmer et al., 2010), and was proved to be more accurate in predicting drivers' actions than unidirectional LSTM model (Olabiyi et al., 2017).

The attention mechanism was introduced in this LSTM network, which helped the model to assign different weights to inputs, extract critical information, and enable the model's ability to make accurate judgments without incurring a heavy load of calculation and storage. This study adopted attention mechanism proposed by Bahdanau et al. (2014), utilizing a context vector to align the source and target inputs. Attention mechanisms are widely used in object detection and document classification (Yang et al., 2016; Chen et al., 2019), as it reduces of training time and error. Besides, as the gap between training and testing error was reduced, the overfitting problem was solved to some extent (Kouchak and Gaffar, 2020). A Self-attention network was used in the developed model (Vaswani et al., 2017). The structure of an attention layer was shown in Fig. 4.

Firstly, the score was calculated considering the output of the previous time step and the current input **Eq. (1)**. Secondly, the alignment score, which is the weight of each step on the output, was calculated using "SoftMax" function **Eq. (2)**. Finally, the context vector c_t was calculated by the sum of hidden states of the input sequence and weighted by alignment scores **Eq. (3)**.

$$e_{ti} = a(s_{t-1}, h_i) \tag{1}$$

$$\alpha_{ti} = \frac{\exp(e_{ti})}{\sum_{j=1}^{T_x} \exp(e_{tj})} \tag{2}$$

$$c_t = \sum_{j=1}^{T_x} \alpha_{ti} h_i \tag{3}$$

The model structure in this study was based on bidirectional LSTM and attention layer as shown in Fig. 5. The first layer had 13 indicators from data processing, which were inputs to the Bidirectional LSTM layer with 36 memory blocks or neurons. An attention

Table 2
Results of Student's *t*-test.

Variables	Indicators	P-Value	Confidence Interval	
Speed	Mean	0.000	0.3008	0.4269
	SD	0.000	−1.4254	−1.3993
	PE	0.000	0.0004	0.0005
Longitudinal Acceleration	Mean	0.000	0.0104	0.0120
	SD	0.000	−0.0079	−0.0075
	PE	0.000	0.0001	0.0002
Lateral Acceleration,	Mean	0.002	0.0002	0.0008
	SD	0.000	0.0016	0.0018
	PE	0.000	0.0002	0.0002
Lane Offset	Mean	0.000	46.2397	84.4140
	SD	0.000	−91.3613	−66.9334
	PE	0.000	11.2970	12.7113
Steering Wheel Rate	Mean	0.206	−0.6731	0.1453
	SD	0.204	−0.1021	0.4782
	PE	0.000	0.0417	0.0454

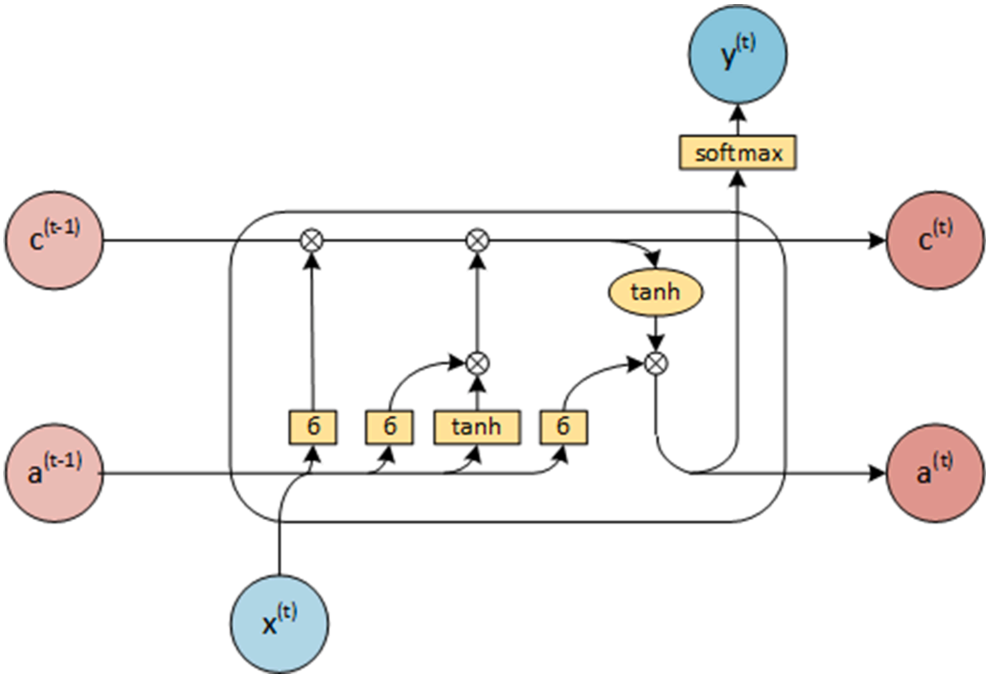


Fig. 3. The workflow of LSTM cell.

layer was added on the top of the LSTM layer with 48 memory neurons. Two dense layers were adopted to generate the output layer. The scores were normalized via "Sigmoid" function to calculate the final weight of each input sample (Han and Moraga, 1995). This model was expected to provide supreme accuracy compared to other machine learning methods.

4. Results

4.1. Model configuration and training

The dataset for model development has a total of 3077 10-second segments, with each segment containing 100 data points. There are 1939 positive samples (distracted driving) and 1138 negative samples (focused driving). A 5-fold cross validation method was used to develop models. The stratified k-fold strategy was used to ensure that the proportion of samples in each category in the training set and test set were the same as the original data set.

Adaptive Moment Estimation (Adam) Optimizer was utilized because of its outstanding performance and little requirement of tuning (Kingma and Ba, 2014). The initial learning rate used in this study was 0.01. The batch size was 32. The total training epoch number is 100. The training platform was a Windows PC with CPU of AMD Ryzen 5 3600 and GPU of Nvidia RTX 2060. The deep learning framework used in this study was TensorFlow.

4.2. Inputs selection

The results of Student's *t*-test to identify significant indicators were shown in Table 2. According to Table 2, there is no significant difference of mean value and standard deviation in steering wheel rate between distracted driving and focused driving ($p = 0.206$, $p = 0.204$). Except for these two indicators, others were significant. It turned out that speed, longitudinal acceleration, lateral acceleration, lane offset, and their standard deviation and PE, along with steering wheel rate PE, were effective to be the input of distracted driving detection models.

4.3. Ablation test of LSTM models

To identify the effectiveness of the attention layer and bidirectional LSTM layer, the ablation tests were conducted for comparison (Newell, 1975). First, a LSTM model was developed as the basic model which contains a LSTM layer and two dense layers. Second, the LSTM model with attention mechanism was built to evaluate the function of attention mechanism. This model added an attention layer before the LSTM layer. Third, the LSTM model with a bidirectional layer was built to evaluate the function of bidirectional LSTM layer. This model used a bidirectional LSTM layer. Fourth, the bidirectional LSTM (Bi-LSTM) model with attention mechanism was expected to be the best model.

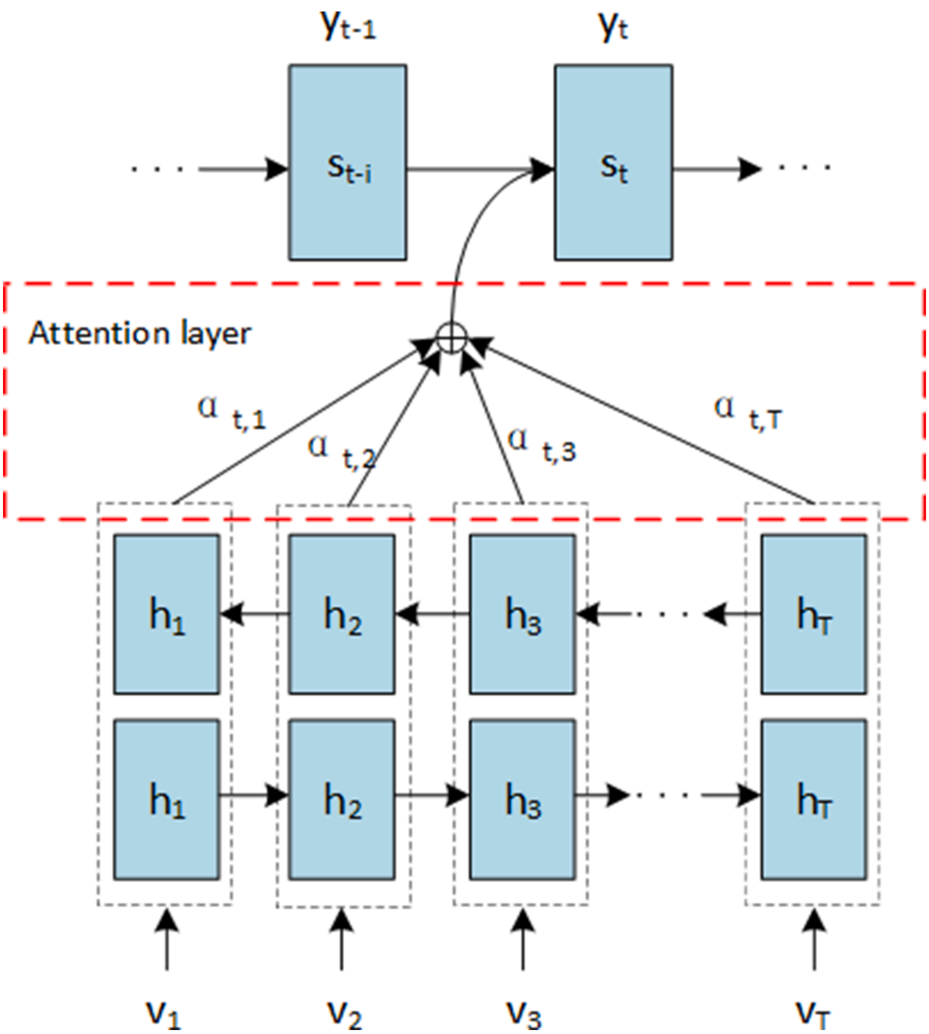


Fig. 4. Attention layer structure.

A confusion matrix was utilized to measure the performance of detection models. The rows of confusion matrix represent the instances in a predicted class while columns represent the instances in an actual class. According to the confusion matrix, the accuracy, precision, recall, and F1 score were further computed. Precision was defined as the ratio of correct positive predictions to the total predicted positives. Recall was defined as the ratio of correct positive predictions to the total positives. F1 score was the harmonic mean of precision and recall, ranging from zero to one with one being the best. The Receiver Operating Characteristic curve (ROC) was also used to present the classification performance, of which the horizontal axis indicates the false positive rate, and the vertical axis indicates the true positive rate. The area under the ROC curve (AUC) was computed to evaluate the capability of distinguishing between classes.

As shown in Table 3, the Bi-LSTM with attention mechanism achieved the highest accuracy, precision rate, F1, and AUC score. Although the recall rate was a little lower, the overall performance was still supreme among the four LSTM models. Besides, the results indicated that the attention mechanism increased the model performance, and the bidirectional LSTM worked better than the regular LSTM model, especially when the attention mechanism was included in models. Even though the running time of the Bi-LSTM model with attention mechanism was relatively longer, the additional time was negligible, and this model is recommended detect drivers' phone use distraction status.

4.4. Model comparison

The comparison among the Bi-LSTM model with attention mechanism, recurrent neural network (RNN), support vector machine (SVM), k-nearest neighbor (KNN), and adaptive boosting (AdaBoost) was conducted. These models were chosen based on their superior predictive capabilities from existing study findings (Atiquzzaman et al., 2018, Tango and Botta, 2013).

RNN is an artificial neural network that retains the sequence information in the hidden state. However, regular RNN is not able to

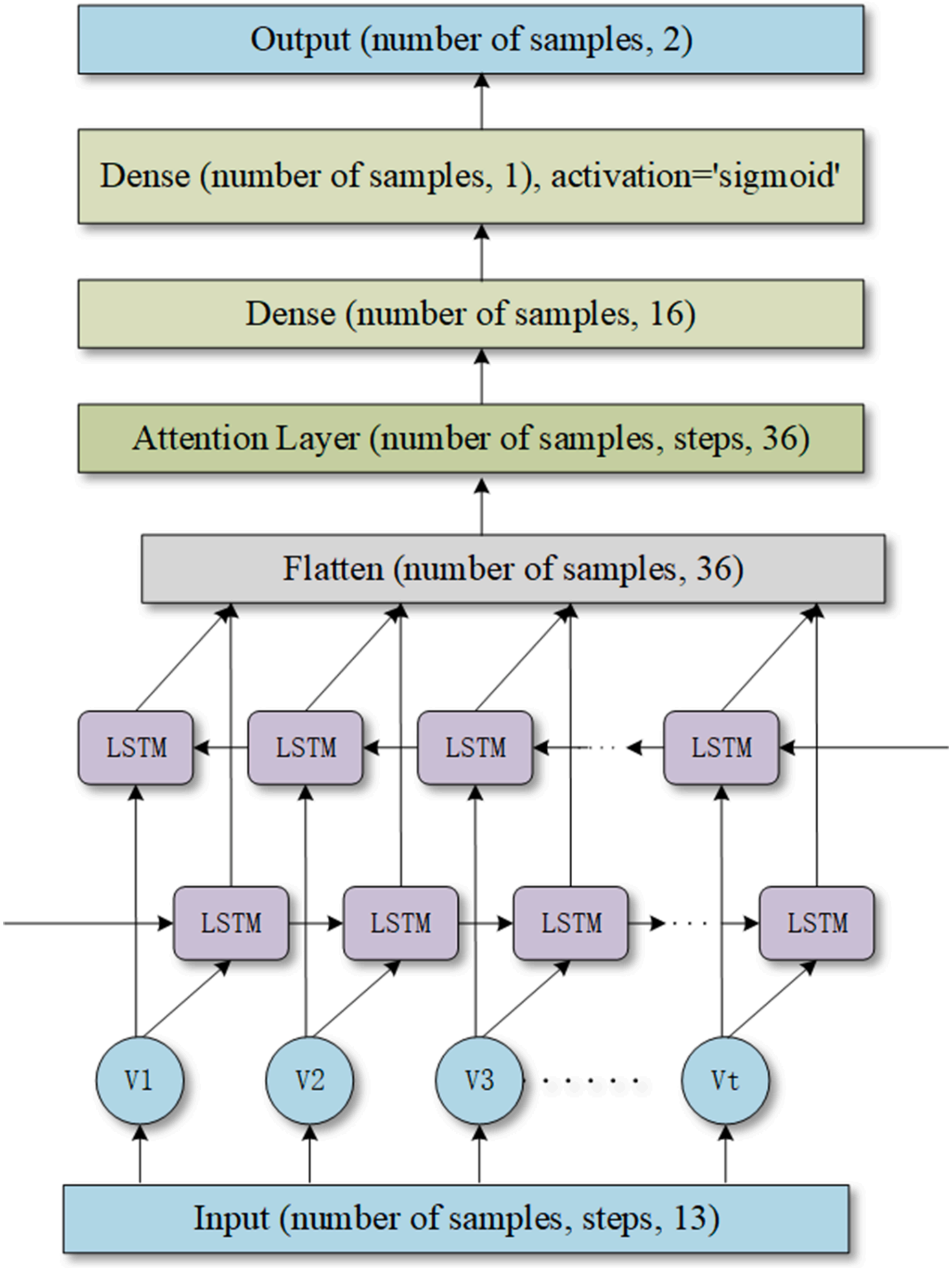


Fig. 5. The bidirectional LSTM network with Attention Mechanism.

Table 3
Ablation test results of four developed LSTM models.

Models	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	AUC (%)	Running time (ms)
LSTM	89.276	91.715	91.904	91.608	97.093	18.25
LSTM with Attention Mechanism	88.465	89.425	93.554	91.211	96.975	18.37
Bi-LSTM	89.438	93.506	91.502	91.829	96.822	19.54
Bi-LSTM with Attention Mechanism	91.226	94.595	91.437	92.924	97.396	19.34

handle long-range context due to the vanishing gradient problem (Hochreiter et al., 2001). A model with a RNN layer was built to evaluate the capability of dealing with long-range contextual information of the proposed LSTM model in this study.

SVM is based on structural risk minimization originated from statistical learning theory (Byun and Lee, 2003). It constructs a hyperplane in high-dimensional space to achieve the maximum separation between different classes. In this study, the radial basis function (RBF) was introduced as the kernel function due to its robustness compared to linear or polynomial SVMs.

KNN is a non-parametric model, which categorizes the target object based on the classes of k nearest training samples in feature space (Dudani, 1976). The most common classification among k nearest neighbors determined the category assigned to the object and k was selected as five in this study.

AdaBoost is a widely used boosting algorithm for pattern classification proposed by Freund and Schapire (Freund and Schapire, 1995). The key point of AdaBoost is improving weak learners and creates an aggregated model to make the prediction more accurate using the training data. The newly created classifiers emphasize more on the observations that were misclassified by the classifiers created before (Miyaji et al., 2008).

The same dataset was used for training and comparing their performance on phone use distraction detection. The comparison results of aforementioned models are shown in Table 4. Their ROC curves are shown in Fig. 6.

The comparison results are shown in Table 4. Contributed by the outstanding time-series data processing ability, the Bi-LSTM model with attention mechanism had better performance than other models. The proposed Bi-LSTM model provided an outstanding classification accuracy of 91.2%, and high precision and recall rate of 94.6% and 91.4%, with F1 score reached 92.9%. As for the AUC score, the Bi-LSTM model reached 97.4% which was the best performance on testing data among these models, and Fig. 6 clearly visualized the AUC. The RNN, SVM, KNN, and AdaBoost models were less accurate in predicting driver distraction, of which accuracy rates were 83.0%, 87.2%, 84.7%, and 84.8% respectively. Other evaluation metrics of RNN, SVM, KNN, and AdaBoost were not as good as LSTM either. The running times were very different among the approaches, which was due to the construction properties of the models and the comparison was not meaningful. The comparison results lead to the conclusion that the Bi-LSTM model with attention mechanism is reliable for driver distraction detection and has superior accuracy and stability than machine learning methods.

5. Conclusions and discussions

This study proposed a Bi-LSTM with attention mechanism to detect drivers' distraction status and evaluated this model using naturalistic driving data from SH-NDS. Five variables of driving performance, (speed, longitudinal acceleration, lateral acceleration, lane offset, and steering wheel rate) were determined to be highly correlated to distracted driving using phones. The phone use events in this study included browsing, texting, phone calls, changing phone position, and combination. They represented cognitive distraction, visual distraction, physical distraction, and auditory distraction, which were representative of distractions in real life. Data were extracted from video recorded during naturalistic driving, and time periods were divided into focused and distracted. With erroneous data removed, the valid dataset included 3077 10-second data segments, with 1138 segments being focused and 1939 segments being distracted. The time unit was 0.1 s, with a one-second moving window processing time-series data. The mean, standard deviation and PE of the five driving performance variables were computed, and all of them, except steering wheel rate and its standard deviation, were significantly related to distraction. Steering wheel rate and its standard deviation were removed from inputs.

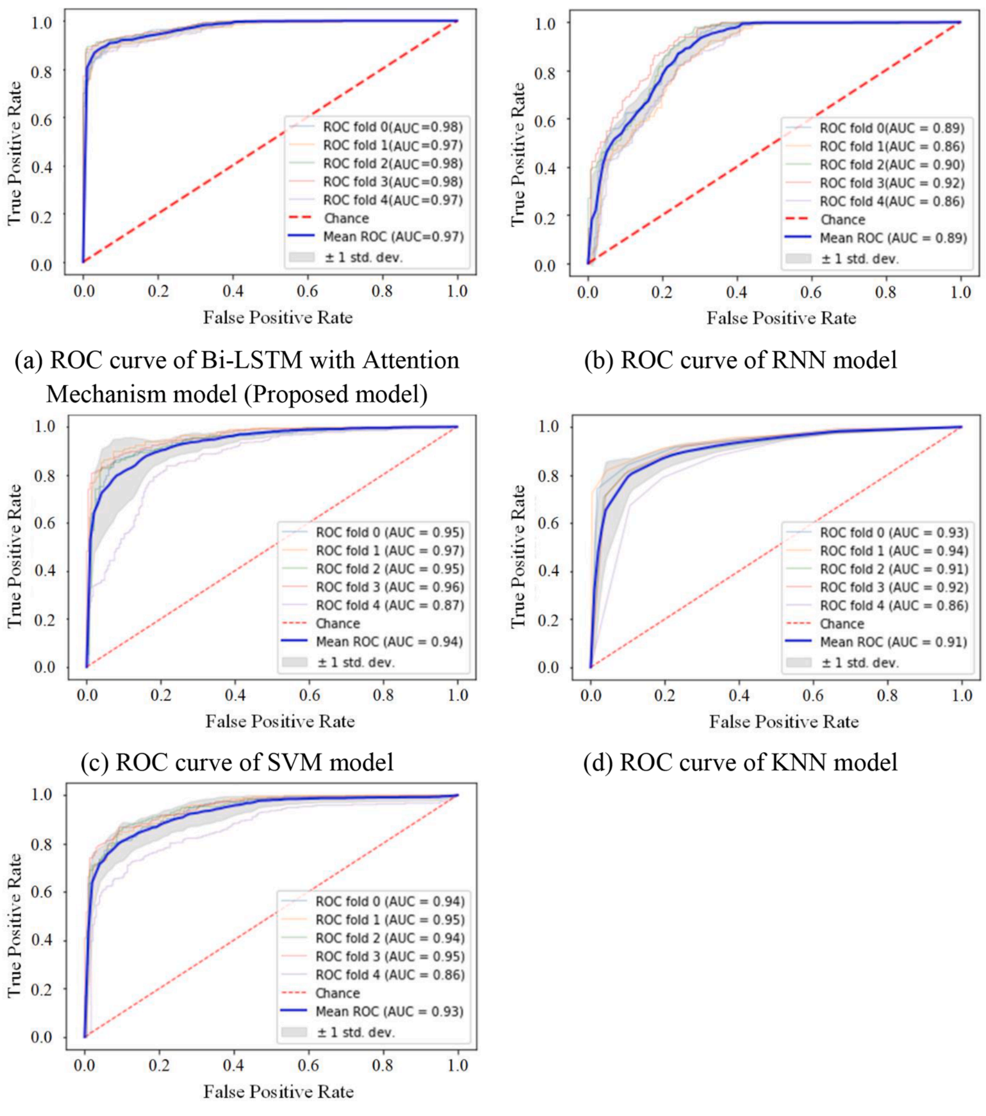
The experiment results show that the Bi-LSTM model with attention mechanism has a promising result in driver phone use distraction detection with higher accuracy than some commonly used machine learning methods. Contributed by the attention mechanism, the Bi-LSTM achieved 91.2% accuracy of distraction detection, which was 8.2%, 4.0%, 6.5%, and 6.4% higher than the RNN, SVM, KNN, and AdaBoost models, respectively. The overfitting problem was resolved because of the attention mechanism. And these indicated that LSTM networks were superior in time-series data processing, which was consistent with the literature (Wollmer et al., 2011; Kouchak and Gaffar, 2020). Attention mechanism was effective in improving the performance of distraction detection algorithms combined with LSTM networks. Besides accuracy, the Bi-LSTM model with attention mechanism in this paper took naturalistic driving and the context of driving time-series data into consideration, which was more realistic than previous studies utilizing driving simulator data only, or solely relying on face recognition.

The results of distraction detection indicated that the driving patterns differed between distracted and focused driving, which was consistent with the literature that driver distraction had significant impacts on vehicle dynamics (Wollmer et al., 2011; Atiquzzaman et al., 2018; Kouchak and Gaffar, 2020). The results of this study implied that speed, longitudinal acceleration, lateral acceleration, lane offset, and their standard deviation and predicted error, along with steering wheel rate predicted error, were significant indicators in distraction detection, which is easily accessible and would be able to replace the effort of intrusive devices like facial video cameras, which could make drivers more relaxed and drive more naturally. This model could be adopted by vehicle manufactures and applied to advanced driving assistant systems, and provide alerts to the driver when distraction was detected using real-time vehicle dynamic data.

This study could be further expanded for safety enhancement. For example, more types of distraction such as visual/physical distraction and cognitive distraction could be examined, as eating and drinking and talking with passengers are as distractive as phone usage. Although the distraction events shared similar features, driving performance patterns may be different for different types of distractions. With rapidly increasing computer power, there may be other deep learning models that are able to process time-series data more efficiently to be investigated.

Table 4
Results of proposed model and other machine learning models.

Models	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	AUC (%)	Running time (ms)
Bi-LSTM with Attention Mechanism	91.226	94.595	91.437	92.924	97.396	19.34
RNN	83.035	83.783	90.922	87.045	88.728	23.74
SVM	87.226	86.196	86.521	86.334	93.594	0.14
KNN	84.692	84.007	82.750	83.262	91.201	0.59
AdaBoost	84.757	83.724	83.567	83.636	92.737	18.94



(a) ROC curve of Bi-LSTM with Attention Mechanism model (Proposed model)

(b) ROC curve of RNN model

(c) ROC curve of SVM model

(d) ROC curve of KNN model

(e) ROC curve of AdaBoost model

Fig. 6. Receiver Operating Characteristic curves of phone use distraction detection models.

CRediT authorship contribution statement

Xuesong Wang: Methodology, Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Rongjiao Xu:** Methodology, Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Siyang Zhang:** Methodology, Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Yifan Zhuang:** Methodology, Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review &

editing. **Yinhai Wang:** Methodology, Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

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.

Acknowledgment

This study was jointly sponsored by the Chinese National Science Foundation(51878498), theScience and Technology Commission of Shanghai Municipality(18DZ1200200), the Clinical Research Center for Mental Disorders, Shanghai Pudong New Area Mental Health Center, School of Medicine, Tongji University (PDJWM-202103), and the Outstanding Clinical Discipline Project of Shanghai Pudong (PWYgy2021-02).

References

Atiquzzaman, M., Qi, Y., Fries, R., 2018. Real-time detection of drivers' texting and eating behavior based on vehicle dynamics. Transp. Res. Part F Traffic Psychol. Behav. 58, 594–604. https://doi.org/10.1016/j.trf.2018.06.027.

Bahdanau, D., Cho, K., Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv: 1409.0473.

Byun, H., Lee, S.-W., 2003. A survey on pattern recognition applications of support vector machines. Intern. J. Pattern Recognit. Artif. Intell. 17 (03), 459–486.

Chen, S., Zhang, S., Shang, J., Chen, B., Zheng, N., 2019. Brain-inspired cognitive model with attention for self-driving cars. IEEE Trans Cogn Dev Syst. 11 (1), 13–25.

Choudhary, P., Velaga, N.R., 2017a. Mobile phone use during driving: effects on speed and effectiveness of driver compensatory behaviour. Accid. Anal. Prev. 106, 370–378.

Choudhary, P., Velaga, N.R., 2017b. Analysis of vehicle-based lateral performance measures during distracted driving due to phone use. Transp. Res. Part F Traffic Psychol. Behav. 44, 120–133.

Chrysostomou, C., Seker, H., Aydin, N., 2011. Effects of windowing and zero-padding on complex resonant recognition model for protein sequence analysis. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4955–4958.

Dudani, S.A., 1976. The distance-weighted k-nearest-neighbor rule. IEEE Trans. Syst. Man Cybern. Syst. SMC-6 (4), 325–327.

Ersal, T., Fuller, H.J.A., Tsimhoni, O., Stein, J.L., Fathy, H.K., 2010. Model-based analysis and classification of driver distraction under secondary tasks. IEEE Trans. Intell. Transp. Syst. 11 (3), 692–701.

Fitch, G.M., Soccolich, S.A., Guo, F., McClafferty, J., Fang, Y., Olson, R.L., et al., 2013. The impact of hand-held and hands-free cell phone use on driving performance and safety-critical event risk (No. DOT HS 811 757).

Freund, Y., Schapire, R.E., 1995. A decision-theoretic generalization of on-line learning and an application to boosting. In: European Conference on Computational Learning Theory. Springer, Berlin, Heidelberg, pp. 23–37.

Graves, A., Schmidhuber, J., 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Netw. 18 (5-6), 602–610.

Haddad, R.A., Akansu, A.N., 1991. A class of fast Gaussian binomial filters for speech and image processing. IEEE Trans. Signal Process. 39 (3), 723–727.

Han, J., Moraga, C., 1995. The influence of the sigmoid function parameters on the speed of backpropagation learning. In: International Workshop on Artificial Neural Networks. Springer, Berlin, Heidelberg, pp. 195–201.

Hochreiter, S., Bengio, Y., Frasconi, P., Schmidhuber, J., 2001. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

Im, S., Lee, C., Yang, S., Kim, J., You, B., 2014. Driver distraction detection by in-vehicle signal processing. In: 2014 IEEE Symposium on Computational Intelligence in Vehicles and Transportation Systems (CIVTS), pp. 64–68.

Jeong, H., Liu, Y., 2019. Effects of non-driving-related-task modality and road geometry on eye movements, lane-keeping performance, and workload while driving. Transp. Res. Part F Traffic Psychol. Behav. 60, 157–171.

Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kircher, K., Ahlstrom, C., 2010. Predicting visual distraction using driving performance data. In: Annals of Advances in Automotive Medicine/Annual Scientific Conference. Association for the Advancement of Automotive Medicine, vol. 54, pp. 333.

Klauer, S.G., Guo, F., Simons-Morton, B.G., Ouimet, M.C., Lee, S.E., Dingus, T.A., 2014. Distracted driving and risk of road crashes among novice and experienced drivers. N. Engl. J. Med. 370 (1), 54–59.

Kouchak, S.M., Gaffar, A., 2019. Driver distraction detection using deep neural network. In: International Conference on Machine Learning, Optimization, and Data Science, pp. 13–23.

Kouchak, S.M., Gaffar, A., 2021. Detecting driver behavior using stacked long short term memory network with attention layer. IEEE Trans. Intell. Transp. Syst. 22 (6), 3420–3429.

Li, Z., Bao, S., Kolmanovsky, I.V., Yin, X., 2018. Visual-manual distraction detection using driving performance indicators with naturalistic driving data. IEEE Trans. Intell. Transp. Syst. 19 (8), 2528–2535.

Liang, Y., Lee, J.D., 2014. A hybrid Bayesian Network approach to detect driver cognitive distraction. Transp. Res. Part C Emerg. Technol. 38, 146–155.

Liang, Y., Reyes, M.L., Lee, J.D., 2007. Real-time detection of driver cognitive distraction using support vector machines. IEEE Trans. Intell. Transp. Syst. 8 (2), 340–350.

Liao, Y., Li, S.E., Li, G., Wang, W., Cheng, B., Chen, F., 2016. Detection of driver cognitive distraction: an SVM based real-time algorithm and its comparison study in typical driving scenarios. In: 2016 IEEE Intelligent Vehicles Symposium (IV), pp. 394–399.

Luo, Y., Gao, Y., You, Z., Shou, R., 2019. A method for driving distraction detection based on ols-rbf neural network model. In CICTP 2019, 289–301.

Miyaji, M., Danno, M., Kawanaka, H., Oguri, K., 2008. Driver's cognitive distraction detection using AdaBoost on pattern recognition basis. In: In 2008 IEEE International Conference on Vehicular Electronics and Safety, pp. 51–56.

Naqvi, R.A., Arsalan, M., Batchuluun, G., Yoon, H.S., Park, K.R., 2018. Deep learning-based gaze detection system for automobile drivers using a NIR camera. Sensor 18 (2), 456. https://doi.org/10.3390/s18020456.

National Center for Statistics and Analysis, 2019. Distracted driving in fatal crashes, 2017: Traffic Safety Facts Research Note. Report No. DOT HS 812 700).

National Highway Traffic Safety Administration, 2017. Distracted driving 2015: traffic safety facts research note (rep. No. DOT HS 812 381).

Newell, A., 1975. A tutorial on speech understanding systems. Speech Recogn. 4–54.

Olabiyi, O., Martinson, E., Chintalapudi, V., Guo. R., 2017. Driver action prediction using deep (bidirectional) recurrent neural network. arXiv preprint arXiv: 1706.02257.

Oviedo-trespalacios, O., Haque, M., King, M., Washington, S., 2016. Understanding the impacts of mobile phone distraction on driving performance: a systematic review. Transp. Res. Part C Emerg. Technol. 72, 360–380. https://doi.org/10.1016/j.trc.2016.10.006.

Ryu, D.W., Jeong, H.B., Lee, S.H., Lee, W.S., Yang, J.H., 2015. Development of driver-state estimation algorithm based on Hybrid Bayesian Network. In: 2015 IEEE Intelligent Vehicles Symposium (IV), pp. 1282–1286.

Schuster, M., Paliwal, K.K., 1997. Bidirectional recurrent neural networks. IEEE Trans. Signal Process. 45 (11), 2673–2681.

Sun, W., Aguirre, M., Jin, J.J., Feng, F., Rajab, S., Saigusa, S., Dsa, J., Bao, S., 2021. Online distraction detection for naturalistic driving dataset using kinematic motion models and a multiple model algorithm. Transp. Res. Part C Emerg. Technol. 130, 103317. https://doi.org/10.1016/j.trc.2021.103317.

Tango, F., Botta, M., 2013. Real-time detection system of driver distraction using machine learning. IEEE Trans. Intell. Transp. Syst. 14 (2), 894–905.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inform. Process. Syst. 30, 5998–6008.

Wang, X., Xing, Y., Luo, L., Yu, R., 2018. Evaluating the effectiveness of Behavior-Based Safety education methods for commercial vehicle drivers. Accid. Anal. Prev. 117, 114–120.

Wang, X., Xu, C., 2016. Driver drowsiness detection based on non-intrusive metrics considering individual specifics. Accid. Anal. Prev. 95, 350–357.

Wang, X., Xu, R., Asmelash, A., Xing, Y., Lee, C., 2020. Characteristics of driver cell phone use and their influence on driving performance: a naturalistic driving study. Accid. Anal. Prev. 148, 105845. https://doi.org/10.1016/j.aap.2020.105845.

Wollmer, M., Blaschke, C., Schindl, T., Schuller, B., Farber, B., Mayer, S., Trefflich, B., 2011. Online driver distraction detection using long short-term memory. IEEE Trans. Intell. Transp. Syst. 12 (2), 574–582.

Wollmer, M., Schuller, B., Eyben, F., Rigoll, G., 2010. Combining long short-term memory and dynamic bayesian networks for incremental emotion-sensitive artificial listening. IEEE J. Sel. Top. Signal Process. 4 (5), 867–881.

Yan, W., Xiang, W., Wong, S.C., Yan, X., Li, Y.C., Hao, W., 2018. Effects of hands-free cellular phone conversational cognitive tasks on driving stability based on driving simulation experiment. Transp. Res. Part F Traffic Psychol. Behav. 58, 264–281.

Yang, M., Wang, X., Quddus, M., 2019. Examining lane change gap acceptance, duration and impact using naturalistic driving data. Transp. Res. Part C Emerg. Technol. 104, 317–331.

Yang, Y., Sun, H., Liu, T., Huang, G. B., Sourina, O., 2015. Driver workload detection in on-road driving environment using machine learning. In: Proceedings of ELM-2014, Springer, Cham., vol. 2, pp. 389–398.

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., Hovy, E., 2016. Hierarchical attention networks for document classification. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1480–1489.

Yannis, G., Laiou, A., Papantoniou, P., Christoforou, C., 2014. Impact of texting on young drivers' behavior and safety on urban and rural roads through a simulation experiment. J. Safety Res. 49, 25-e1.

Ye, M., Osman, O.A., Ishak, S., Hashemi, B., 2017. Detection of driver engagement in secondary tasks from observed naturalistic driving behavior. Accid. Anal. Prev. 106, 385–391. https://doi.org/10.1016/j.aap.2017.07.010.

Zhang, X., Wang, X., Yang, X., Xu, C., Zhu, X., Wei, J., 2020. Driver drowsiness detection using mixed-effect ordered logit model considering time cumulative effect. Anal. Methods Accid. Res. 26, 100114. https://doi.org/10.1016/j.amar.2020.100114.

Zhang, Y., Owechko, Y., Zhang, J., 2004. Driver cognitive workload estimation: a data-driven perspective. In: Proceedings. 7th Int. IEEE Conf. Intell. Transp. Syst., pp. 642–647.

Zhu, M., Wang, X., Tarko, A., Fang, S., 2018. Modeling car-following behavior on urban expressways in Shanghai: a naturalistic driving study. Transp. Res. Part C Emerg. Technol. 93, 425–445.