scientific reports

Check for updates

OPEN

Classifying metro drivers' cognitive distractions during manual operations using machine learning and random forest-recursive feature elimination

Haiyue Liu[1], Yue Zhou[2] & Chaozhe Jiang[1]✉

Metro drivers are more likely to trigger accidents if they suffer from cognitive distractions during manual driving. However, identifying metro drivers' cognitive distractions faces challenges as generally no obvious behavior can be found during the distractions. To address the challenge, this paper identifies metro drivers' cognitive distractions based on Electrocardiogram (ECG) signals collected by wearable devices in simulated driving experiments. The ECG signals are processed to generate ultra-short-term heart rate and heart rate variability (HR-HRV) features. The HR-HRV features are extracted by 30-s and 60-s time-windows in driving phase, and 25-s time-windows in parking phase, respectively. Machine learning approaches are developed to identify distractions (binary) and distinguish the degrees of distractions (multi-class). The optimal input features are determined by a random forest and recursive feature elimination (RF-RFE) algorithm. Results show that the DT with only one HR-HRV feature extracted from 30-s time-windows and XGBoost with 20 h-HRV features extracted from 60-s time-windows are optimal models for binary and multi-class classification for distractions during driving phase, respectively. The features including NN20, pNN20, SD1/SD2, Max-HR, Min-HR, and MEDNN are the most critical HR-HRV features associated with distractions. Cognitive distractions in parking phase are difficult to be detected using HR-HRV features.

Keywords Metro drivers, Cognitive distractions, Recursive feature elimination, Machine learning, HR-HRV features, Shapley additive explanations

Driving distraction is regarded as one of the riskiest physiological states for drivers. It impairs drivers' reaction time and ability to control the vehicle, which significantly increases the likelihood of accidents. In addition to the impact on road safey[1], this detrimental physiological state also compromises the safety of railway systems. Past studies reveal that rail drivers who suffer from distractions are more likely to trigger incidents such as overruns or collisions, especially metro drivers[2]. For instance, on December 15, 2023, a distracted metro driver in Beijing failed to respond timely to a fault-warning signal, resulting in a rear-end collision that caused over 500 injuries[3]. The severe consequences of metro accidents triggered by driving distractions make it crucial to prevent metro drivers from becoming distracted. However, to our knowledge, there are quite limited studies that focus on this issue among metro drivers.

Driving distractions are typically categorized into visual, auditory, manual, and cognitive distractions, corresponding to the engagement of sight, hearing, hands, and mental focus away from driving tasks[4]. Among these types, cognitive distractions, such as daydreaming, mind wandering, and decision-making challenges, are particularly difficult to detect, as no obvious behavior (e.g., talking, texting, or drinking) or emotion changes appear on the drivers during cognitive distractions. In manual metro operations, drivers must repeatedly perform monotonous tasks in enclosed environments, e.g., underground tunnels. Hence, cognitive distractions are more commonly found among metro drivers than other driving cohorts. To reduce driving distractions and ensure operational safety, metro companies have implemented various countermeasures and regulations to combat visual, auditory, and manual distractions. Moreover, unlike motor drivers, metro is running in a fixed

[1]School of Transportation and Logistics, Southwest Jiaotong University, 610097 Chengdu, People's Republic of China. [2]Civil Aviation Flight University of China, Flight Technology College, 618300 Guanghan, People's Republic of China. ✉email: jiangchaozhe@swjtu.edu.cn

track and have small changes in speeds during cruises. As a result, previous studies that use driver's behavior or vehicle-based data (e.g., speed, acceleration, lane offset, or steering) to detect drivers' distraction[5,6] may not be effective for detecting metro drivers' cognitive distractions.

To solve the challenges, studies recommend exploiting physiological signals for detecting cognitive distractions. Among the physiological signals, Electroencephalogram (EEG)[7] and electrocardiogram (ECG)[7] are highly related to cognitive distraction. Therefore, analysts can develop methods to identify cognitive distractions using EEG or ECG signals. For instance, Wu, et al.[8] used GRU-EEGNet model based on EEG signals and achieved 78% accuracy. Taherisadr, et al.[9] developed a deep learning model and correctly classified 95.51% of drivers' distractions using ECG signals. While EEG data have been confirmed as closely associated with driving distraction, ECG data are also favored by analysts. Compared with other physiological signals, ECG data can be measured by lightweight and wearable devices that offer a high signal-to-noise ratio and minimal invasiveness in driving experiments, making ECG more popular and feasible for identifying cognitive distractions[10,11]. This is particularly important because of the cockpit of a metro train is a small and confined space. Conventional EEG, Functional Near-Infrared Spectroscopy (fNIRS), and Magnetic Resonance Imaging (MRI) and other physiological acquisition devices are either relatively larger or require sensitive contact with the driver's skin. These devices may affect the driver's manipulations and provide unreliable physiological responses during the measurement. ECG signals may also provide solutions for detecting cognitive distraction of metro drivers, even similar works are not found in past studies.

To address the gaps above, this paper aims to identify cognitive distractions in metro drivers and classify the degrees of these distractions using HR-HRV features during manual operations. The HR-HRV features are extracted from ECG signals collected during simulated driving experiments. The physiological features of metro drivers under cognitive distractions are compared between driving and parking phase. Additionally, the study employs RF-RFE to select the most critical HR-HRV features, which are then inputted into machine learning (ML) models to identify the distraction and classify the distraction degrees. The contributions of this study to the literature are: (1) it demonstrates the feasibility of using HR-HRV features to identify cognitive distractions of metro drivers during the manual driving; (2) it compares the model performance in identifying cognitive distractions between driving and parking phase. (3) it uncovers the critical HR-HRV features related to metro drivers' cognitive distractions and distractions in different degrees and visually explain the associations using SHAP technique.

Literature review

Analysis for HR-HRV features

The detection of physiological states of drivers using ECG usually computes the raw signals into frequency-domain, time-domain, or non-linear features of the beat-to-beat variations in heart rates (R-R intervals)[12], denoted heart rate (HR) and heart rate variability (HRV). The HR-HRV features provide more detailed characteristics of heartbeats compared to raw ECG signals, thereby enhancing the identification of cognitive distractions[13,14]. For instance, past studies indicate that distractions are significantly associated with increased MHR[14], decreased LF/HF power ratio[15] and HF[16,17]. Studies of Arutyunova, et al.[18] showed that PermEn, LF, TP, and LF/HF tend to increase as drivers become mentally distracted, while SDNN tend to decrease[19]. The feature of pNN50 is found to decrease with cyclists being distracted[17]. Wang, et al.[20] and Siennicka, et al.[21] found MaxHR, MHR, RMSSD, CVSD and pNN50 are significantly associated with distraction of participants.

The performance of identification depends on the time-windows used for extracting the HR-HRV features. Traditional medical time-window of HR-HRV extraction is recommended to be 5–10 min[22,23], while recent studies also shown that shorter HR-HRV time-windows can be considered to become an important direction of driver state analysis[24,25], especially for the analytic scenarios with a short time. Short-term HRV analysis is a convenient method for tracking dynamic changes of cardiac autonomic function within minutes[26,27], which are especially suitable for experiment that data collection time cannot be too long. For instance, Jiao, et al.[28] analyze the statistical differences of HR-HRV features extracted from different time-windows across the drivers' status, finding that the differences of features extracted from the 1-min time-windows show the most significances, while the features extracted from the 5-min time-windows have the least significances. Similar results are also shown in the study of Castaldo, et al.[24] who argue that ultra-short HRV features (1 to 3 min time-windows) are proved to be valid surrogates for mental stress investigation. It should be mentioned that, in our study, metro has a short travel time between stops (1–2 min), implying that long time-windows are not suitable. Therefore, selecting a proper length for the short time-window for HR-HRV extraction is needed for guaranteeing the validity of our study.

Feature selection for physiological signal analysis

Physiological responses of drivers are typically reflected by numerous physiological features. In order to lower the data noises, it is not advisable to include all features as input variables of ML models[29]. Therefore, applying feature selection methods before modeling can raise the effectiveness and validity of the model. Feature selection methods are generally categorized into three types: filter, wrapper, and embedded methods[30]. Filter methods, such as Chi-square tests[31], Pearson correlation coefficients[32] and Mutual information[33], are used to select appropriate features from the entire set before model training. Wrapper methods contain training classifiers (e.g., support vector machine[34], random forest) with different subsets of features and selecting the best subset by optimizing model performance. Embedded methods incorporate penalties within ML models to reduce the contributions of less important features[35].

Compared to filters and embedded methods, wrapper methods are more intuitive, flexible, and effective. Recursive feature elimination (RFE) is a well-known wrapper method widely used to address the drawbacks of traditional feature selection methods[36]. To be specific, RFE repeatedly evaluates the importance of features in

Reference	Research objects	Physiological signals	Features selection method	Machine learning models	Model accuracy
Chen, et al.[40]	Road users	Electrocardiogram (EKG), Galvanic skin response (GSR), EEG, Electrooculography (EOG)	Single filtering feature selection; Multiple filtering feature selection	SVM, Transfer joint matching (TJM), Adaptation regularization based transfer learning (ARTL)	94.30% from ARTL model with multiple filtering feature selection; 93.23% from ARTL model with singe filtering feature selection
Zuo, et al.[41]	Road users	EEG, ECG, Electromyogram (EMG)	ReliefF, Non-Negative matrix factorization (NMF), Mutual information, Neighborhood component analysis (NCA), Sequential forward selection (SFS)	RF, BiLSTM	The accuracy of models with ReliefF selection are 3% higher than model with other feature selections
Misra, et al.[42]	Road users	ECG, Photoplethysmography (PPG), Eye-Tracking	RFE	Naive bayes, KNN RF, DT, SVM	91.54% from RF model using RFE
Dehzangi, et al.[43]	Road users	GSR, Skin conductance (SC)	Relief, RF, Neighborhood component analysis	SVM, KNN, Ensemble bagged classifiers	93.5% from ensemble bagged classifier using RF
Papakostas, et al.[44]	Polits	Blood volume pulse, Skin conductance, Skin temperature, Respiration	DT	RF, SVM, KNN, CNN-LSTM	70% from CNN-LSTM using DT feature selection

Table 1. Comparative study of recent work for feature selection analysis.

Reference	Participants	Driving status	Features	Algorithm model	Model accuracy
Zhou, et al.[49]	Road users	Fatigue	HR-HRV	GPBoost	89.2% from the unique model
Strle, et al.[50]	Road users	Cognitive distraction	HRV, EDA, skin temperature, pupil dilation	LGBM, HGBC, XGB	63% from best model (LGBM) with four kinds of feature fusion
Huang, et al.[51]	Road users	Emotional state	EEG, EDA	XGBoost, CatBoost, LightGBM, KNN	94.88% from the unique model
Fasanmade[52]	Road users	Driving stress	HR-HRV	GBDT	90.92% from the unique model
Alreshidi, et al.[53]	Polit	Cognitive distraction	EEG	CNN	96% from the unique model

Table 2. Comparative study of recent work for SHAP analysis.

different subsets using a classifier and ultimately identifies the best feature subset. This iterative process allows the RFE algorithm to assess feature contributions to model outputs in various combinations of feature sets, which traditional feature selection methods may not achieve. Recently, the development of RFE manifests promise for predictions involving numerous features, such as identifying driving behaviors with a large number of physiological features[37]. While RFE provides an effective approach for feature selection, its performance heavily depends on the underlying classifier used. When applied with linear models, it may fail to capture complex non-linear relationships, which are common in physiological data like heart rate variability (HRV) signals. To address these limitations, Random Forest can be combined with Recursive Feature Elimination (RF-RFE) to better capture the non-linear relations[38], which is particularly effective for physiological data analysis. For instance, Gómez-Ramírez, et al.[39] indicated that Random Forest's built-in feature importance mechanism is more robust and reliable compared to that of linear models.

This study provides a comparative list including studies with different feature selection methods in analyzing drivers' physiological states (Table 1). Note that those studies mainly focus on road users or pilots, while no metro drivers are concerned.

Explanation of machine learning models

Machine learning models can achieve high accuracy; however, they are more difficult to interpret than simple approaches such as statistical regression models. The restriction of machine learning models may hinder the understanding of how driver response to cognitive distractions.

To provide interpretability, Friedman[45] introduced a tree ensemble machine learning algorithm named Gradient Boosting Decision Tree (GBDT) to provide feature importance and explain the impact of each factor on the results in terms of feature importance. Recently, Lundberg, et al.[46] introduce an explanation framework named SHapley Additive exPlanation (SHAP). Using the idea of the Shapley value represents the outcome as the sum of each feature contribution. SHAP values have proved to be consistent and can be plotted to illustrate the effect of different features on model outputs[47]. At present, there are few studies on visually interpretating the associations between driver distraction and their physiological features, while more studies apply SHAP in analyzing other aberrant physiological states, e.g., fatigue. For example, Zhou, et al.[48] used GPBoost and SHAP to predict driver fatigue in monotonous automated driving based on HR-HRV features. Table 2 lists some recent studies that visually interpret the associations between physiological signals and aberrant driving states using SHAP framework.

Fig. 1. The cockpit of driving simulator and real metro.

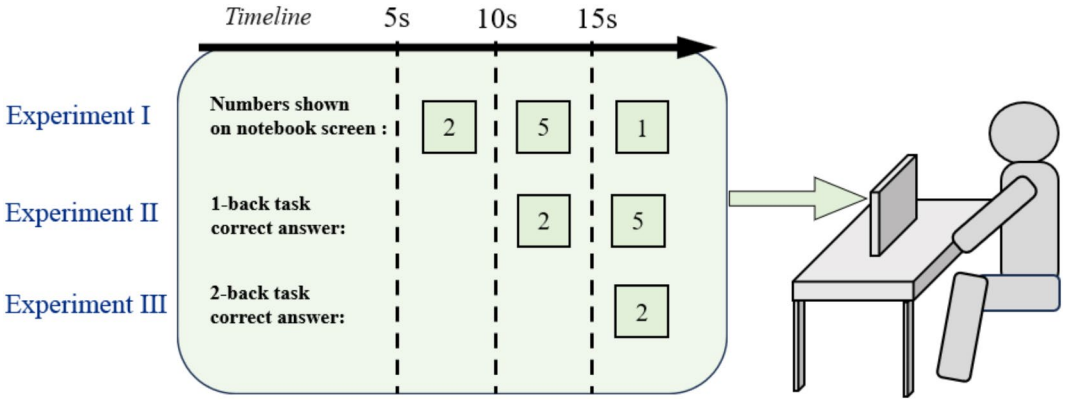


Fig. 2. The detailed of the designed N-backs tests.

Data collection

Participants and equipment

The analysis of metro drivers' cognitive distractions is based on several physiological experiments approved by the Medical Ethics Committee at Southwest Jiaotong University (SWJTU-2103-039). Thirty-two professional male drivers from Lanzhou Rail Transit Cooperation participated in these driving experiments. The average age and working experience of the drivers are 30 years (SD = 3.08) and 6 years (SD = 2.87), respectively. All participants were healthy and reported no discomfort before or during the experiments. Also, informed consents were obtained from all participants before the experiments.

The experiments were conducted using a metro simulator maintained by Lanzhou Rail Transit Cooperation. This simulator comprises a train cockpit with a 150-inch screen displaying real-time driving scenes. The simulator's layout replicates that of a real train cockpit (Fig. 1). The simulated driving scenes, including environments, objects, sounds, operating interface, infrasound, and stations, closely resemble those in real operations.

ECG signals of the metro drivers were collected using a Polar H10 device working at a frequency of 130 Hz. The Polar H10 is a medical-grade, lightweight, and wearable device that collects high-quality ECG signals. Due to its advantages, the Polar H10 has been widely used in studies regarding physiology and behavior analysis[54,55].

Experiments design

This study confirms that all experiments and methods were performed in accordance with relevant guidelines and regulations. All recorders in this experiment have received professional training for collecting and recording physiological signal. The participants in this study performed four driving experiments (see Fig. 2): pre-experiments that help drivers adapt to the metro simulator, driving with non-distraction (Experiment I), driving with low-degree distraction (Experiment II), and driving with high-degree distraction (Experiment III).

The low- and high-degree cognitive distractions in this study were induced by a set of N-backs tasks[56,57]. N-backs is a widely utilized cognitive exercise designed to assess working memory and have extensively employed to evoked participants' cognitive distraction[58]. The N-backs test ensures that the driver is mentally occupied. In our study, an electronic notebook was placed next to the driver's seat in the cockpit. During the Experiment II and III, a random number ranging from 1 to 9 was displayed on the notebook screen every 5 s (see Fig. 2). Drivers were asked to recall and report the number displayed in the N-th 5-second period prior to

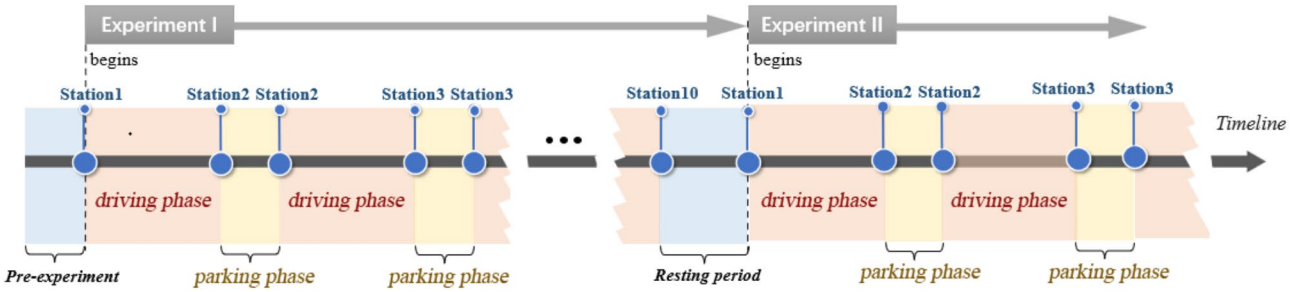


Fig. 3. The detailed schematic diagram of the experimental process.

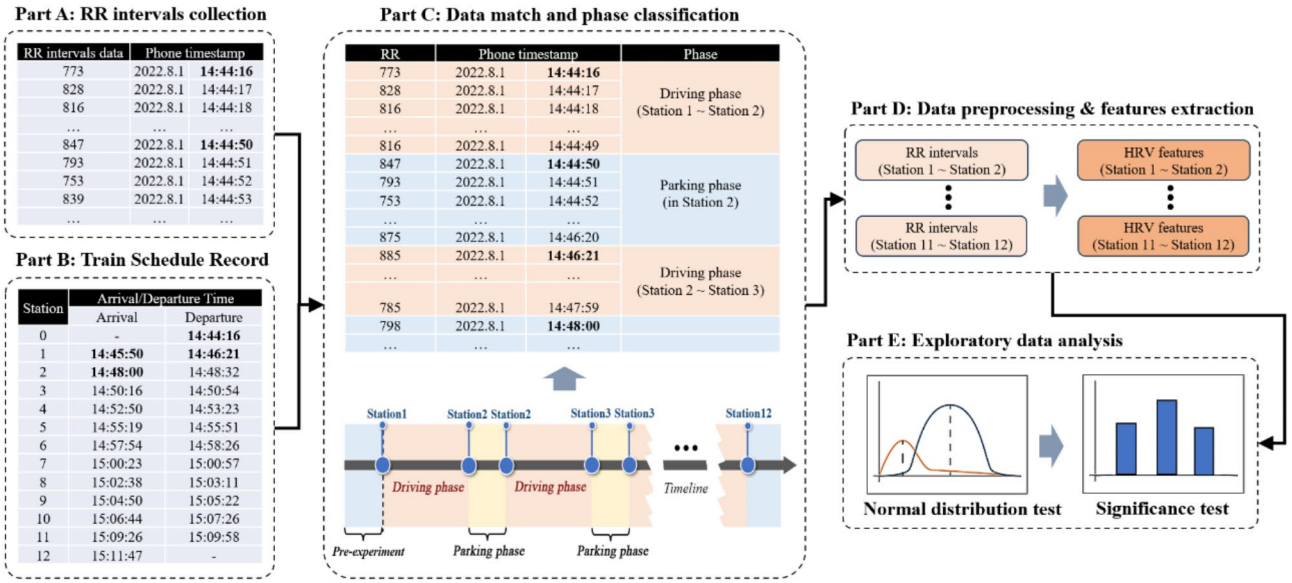


Fig. 4. The workflow of processing and analyzing the ECG signals.

the current one. A 5-second countdown was displayed below the number to inform the driver when the current number would change. Driving without N-back tasks (non-distraction) corresponds to the Experiment I, while driving with 1-back tasks (low-degree distraction) and 2-back tasks (high-degree distraction) correspond to Experiment II and III, respectively. In order to ensure the correctness of the numbers, the driver must remember the numbers while cognitively distracted during the driving task. Therefore, drivers performing the N-backs task will always be distracted.

All the experiments were conducted on a fixed metro line including 10 stops in the driving simulator, from Donggang station to Wenhuagong station. The expected driving time to complete the line was approximately 30 min. Drivers were asked to wait at each stop for 30 s to board virtual passengers and then keep going. After each driving experiment, participants were given a 10-minute rest to prevent fatigue. Each driving experiment is divided into driving phase (i.e., the intervals between two adjacent stations) and parking phase (i.e., when the metro was stopped at a platform), as the physiological responses of the drivers may vary between the two types of phases (see Fig. 3). The experiment recorder recorded the schedule of each driver's train stopping and departing at each station in order to correspond to the subsequent ECG data timeline and determine the driver's ECG data under driving phase/parking phase.

Methods

Feature extraction and statistical analysis

The framework of feature extraction and statistical analysis are illustrated in Fig. 4. It includes Part A: RR interval collection, Part B: Train schedule record, Part C: Data match and phase classification, Part D: Data preprocessing & feature extraction, and Part E: Exploratory data analysis.

1) In Part A, we collect the raw RR intervals (RRI, the time interval between adjacent R-wave peaks of the driver) of metro drivers by the smart phone application connected with Polar H10 during the experiments. The RR intervals are recorded per second.

2) In Part B, the train (metro) schedule is obtained, which records the stop and departure of the metro in each station during the experiments. Note that the Station 0 means the start of the route and this is not a real metro station.
3) In Part C, the phases of metro operation are classified as driving phases and parking phases by examining the train schedule. The parking phase means the metro fully stops at the platform, while driving phase refers to the movement of metro between the adjacent stations. Then, the schedule is used to mark the RR intervals to distinguish the data under driving or parking phases.
4) In Part D, RR intervals are visually inspected and then processed using a Python package named "hrvanalysis" to remove ectopic beats and outliers (outside the range of 280 to 1500 milliseconds, suggested by Jiao, et al.[28]). The 'hrvanalysis' package applies a linear interpolation algorithm to replace null points with interpolated intervals. This package is also used to extract HRV features from the processed RR intervals using 30-s/60-s time-windows (driving phases) and 25-s time-windows (parking phases) with 50% overlap. All the features are scaled using Z-score method before being inputted to prediction models.
5) In part E, the normality of each HRV feature and its distribution differences across the degrees of distraction are statistically examined by Kolmogorov-Smirnov (K-S) tests and Kruskal-Wallis (K-W) tests, respectively.

In total, 29 h-HRV features were extracted from the processed ECG data. Table 3 illustrates the definitions of the HR-HRV features. Meanwhile, it should be mentioned that the experiments for both distracted and non-distracted driving include the same metro stops, and the total durations in both cases are nearly equal. Therefore, sample size variability or data imbalance was not explicitly addressed in training the model.

Feature selection based on RF-RFE

This study uses RFE algorithm to identify the most critical HR-HRV features that are associated with cognitive distractions. In this study, we use a Random Forest (RF) classifier to rank the contribution of each feature subset[61]. The combination of RF and RFE, known as RF-RFE, provides an effective overall feature selection method[62,63]. The main steps of the RF-RFE algorithm are as follows:

(1) An RF model is trained using the full feature set$\{A_1, A_2, \ldots, A_m\}$, where m is the total number of available features. Then, the analysts will predetermine an expected number of the involved features, i.e., K

Variables	Features types	Description
MNN	Time-domain features	Mean of RR intervals
SDNN		Standard deviation of RR intervals
SDSD		Standard deviation of the differences between adjacent RR intervals
NN50		Number of pairs of differences between adjacent RR intervals more than 50 milliseconds
pNN50		Percentage of successive RR intervals differed by more than 50 milliseconds
NN20		Number of pairs of differences between adjacent RR intervals more than 20 milliseconds
pNN20		Percentage of successive RR intervals differed by more than 20 milliseconds
RMSSD		Square root of the mean of the sum of the squares of differences between adjacent RR intervals
MEDNN		Median of RR intervals
RANNN		Difference between the maximum and minimum RR intervals
RMSSD/NN		The ratio of RMSSD to RR intervals
CVNN		The coefficient of variation of normal RR intervals
MHR		Mean heart rate
Max-HR		Maximum of heart rate
Min-HR		Minimum of heart rate
Std-HR		Standard deviation of heart rate
LF	Frequency-domain features	Power of the low frequency band (0.04–0.15 Hz)[59]
HF		Power of the high frequency band (0.15–0.4 Hz)[60]
LF/HF		Ratio of LF to HF
LF norm		LF power in normalized
HF norm		HF power in normalized
TP		Power of the total frequency band
VLF		Power of the very low frequency band (0.003–0.04 Hz)
SD1	Non-linear features	The standard deviation of projection of the Poincaré plot on the line perpendicular to the line
SD2		The standard deviation of the projection of the Poincaré plot on the line
SD1/SD2		Ratio of SD1 to SD2
CVI		Cardiac vagal index
Modified-CSI		The modified cardiac sympathetic index
TRI-index		The triangular index

Table 3. The defines of HR-HRV features.

($K = 1, 2, \ldots, m$). The *K* number is then dynamically changed through cross-validation at each iteration. To clearly see the change of model performance with each removal of an eigenvalue, the iteration *K* is specified as $m - j$ in each iteration, where *j* is the number of iterations.

(2) After the initial training, the RF model ranks the importance of each feature based on criteria such as mean decrease in accuracy. The least important featureA_I ($I \in [1, m]$) among the *m* features is therefore removed from the feature set $\{A_1, A_2, \ldots, A_m\}$. The performance of the model, evaluated using cross-validation, is recorded as the average precision (denoted as P_1)

(3) In subsequent iterations (from 2 to *R*, where *R* is the number of rounds), the least important features, i.e., $A_{II}, A_{III}, ?, A_R$, is sequentially extracted and removed from the feature set. After each removal, the RF model is retrained, and its performance is evaluated. The performance metrics for each round are recorded, denoted as $P_2, P_3, ?, P_R$.

(4) After performing *R* iterations and generating *R* feature subsets (with one feature removed at each step), the subset of features that yields the highest average precision during cross-validation is selected as the optimal feature set.

(5) To ensure robustness, the process is repeated for different values of *K* (the number of features retained), ranging from 1 to *m*. For each value of *K*, a total of *m* RF models are trained and evaluated. The final optimal feature subset is selected based on the RF model that provides the highest precision across all trained models.

.

Machine learning approaches

XGBoost classifier

In this study, XGBoost is used to identify distractions and classify the degrees of distraction. This approach is one of the most popular ensemble machine learning algorithms in transportation safety studies[64]. XGBoost develops a sequence of decision trees in a boosted, parallel structure to achieve an effective evolution process[65]. The objective function of XGBoost usually consists of two parts, including a training loss and a regularization, expressed as:

$$Obj(\theta) = L(\theta) + \Omega(\theta) \tag{1}$$

$$\Omega(\Theta) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{2}$$

where $L(\theta)$ is the training loss function used to measure the model performance on training data; and $\Omega(\Theta)$ is the regularization term used to control the complexity of the model to avoid overfitting; T is the number of leaves; λ and γ are regularization parameters; w_j are the weights of the leaf nodes *j* of the trees.

Traditional machine learning

In addition to the XGBoost, this study also trains several traditional machine learning models to be alternative approaches for identifying the cognitive distractions of metro drivers. These machine learning models include Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Decision Tree (DT). All the developed traditional machine learning models are compared with XGBoost in terms of model metrics.

Model metrics

The performances of the above machine learning models on the identifications or classifications are compared by several metrics, including accuracy, precision, recall, and F1-score. These model metrics are drawn from the confusion matrix with true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). The defines of the model metrics are given by Eqs. (7)-(10):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1 - Score = \left(\frac{Precision \times Recall}{Precision + Recall}\right) \times 2 \tag{6}$$

SHAP technique

Our study uses SHAP technique to illustrate the associations between HR-HRV features and cognitive distractions. In the SHAP framework, the contribution of the feature *i* on the model outputs is allocated based on its marginal contribution. The Shapely values of the *i*-th feature can be computed by:

$$\phi_i = \sum_{S \in M\{i\}} \frac{|S|!(m - |S| - 1)!}{m!} \left[v(S \bigcup \{i\}) - v(S)\right] \tag{7}$$

where M denotes the testing group with m features; S is a subset of features without the i-th feature; $v(S \bigcup \{i\})$ denotes the model outputs with both S feature set and the i-th feature, $v(S)$ is the model outputs with the S feature set only,

Framework of the methodology

The goals of this study are to: (1) identify that whether a driver is experiencing a distraction, which constitutes a binary classification, and (2) classify the degree of the driving distraction, which constitutes a multi-class classification. Figure 5 illustrates the overall framework of the methodology to achieve the goals.

First, HR-HRV features are extracted from raw RR intervals measured in both driving and parking phases. The HR-HRV features extracted using either 30-s/60-s time-windows in driving phase and 25-s time-windows in parking phase yield different classification models. Hence, HR-HRV features from the three types of time-windows are simultaneously examined to determine the most suitable feature subsets using RF-RFE. The full dataset is randomly divided in a 70:30 ratio, corresponding to training data and testing data, respectively.

The training data are then used for model training and parameter tuning with 5-fold cross-validations. Moreover, this study also tests the performance of models trained with 10-fold cross-validations and LOOCV approach in addition to the 5-fold cross-validations.

The testing data used for model evaluation are never involved in the training process. Finally, several machine learning models are then trained to make identifications or classifications for different operational phases using the training data. The outputs of the models that respectively provided the best classifications for driving and parking phases, are interpreted using the SHAP framework.

Results

Results of statistical analysis for HR-HRV features

The study first examines the differences in HR-HRV features across degrees of distraction. K-S tests indicate that all features are not normally distributed, thus K-W tests are then used to unveil the differences. Table 4 summarizes the results of the K-W tests for HR-HRV features in both driving and parking phase at the 95% confidence interval.

The distributions of all HR-HRV features significantly differ across the degrees of distraction in driving phase. As the degree of distraction increases, the features including MNN, SDSD, pNN20, RMSSD, MEDNN, RMSSD/NN, SD1, and CVI decrease, while MHR and Max-HR increase. For parking phase, only nine HR-HRV features show significantly differences across the degrees of distraction. However, all the statistical significances

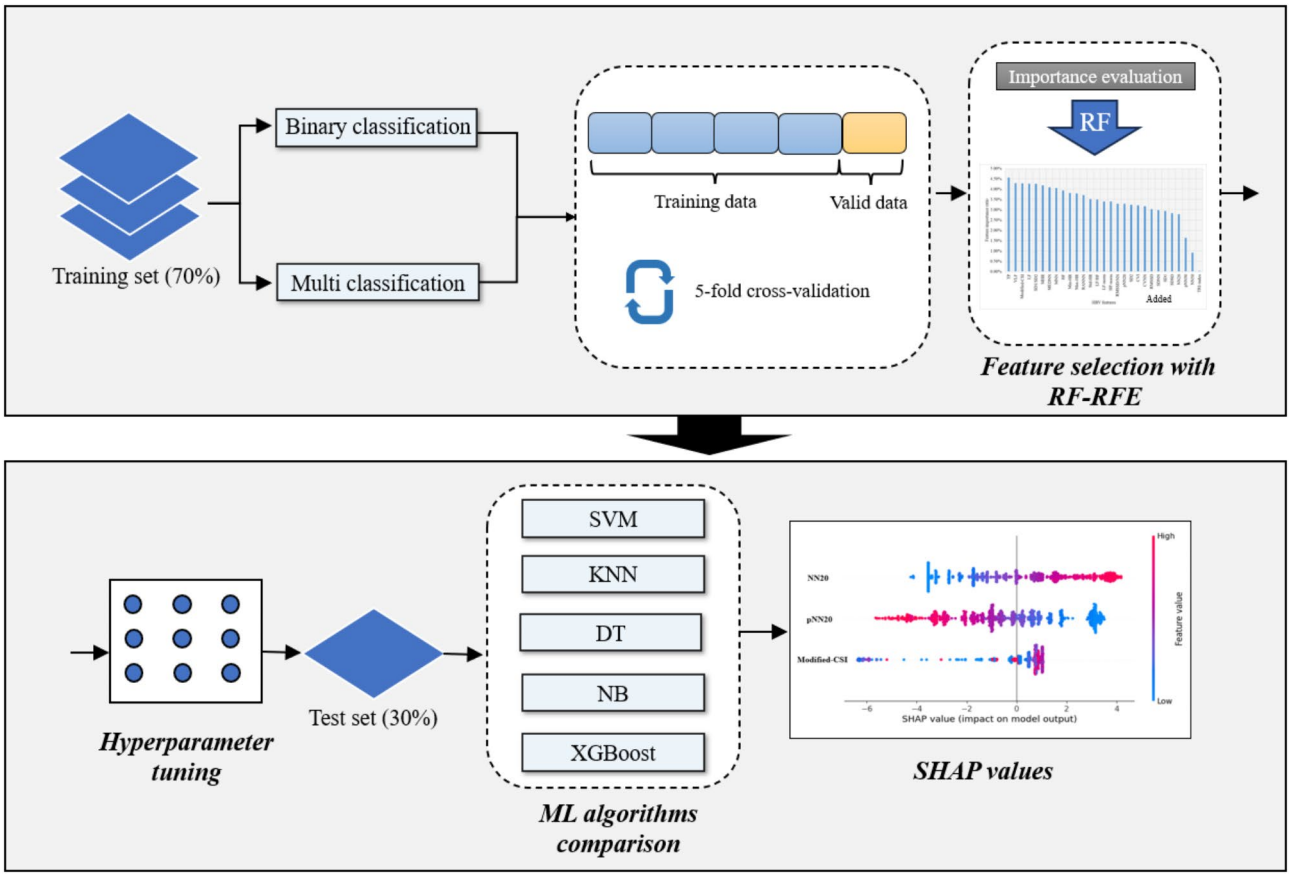


Fig. 5. The proposed analytical framework of the methodology.

Metrics	Driving period							Parking period						
	Non-distraction		Low-degree distraction		High-degree distraction			Non-distraction		Low-degree distraction		High-degree distraction		
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Sig	Mean	S.D.	Mean	S.D.	Mean	S.D.	Sig
MNN	709.76	87.18	698.68	83.36	679.06	93.27	< 0.001*	677.36	69.64	668.71	69.80	657.82	81.98	0.017*
SDNN	33.43	18.93	35.82	17.89	32.60	18.44	< 0.001*	42.24	18.10	39.12	17.18	38.79	18.91	0.018*
SDSD	25.38	13.27	22.13	12.17	21.63	14.22	< 0.001*	22.96	11.13	22.31	10.98	22.48	14.19	0.395
NN50	2.33	3.15	2.02	2.88	2.09	3.61	< 0.001*	2.11	3.17	1.85	2.38	1.86	2.80	0.541
pNN50	7.76	10.51	5.11	7.58	5.30	9.75	< 0.001*	5.86	8.00	5.28	7.03	5.35	8.50	0.491
NN20	10.29	6.30	11.74	7.35	11.25	7.74	0.001*	10.82	7.55	9.91	5.79	10.14	6.39	0.560
pNN20	34.31	21.04	28.45	19.24	26.90	20.02	< 0.001*	29.53	18.83	27.41	17.26	28.02	19.13	0.428
RMSSD	25.46	13.36	22.17	12.19	21.67	14.24	< 0.001*	23.05	11.14	22.36	11.00	22.54	14.20	0.374
MEDNN	708.39	87.66	697.65	84.68	678.48	95.07	< 0.001*	674.15	70.07	665.47	69.95	655.07	83.16	0.018*
RANNN	129.96	77.40	146.38	77.89	136.41	87.83	< 0.001*	159.04	68.33	152.02	67.67	151.97	89.77	0.115
RMSSD/NN	0.03	0.02	0.03	0.02	0.03	0.02	< 0.001*	0.03	0.01	0.03	0.01	0.03	0.02	0.655
CVNN	0.05	0.02	0.05	0.02	0.05	0.02	< 0.001*	0.06	0.02	0.06	0.02	0.06	0.03	0.043*
MHR	85.99	10.19	87.33	10.17	90.31	12.71	< 0.001*	89.85	8.81	91.01	9.12	93.06	12.09	0.016*
Max-HR	93.82	10.56	96.34	11.33	99.61	15.44	< 0.001*	99.56	9.72	100.64	10.83	104.01	16.99	0.020*
Min-HR	78.70	10.86	78.70	10.78	81.92	13.37	< 0.001*	79.34	9.65	80.68	9.75	82.71	12.54	0.016*
Std-HR	3.81	1.77	4.26	1.83	4.12	2.29	< 0.001*	5.33	1.95	5.09	2.04	5.30	2.85	0.182
LF	710.56	1098.85	852.45	1291.23	655.35	1143.89	< 0.001*	1094.21	1743.87	1017.65	1676.20	1030.80	2364.39	0.137
HF	269.82	417.49	231.13	572.01	279.86	634.47	< 0.001*	236.93	293.26	223.24	273.66	317.65	1072.69	0.920
LF/HF	3.90	5.62	6.51	8.33	4.95	6.33	< 0.001*	7.05	7.35	7.57	9.13	7.09	10.92	0.148
LF-norm	65.85	19.34	75.52	17.15	70.62	18.96	< 0.001*	77.66	15.78	75.81	18.71	74.77	17.83	0.148
HF-norm	34.15	19.34	24.48	17.15	29.38	18.96	< 0.001*	22.34	15.78	24.19	18.71	25.23	17.83	0.148
TP	1123.83	1608.94	1342.54	2035.34	1127.48	1938.61	< 0.001*	1702.55	2303.46	1599.33	2099.51	1671.39	3633.93	0.273
VLF	143.46	296.97	258.96	541.01	192.27	438.60	< 0.001*	371.41	498.27	358.44	457.81	322.95	500.67	0.059
SD1	18.25	9.55	15.84	8.73	15.48	10.20	< 0.001*	16.46	7.99	15.99	7.89	16.12	10.19	0.393
SD2	43.30	25.55	47.91	24.16	43.18	24.44	< 0.001*	57.21	24.82	52.75	23.48	52.19	25.25	0.012*
SD1/SD2	2.46	0.76	3.23	1.02	3.13	1.19	< 0.001*	3.73	1.21	3.51	1.05	3.66	1.33	0.143
CVI	3.98	0.44	3.97	0.43	3.88	0.51	< 0.001*	4.09	0.38	4.04	0.39	4.01	0.46	0.102
Modified-CSI	449.95	360.93	631.69	398.44	537.01	380.87	< 0.001*	876.35	525.59	762.64	472.01	752.27	423.17	0.002*
TRI-index	5.86	2.11	6.40	2.13	6.08	2.16	< 0.001*	6.85	2.04	6.66	2.01	6.46	2.03	0.075

Table 4. HR-HRV features under different degrees of distraction and results of Kruskal-Wallis tests. * means the variable is significant in 99% confidence interval.

of the differences, except for Modified-CSI, are above 0.01, indicating that these differences are relatively weak. In parking phase, HR-HRV features including MNN, SDNN, MEDNN, RANNN, HF, LF-norm, VLF, SD2, CVI, Modified-CSI, and TRI-index decrease with an increase in distraction degree, while MHR, Max-HR, Min-HR, and HF-norm increase.

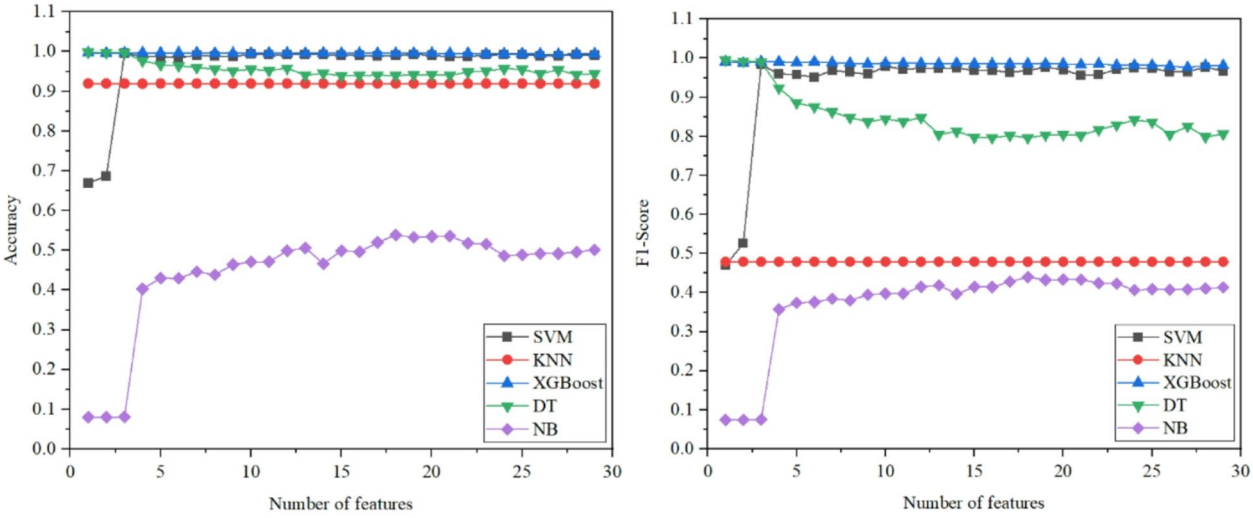
Classifications of driving distractions during driving period

Binary classification

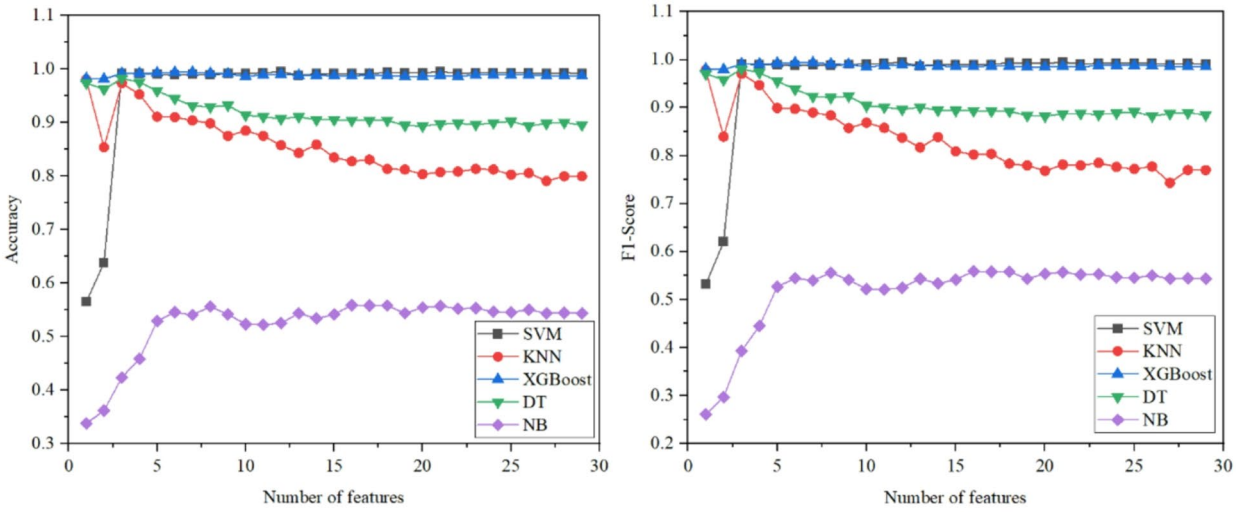
The study first identifies that whether the driver is distracted during driving phase using machine learning models. Figure 6 shows the accuracy and F1-score of the binary classifications with different numbers of inputted HR-HRV features. Note that the number of input features is sequentially determined by the RF-RFE. The selected HR-HRV features used in driving phase are extracted from either 30-s or 60-s time-windows.

For the features from 30-s time windows (Fig. 6(a) and (b)), it is found that SVM, XGBoost, and DT models generally perform better in identifying cognitive distractions compared to other models. Moreover, incorporating the three most critical HR-HRV features yields the best performance for the XGBoost, SVM, and DT models. However, adding more HR-HRV features does not improve performance for these models. In fact, the performance metrics for the DT model decrease as more features are added.

Figure 6(c) and (d) show the accuracy and F1-score of models with different numbers of selected features using HR-HRV features extracted from 60-s time-windows. Among the models, XGBoost also outperforms the others in terms of accuracy and F1-score, followed by the SVM approach. The comparisons between metrics



(a) Accuracy for 30-s time-windows features (b) F1-score for 30-s time-windows features



(c) Accuracy 60-s time-windows features (d) F1-score for 60-s time-windows features

Fig. 6. Metrics of models in binary classifications.

of the developed models using features from 60-s time-windows are similar to those of models using HR-HRV features from 30-s time-windows. However, the F1-score of the KNN model with HR-HRV features from 60-s time windows is significantly better when only a few features are included than the KNN model with features from 30-s time-windows.

The detailed model metrics and the corresponding number of inputted HR-HRV features in fine-tuned models are summarized in Table 5. It is observed that models using HR-HRV features extracted from 30-s time-windows are similar to those using features extracted from 60-s time-windows in terms of model metrics, except for the KNN and NB models. The model comparisons show that the DT with only one HR-HRV feature extracted from 30-s time-windows (Accuracy = 0.998, F1-score = 0.995) and the SVM with 12 h-HRV features extracted from 60-s time-windows (Accuracy = 0.995, F1-score = 0.995) provide the best identifications for cognitive distractions.

Moreover, this study presents the frequency of which each HR-HRV feature is selected in the optimal feature subsets in the fine-tuned ML models. A higher frequency of a feature being included in the optimal subsets, according to the RFE algorithm, indicates a greater importance of this feature. As shown in Fig. 7, TRI-index is the most important feature as it is selected in the optimal feature subsets by all five models using HR-HRV features from both 30-s and 60-s time-windows. Additionally, NN20 and pNN20 from 30-s and 60-s time-

Models	Window length	Number of features	Accuracy	Recall	Precision	F1-Score
SVM	30-s	3	0.995	0.971	0.997	0.984
KNN		all	0.910	0.500	0.460	0.479
XGBoost		1	0.997	0.999	0.984	0.991
DT		1	0.998	0.999	0.990	0.995
NB		18	0.539	0.631	0.539	0.440
SVM	60-s	12	0.995	0.993	0.996	0.995
KNN		1	0.981	0.983	0.975	0.979
XGBoost		6	0.994	0.994	0.994	0.994
DT		3	0.983	0.983	0.979	0.981
NB		16	0.559	0.620	0.619	0.559

Table 5. Metrics of fine-tuned models for binary classifications in driving phase.

windows are respectively selected in the optimal feature subsets by four models, making them the second and third most important features. The remaining features are less frequently selected when extracted from 60-s time-windows compared to those from 30-s time-windows. This suggests that HR-HRV features from 60-s time-windows are generally more important, as they can capture more information than those from 30-s time-windows.

Multi-class classification

The cognitive distractions are further classified by degrees (non-distraction, low-degree distraction, and high-degree distraction) using HR-HRV features. Figure 8 shows the accuracy and F1-score of the classification models with different numbers of input HR-HRV features (30-s and 60-s time-windows).

For the HR-HRV features from 30-s time-windows, the performance of XGBoost and SVM improves as the number of included features increases. Among the models, XGBoost are significantly superior to SVM, KNN, DT, and NB models in terms of model metrics after inputting the 17th feature (i.e., TRI-index). However, other models do not benefit from this input feature, suggesting that XGBoost can more effectively exploit the information from this important feature to distinguish the degrees of distractions.

For the models with HR-HRV features extracted from 60-s time-windows (see Fig. 8), the SVM and XGBoost models are superior to other models. For instance, increasing the number of input HR-HRV features from 1 to 5 significantly enhances the performance of the SVM and XGBoost models, while this is not the case for other models. This demonstrates that the SVM and XGBoost models are more appropriate in classifying the degrees of cognitive distractions with only a few important features available.

Table 6 presents the detailed model metrics and the number of features included in the fine-tuned ML models for distinguishing the degrees of distractions. The results show that models with HR-HRV features from 60-s time-windows outperform those using features from 30-s time-windows, probably because the longer time-windows capture more information about heart rates during cognitive distractions. Additionally, the comparisons between models indicate that XGBoost outperforms other models in classifying the degrees of cognitive distraction using HR-HRV features from both 30-s (accuracy: 0.677; F1-score: 0.744) and 60-s (accuracy: 0.779; F1-score: 0.777) time-windows.

Figure 9 presents the frequency of each HR-HRV feature selected in the optimal feature subsets by the five ML models in the multi-class classifications. It shows that MNN, RMSSD/NN, Max-HR, Min-HR, and SD1/SD2 extracted from 30-s time-windows, and NN20, pNN20, and SD1/SD2 extracted from 60-s time-windows are included in the optimal feature subsets by all five models. This indicates that these features are more important for classifying the degrees of distractions than others. Consistent with the differences observed in binary classifications, there are fewer HR-HRV features from 60-s time-windows included in the optimal feature subsets compared with features from 30-s time-windows.

Association analysis with SHAP

This study uses the SHAP framework to illustrate the associations between the HR-HRV features extracted in driving phase and cognitive distractions by degrees. Note that the interpretations for the binary classifications with SHAP are not provided, as there is only one feature included in the best model.

Figure 10 shows the SHAP summary plots that list the SHAP values of the HR-HRV features in descending order based on their importance in classifying drivers' non-, low-degree, and high-degree distractions. Specifically, a red dot represents a high value for a certain feature in a sample, while a blue dot indicates a low value. The horizontal axis represents the magnitude of the SHAP values of the features in a sample. A positive SHAP value means that the corresponding feature increases the model outputs, i.e., the probability of either non-, low-, or high-degree cognitive distraction, while a negative SHAP value implies a decrease in the model outputs.

According to the SHAP values of the 20 h-HRV features involved in the fine-tuned XGBoost model (Fig. 9a), NN20 and pNN20 are the two most important features associated with non-distractions. NN20 and pNN20 have opposite associations with non-distractions: a higher value of NN20 or a lower value of pNN20 indicates a higher probability of non-distractions. Compared to these two critical HR-HRV features, other features have directional but smaller absolute SHAP values, indicating that they are less important in signaling non-distractions.

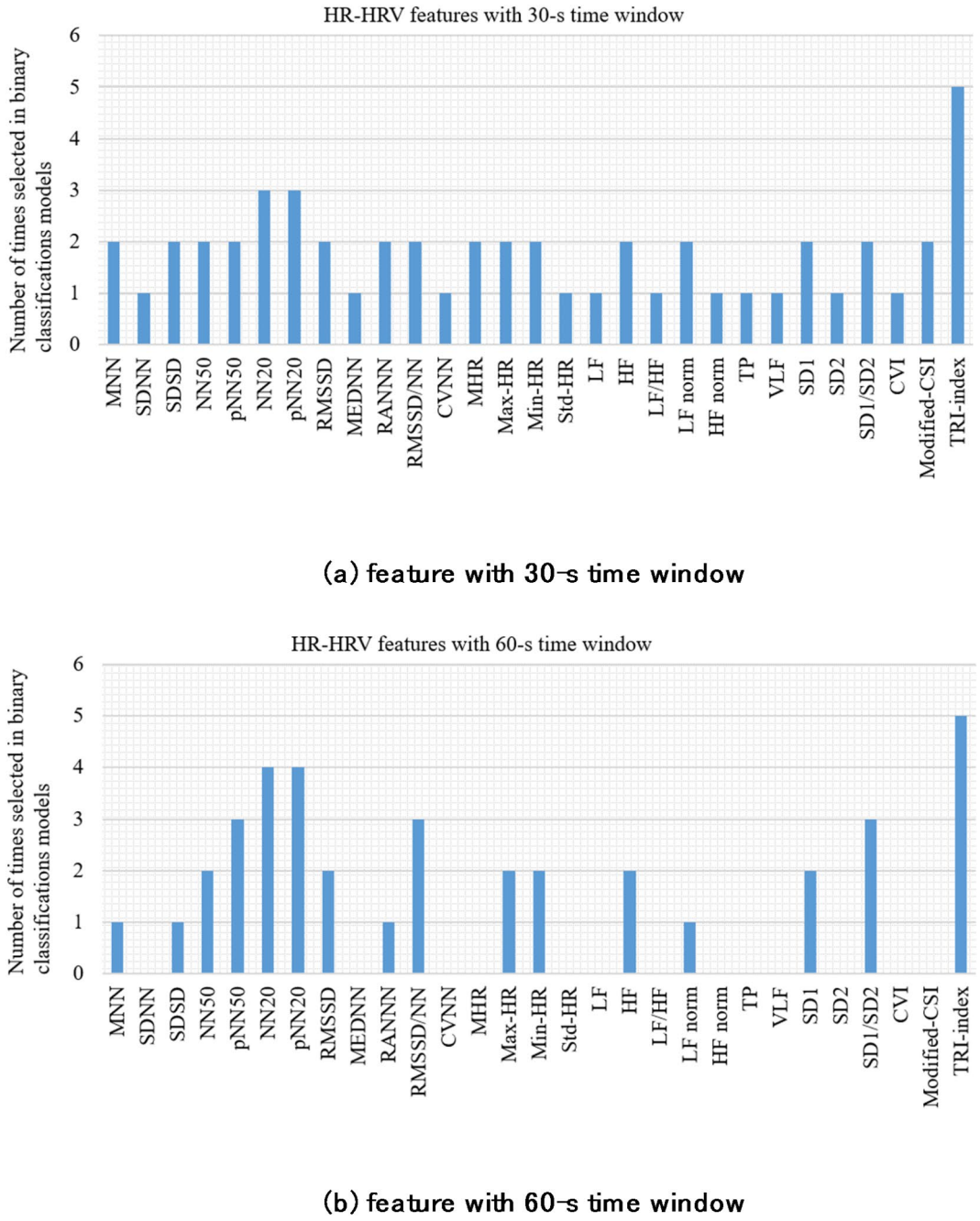
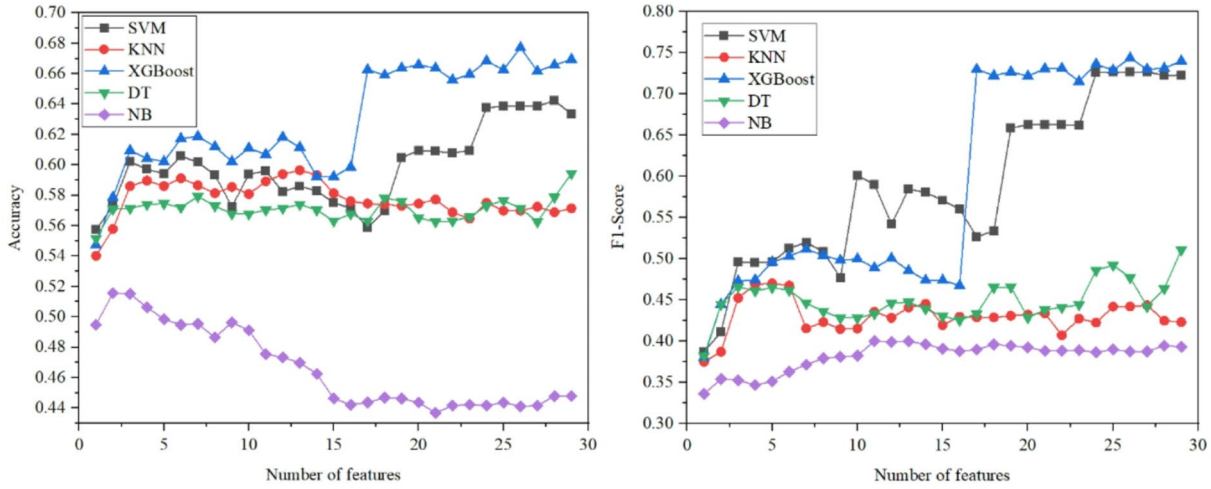
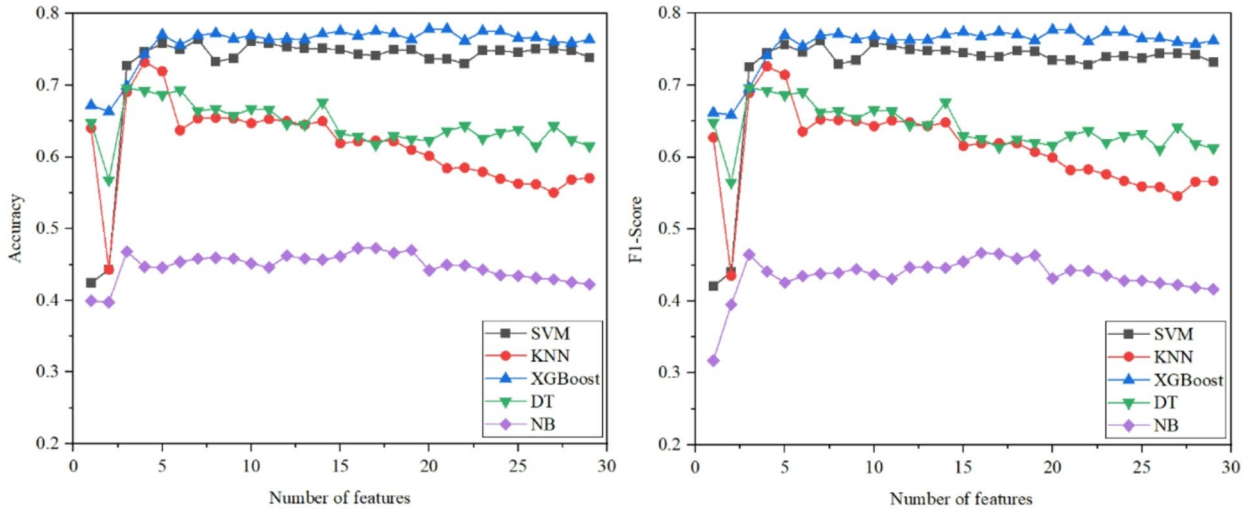


Fig. 7. The frequency of selection for each HR-HRV feature selected in optimal models for binary classifications.

Figure 9 (b) and (c) illustrate that NN20, pNN20, SD1/SD2, Max-HR, MEDNN, and Min-HR are the most important HR-HRV features for indicating low-degree and high-degree distractions. For most samples, NN20 have strictly positive associations with low-degree and high-degree distractions, while pNN20 is negatively linked to the two types of distractions. pNN50 is also found to decrease as degree of distraction increases. However, the associations between other features and distractions vary across the degrees of distraction. For instance, SD1/SD2 is generally positively linked to low-degree distractions, but has a mixed relationship with high-degree distractions. Similarly, Max-HR is negatively associated with low-degree distractions while being positively linked to high-degree distractions in most samples. Other HR-HRV features, such as MEDNN, MNN, Modified-CSI, MHR, RMSSD, also show mixed associations between the low-degree and high-degree distraction and less importances, indicating that the relationships are more uncertain.



(a) Accuracy for 30-s time-windows features (b) F1-score for 30-s time-windows features



(c) Accuracy for 60-s time-windows features (d) F1-score for 60-s time-windows features

Fig. 8. Metrics of models in multi-class classifications.

Models	Window length	Number of features	Accuracy	Recall	Precision	F1-score
SVM	30-s	25	0.639	0.718	0.738	0.727
KNN		5	0.586	0.459	0.555	0.470
XGBoost		26	0.677	0.731	0.758	0.744
DT		all	0.594	0.486	0.639	0.511
NB		11	0.475	0.436	0.418	0.400
SVM	60-s	7	0.764	0.762	0.762	0.762
KNN		4	0.732	0.733	0.740	0.726
XGBoost		20	0.779	0.776	0.778	0.777
DT		3	0.696	0.694	0.699	0.696
NB		16	0.473	0.474	0.472	0.467

Table 6. Metrics of fine-tuned models for multi-class classifications in driving phase.

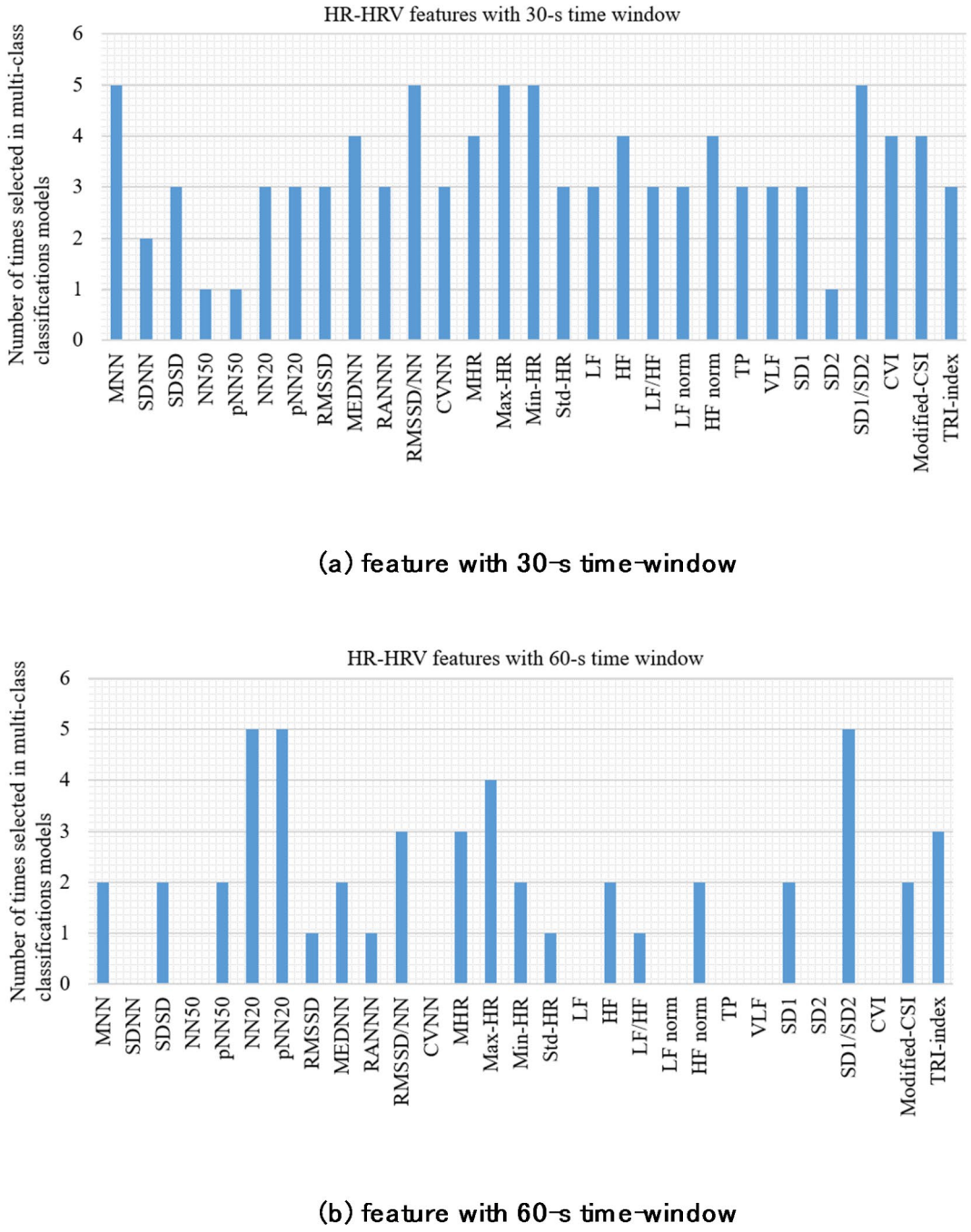


Fig. 9. The frequency of selection for each HR-HRV feature selected in each optimal model for multi-class classifications.

Classifications of driving distraction in parking period

In addition to driving phase, the current study identifies distractions (binary) and distinguishes the degrees of distractions (multi-class) in parking phase. Figure 11 presents the F1-scores of the five ML models used for binary and multi-class classifications with different numbers of input features. The results show that all models perform poorly in both types of classifications, regardless of the number of features. The F1-scores of the models are below 0.6 for binary classifications and are below 0.42 for multi-class classifications.

Additionally, Table 7 lists the detailed information and metrics of the fine-tuned models used for identifications and classifications in parking phase. The results indicate that the SVM model with 6 h-HRV features and the DT model with 17 h-HRV features outperform other models in identifying distractions and classifying the degree of distractions, respectively. However, F1-scores of the two models are only 0.596 and 0.411. Given the poor performance of the models, it can be concluded that the current HR-HRV features may not be effective in identifying drivers' cognitive distractions in parking phase.

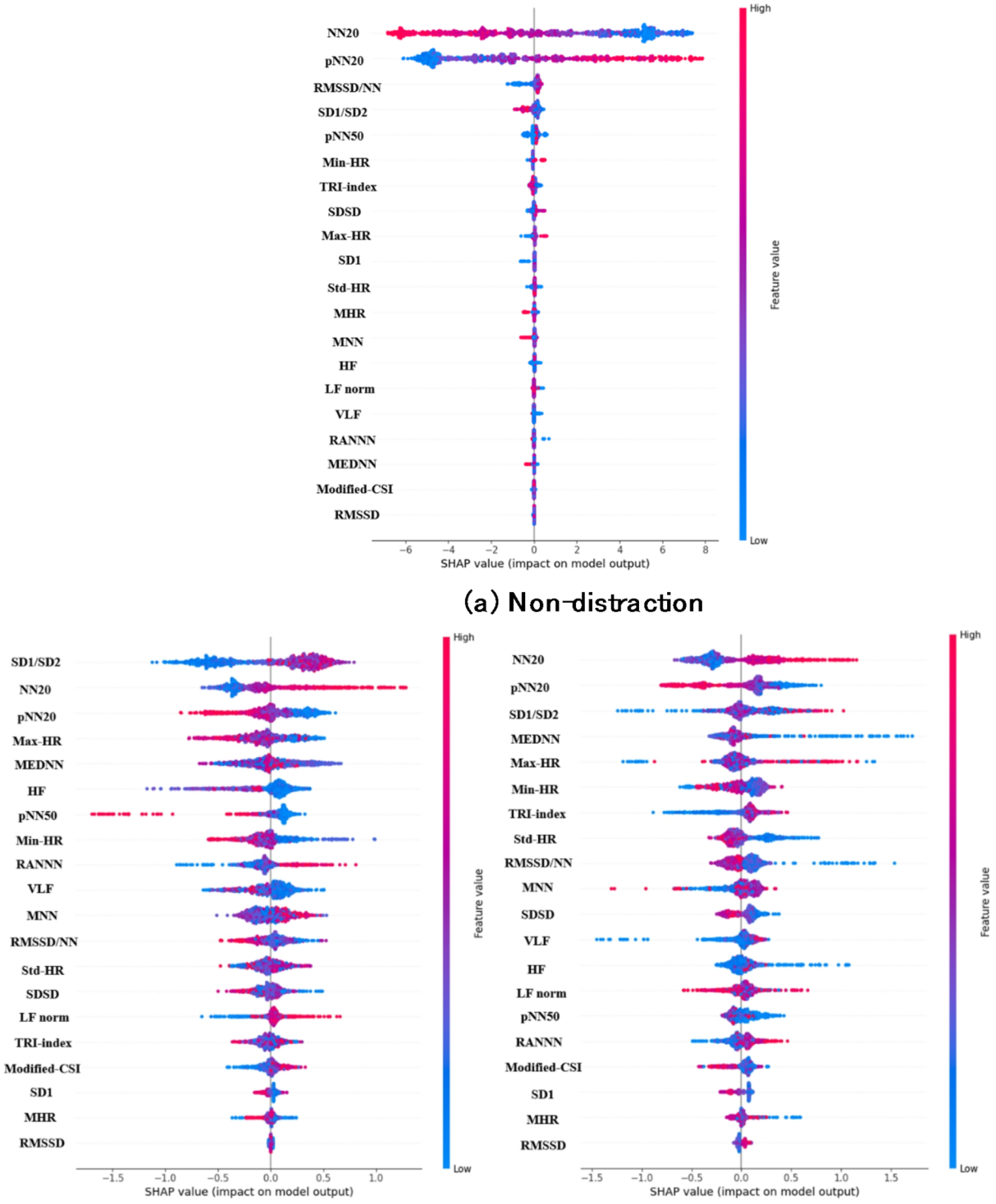


Fig. 10. SHAP summary plots for multi-class classification during driving phase.

Discussion

This study conducts an analysis and proposes detection methods for cognitively distracted metro drivers. According to the results, HR-HRV features are related to metro drivers' cognitive distractions during driving phase. Specifically, HR-HRV features, such as MNN, SDSD, RMSSD, pNN20, and pNN50 decrease if the degree of cognitive distraction increases, while Max-HR and MHR increases. The significant associations are due to that cognitive distractions tend to manifest in altered autonomic nervous system (ANS) regulation reflected in HR-HRV indices[66]. Hence, HR-HRV features can be used for identifying and classifying cognitive distractions in

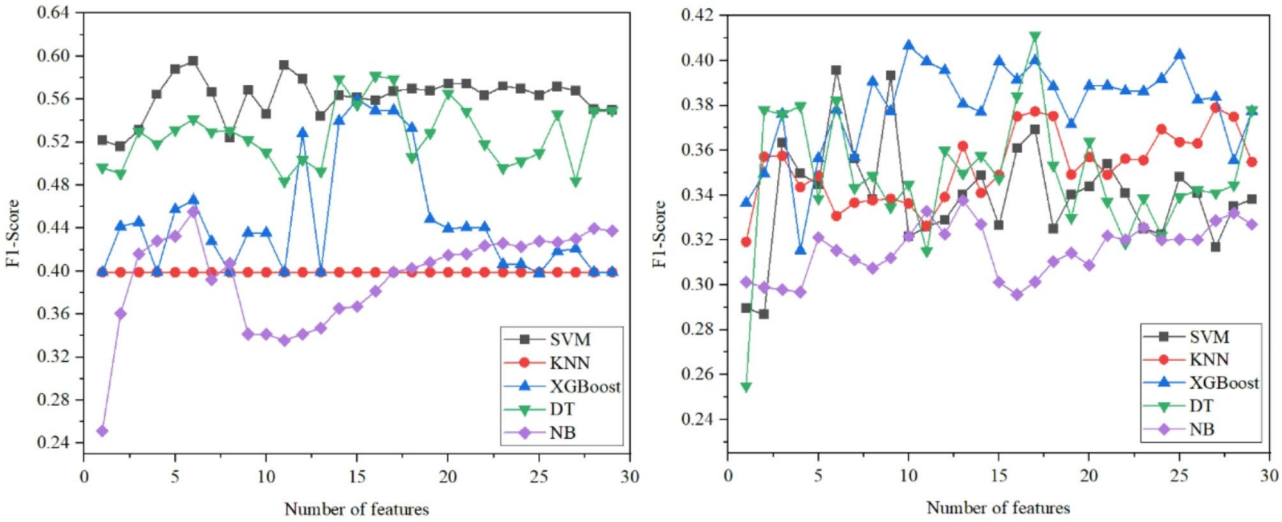


Fig. 11. F1-score of models in binary and multi-class classifications during parking phase (a) Binary classification (b) Multi-class classification.

Classification type	Models	Number of features	Accuracy	Recall	Precision	F1-score
Binary	SVM	6	0.624	0.602	0.595	0.596
	KNN	all	0.664	0.5	0.332	0.399
	XGBoost	15	0.672	0.567	0.613	0.560
	DT	16	0.624	0.582	0.581	0.582
	NB	6	0.462	0.560	0.577	0.455
Multi	SVM	7	0.360	0.359	0.356	0.356
	KNN	18	0.387	0.386	0.389	0.375
	XGBoost	10	0.411	0.410	0.408	0.407
	DT	17	0.411	0.412	0.418	0.411
	NB	13	0.371	0.370	0.406	0.338

Table 7. Metrics of fine-tuned models for binary and multi-class classifications in parking phase.

metro driving phase. However, the cognitive distractions occurring in parking phase are less related to HR-HRV features, with fewer HR-HRV features showing significant variation. This difference can be interpreted through the differences of workload between driving and parking. Drivers typically have fewer manipulations and attention allocations when the metro is parked at platform. Therefore, the cognitive demands are lower during parking phase as the drivers are more relax in coping with the N-backs tasks, demonstrating that HR-HRV features are more sensitive to reflect distractions during high mental demand scenarios than during relatively less demand ones. This finding also implies that reflecting cognitive distractions with HR-HRV features should be restricted in specific scenarios of operation.

Results of machine learning models demonstrate that HR-HRV features perform well in detecting cognitive distractions of metro drivers. Models like XGBoost, SVM, and DT have excellent model metrics in detecting cognitive distractions (binary-class classification) during driving phase. However, the performances of machine learning models are less as expected in classifying the degrees of distractions (multi-class classification), especially for distinguishing the low-degree distractions and high-degree distractions. Among the models, XGBoost achieves the best performance in both binary and multi-class classifications, aligning with previous research highlighting its robustness and ability to handle complex, high-dimensional data[67].

It is also found that adding extra features did not always enhance the performance of the classification model. The reduction in model performance implies that the models are affected by unnecessary features and noises, which also implies that a few features selected play a crucial role in determining the cognitive distractions. These results emphasize the importance of applying feature engineering and selection techniques like RF-RFE to filter the most critical HR-HRV features for constructing distraction detection models.

The effect of 30-s and 60-s time-windows used for capturing HR-HRV features on model performance are also noteworthy to be discussed. According to the results, while the 60-s time-window generally captured more information about the physiological state[26], features extracted from 30-s time-windows are still highly informative for distraction detection. This is consistent with some research findings that short time segments can also extract enough heart rate characteristics and gradually apply to the measurement field[58]. This finding

underscores the flexibility of the proposed classification methods, as it suggests that even HR-HRV features from short time windows can provide valuable insights in detecting cognitive distraction by degrees. This is particularly relevant for constructing real-time detection methods in practical metro driving based on timely responses.

Among the HR-HRV features, TRI-index is consistently selected as the most important feature across all models, indicating its central role in reflecting cognitive workload and mental stress. However, SHAP values of TRI-index in classifying the low-degree and high-degree distraction are relatively small due to the mixed associations. Additionally, NN20 and pNN20 are frequently selected as critical features, with NN20 showing positive correlations with non-distractions and pNN20 showing negative associations with distractions. These trends are in line with previous studies that indicate lower heart rate variability correlates with higher levels of cognitive distraction[68] and cognitive load or stress[69]. Statistical and SHAP analysis also reveal that features like SD1/SD2, Max-HR, and MEDNN are particularly important in distinguishing low-degree and high-degree distractions. The positive relationship between SD1/SD2 and low-degree distractions suggests that subtle changes in heart rate patterns could be indicative of less severe distractions, whereas this feature also has a mixed association with high-degree distraction. Guo, et al.[17] indicate that drivers who in higher distraction have significantly lower pNN50, which is in line with our finding. However, the association or importance of some HR-HRV features in our study is inconsistent with past studies. For instance, Siennicka, et al.[21] argue that higher RMSSD is critical in predicting distractions, while it is less important and has a mixed impact in our study. We also found that LF and LF/HF are less important in classifying the degree of distraction, which is not the case in previous studies[70,71]. This is probably because LF and LF/HF are unable to effectively reflect the drivers' tiny cognitive demand[72]. If the cognitive load caused by the distraction is mild or sporadic, changes in LF and LF/HF may be minimal, leading to reduced importance of these features. Also, the mechanism of distractions of road users are quite different from those of metro drivers, as the road users may suffer more environmental stressors that lead to significant LF and LF/HF. These findings demonstrate HR-HRV features, used as biomarker, may be quite different if the distraction occurs in different scenarios, implying that comprehensive and considerable analyses for cognitive distractions of driving cohorts excluding motor vehicle drivers are necessary.

Conclusions

This study uses HR-HRV features extracted from the ECG signals of metro drivers to determine whether they are experiencing cognitive distractions during manual operations and to distinguish the degrees of these distractions. The ECG data are collected through driving simulation experiments. HR-HRV features are separately extracted using 30-s and 60-s time-windows in driving phase and 25-second time-windows in parking phase. Five ML models, including XGBoost, SVM, KNN, DT, and NB, are employed to identify distractions and classify the degrees of distractions using the optimal HR-HRV features selected by an RF-RFE algorithm.

Several key findings are presented as follows:

1) During driving phase, classifying whether the drivers are distracted is easier than distinguishing the degrees of distraction;
2) During driving phase, models using HR-HRV features from 30-s time-windows outperform others in identifying distractions, while those using HR-HRV features from 60-s time-windows are better at distinguishing the degrees of distraction;
3) The SHAP values indicate that NN20, pNN20, SD1/SD2, Max-HR, Min-HR, and MEDNN are the most important features for classifying the degrees of distraction during driving phase. However, the associations between some features and distractions are mixed and vary across distraction degrees.
4) It is difficult to identify drivers' cognitive distractions in parking phase using HR-HRV features.

Based on the above results, several practical applications can be considered to improve the safety of metro operations:

1) Wearable devices can be considered to collect ECG data from metro drivers and detect their cognitive distractions in real-time, particularly for those frequently experiencing cognitive distractions;
2) Physical countermeasures can be considered. For instance, installing alert systems in the cockpit to warn drivers of cognitive distractions once detected by the wearable devices;
3) Management countermeasures can be considered, e.g., rewarding drivers who are less frequently distracted and penalizing those who are more frequently detected with cognitive distractions. Additionally, special training programs focused on attention allocation and metro operations can be implemented to help metro drivers focus on their operations.

In addition to the findings, several limitations are noted. First, the ECG data is not effective in distinguishing the degrees of distraction, especially in parking phase. Second, distractions are more closely related to brain activities, but cerebral signals of the drivers are not available in this study. Last, since the HR-HRV features used in this study are recorded in a simple data structure, the comparation between machine learning and more advanced deep learning approaches are not considered. Future research may consider incorporating higher-quality physiological signals and developing multimodal models (e.g., a model combining ECG and EEG) with advanced structure to improve the model performance of identifying cognitive distractions in metro drivers.

Data availability

The datasets generated and/or analyzed during this study are not publicly available due to the requirements for the confidentiality of the subjects, but are available from the corresponding author on reasonable request.

Received: 8 December 2024; Accepted: 26 February 2025
Published online: 04 March 2025

References

1. Beanland, V., Fitzharris, M., Young, K. L. & Lenné, M. G. Driver inattention and driver distraction in serious casualty crashes: Data from the Australian National crash in-depth study. *Accid. Anal. Prev.* **54**, 99–107. https://doi.org/10.1016/j.aap.2012.12.043 (2013).
2. Klauer, S., Dingus, T., Neale, T., Sudweeks, J. & Ramsey, D. The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data. *Natl. Highway Traffic Saf. Adm. DOT HS.* **810**, 594 (2006).
3. Global Times. Beijing authority releases investigation result of subway rear-end collision accident; 18 people to be held accountable. (2024). https://www.globaltimes.cn/page/202402/1306759.shtml [Accessed September 7, 2024].
4. Li, N. & Busso, C. Predicting perceived visual and cognitive distractions of drivers with multimodal features. *IEEE Trans. Intell. Transp. Syst.* **16** (1), 51–65. https://doi.org/10.1109/tits.2014.2324414 (2015).
5. Shi, X. Driver distraction behavior detection framework based on the dwpose model, Kalman filtering, and multi-transformer. *IEEE Access.* **12**, 80579–80589. https://doi.org/10.1109/access.2024.3406605 (2024).
6. Wang, X., Xu, R., Zhang, S., Zhuang, Y. & Wang, Y. Driver distraction detection based on vehicle dynamics using naturalistic driving data. *Transp. Res. Part. C: Emerg. Technol.* **136**, 103561. https://doi.org/10.1016/j.trc.2022.103561 (2022).
7. Yu, L., Sun, X. & Zhang, K. Driving distraction analysis by ecg signals: An entropy analysis. In *proceedings of 4th International Conference on Internationalization, Design and Global Development* 6775, Orlando, FL, 258–264 (2011).
8. Wu, X., Shi, C. & Yan, L. Driving attention state detection based on GRU-EEGNet. *Sensors* **24** (16), 5086. https://doi.org/10.3390/s24165086 (2024).
9. Taherisadr, M., Asnani, P., Galster, S. & Dehzangi, O. ECG-Based driver inattention identification during naturalistic driving using Mel-frequency cepstrum 2-D transform and convolutional neural networks. *Smart Health*. **9–10**, 50–61. https://doi.org/10.1016/j.smhl.2018.07.022 (2018).
10. Miyaji, M. Method of driver state detection for safety vehicle by means of using pattern recognition. In *{roceedings of 6th International Conference on Computational Collective Intelligence (ICCCI)* 8733, Seoul, South Korea 194–203.
11. Han, S. Y., Kim, J. W., Lee, S. W. & Ieee Recognition of pilot's cognitive states based on combination of physiological signals. In *Proceedings of 7th International Winter Conference on Brain-Computer Interface (BCI)*, Korea Univ, Res Ctr Brain & Artificial Intelligence, Gangwon, South Korea 232–236 (2019).
12. Shaffer, F. & Ginsberg, J. P. An overview of heart rate variability metrics and norms. *Front. Public. Health*. **5**, 258. https://doi.org/10.3389/fpubh.2017.00258 (2017).
13. Monkaresi, H., Calvo, R. A. & Yan, H. A machine learning approach to improve contactless heart rate monitoring using a webcam. *IEEE J. Biomed. Health Inf.* **18** (4), 1153–1160. https://doi.org/10.1109/JBHI.2013.2291900 (2014).
14. Yang, S. et al. The impacts of Temporal variation and individual differences in driver cognitive workload on ECG-based detection. *Hum. Factors*. **63** (5), 772–787. https://doi.org/10.1177/0018720821990484 (2021).
15. Sogabe, M. et al. Effects of audio and visual distraction on patients' vital signs and tolerance during Esophagogastroduodenoscopy: A randomized controlled trial. *BMC Gastroenterol.* **20** (1), 122. https://doi.org/10.1186/s12876-020-01274-3 (2020).
16. Ramírez, E., Ortega, A. R. & Del Paso, G. A. R. Anxiety, attention, and decision making: The moderating role of heart rate variability. *Int. J. Psychophysiol.* **98** (3), 490–496. https://doi.org/10.1016/j.ijpsycho.2015.10.007 (2015).
17. Guo, X., Tavakoli, A., Chen, T. D. & Heydarian, A. Unveiling the impact of cognitive distraction on cyclists psycho-behavioral responses in an immersive virtual environment. *IEEE Trans. Intell. Transp. Syst.* **25** (8), 10274–10285. https://doi.org/10.1109/TITS.2024.3366777 (2024).
18. Arutyunova, K. R. et al. Heart rate dynamics for cognitive load estimation in a driving simulation task. *Sci. Rep.* **14**, 31656. https://doi.org/10.1038/s41598-024-79728-x (2024).
19. Ouyang, Y. et al. Monitoring inattention in construction workers caused by physical fatigue using electrocardiograph (ECG) and galvanic skin response (GSR) sensors. *Sensors* **23** (17), 7405. https://doi.org/10.3390/s23177405 (2023).
20. Wang, J., Yuan, S., Lu, T., Zhao, H. & Zhao, Y. Fusing YOLOv5s-MediaPipe-HRV to classify engagement in E-learning: From the perspective of external observations and internal factors. *Knowl. -Based Syst.* **305** (3), 112670. https://doi.org/10.1016/j.knosys.2024.112670 (2024).
21. Siennicka, A. et al. Resting heart rate variability, attention and attention maintenance in young adults. *Int. J. Psychophysiol.* **143**, 126–131. https://doi.org/10.1016/j.ijpsycho.2019.06.017 (2019).
22. Wu, L., Shi, P., Yu, H. & Liu, Y. An optimization study of the ultra-short period for HRV analysis at rest and post-exercise. *J. Electrocardiol.* **63**, 57–63. https://doi.org/10.1016/j.jelectrocard.2020.10.002 (2020).
23. Camm, A. J. et al. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. *Eur. Heart J.* **17** (3), 354–381. https://doi.org/10.1161/01.CIR.93.5.1043 (1996).
24. Castaldo, R., Montesinos, L., Melillo, P., James, C. & Pecchia, L. Ultra-short term HRV features as surrogates of short term HRV: A case study on mental stress detection in real life. *BMC Med. Inf. Decis. Mak.* **19**, 1–13. https://doi.org/10.1186/s12911-019-0742-y (2019).
25. Liu, K. et al. Driver stress detection using ultra-short-term HRV analysis under real world driving conditions. *Entropy* **25** (2), 194. https://doi.org/10.3390/e25020194 (2023).
26. Li, K., Rüdiger, H. & Ziemssen, T. Spectral analysis of heart rate variability: Time window matters. *Front. Neurol.* **10**, 545. https://doi.org/10.3389/fneur.2019.00545 (2019).
27. Baek, H. J., Cho, C. H., Cho, J. & Woo, J. M. Reliability of ultra-short-term analysis as a surrogate of standard 5-min analysis of heart rate variability. *Telemed e-Health*. **21** (5), 404–414. https://doi.org/10.1089/tmj.2014.0104 (2015).
28. Jiao, Y. et al. Physiological responses and stress levels of high-speed rail train drivers under various operating conditions-a simulator study in China. *Int. J. Rail Transp.* **11** (4), 449–464. https://doi.org/10.1080/23248378.2022.2086638 (2023).
29. Wu, X. et al. Prediction of the Frost resistance of high-performance concrete based on RF-REF: A hybrid prediction approach. *Constr. Build. Mater.* **333**, 127132. https://doi.org/10.1016/j.conbuildmat.2022.127132 (2022).
30. Al-Tashi, Q., Abdulkadir, S. J., Rais, H. M., Mirjalili, S. & Alhussian, H. Approaches to multi-objective feature selection: A systematic literature review. *IEEE Access.* **8**, 125076–125096. https://doi.org/10.1109/ACCESS.2020.3007291 (2020).
31. Yan, C. et al. A novel hybrid filter/wrapper feature selection approach based on improved fruit fly optimization algorithm and chi-square test for high dimensional microarray data. *Curr. Bioinform.* **16** (1), 63–79. https://doi.org/10.2174/1574893615666200324125535 (2021).
32. Saidi, R., Bouaguel, W. & Essoussi, N. Hybrid feature selection method based on the genetic algorithm and pearson correlation coefficient. *Mach. Learn. Paradig Theory Appl.* **801**, 3–24. https://doi.org/10.1007/978-3-030-02357-7_1 (2019).
33. Beraha, M., Metelli, A. M., Papini, M., Tirinzoni, A. & Restelli, M. Feature selection via mutual information: New theoretical insights. In *Proceedings of 29th international joint conference on neural networks (IJCNN)*, Budapest, Hungary (2019).
34. Albashish, D., Hammouri, A. I., Braik, M., Atwan, J. & Sahran, S. Binary biogeography-based optimization based SVM-RFE for feature selection. *Appl. Soft Comput.* **101**, 107026. https://doi.org/10.1016/j.asoc.2020.107026 (2021).
35. Chandrashekar, G. & Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **40** (1), 16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024 (2014).

36. Wang, C. et al. Indicator element selection and geochemical anomaly mapping using recursive feature elimination and random forest methods in the Jingdezhen region of Jiangxi Province, South China. *Appl. Geochem.* **122**, 104760. https://doi.org/10.1016/j.apgeochem.2020.104760 (2020).
37. Liao, Y. et al. Detection of driver cognitive distraction: A comparison study of stop-controlled intersection and speed-limited highway. *IEEE Trans. Intell. Transp. Syst.* **17** (6), 1628–1637. https://doi.org/10.1109/TITS.2015.2506602 (2016).
38. Speiser, J. L., Miller, M. E., Tooze, J. & Ip, E. A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst. Appl.* **134**, 93–101. https://doi.org/10.1016/j.eswa.2019.05.028 (2019).
39. Gómez-Ramírez, J., Ávila-Villanueva, M. & Fernández-Blázquez, M. Á. Selecting the most important self-assessed features for predicting conversion to mild cognitive impairment with random forest and permutation-based methods. *Sci. Rep.* **10** (1), 20630. https://doi.org/10.1038/s41598-020-77296-4 (2020).
40. Chen, L. L., Zhang, A. & Lou, X. G. Cross-subject driver status detection from physiological signals based on hybrid feature selection and transfer learning. *Expert Syst. Appl.* **137**, 266–280. https://doi.org/10.1016/j.eswa.2019.02.005 (2019).
41. Zuo, X., Zhang, C., Cong, F., Zhao, J. & Hämäläinen, T. Mobile phone use driver distraction detection based on MSaE of multi-modality physiological signals. *IEEE Trans. Intell. Transp. Syst.* **99**, 1–16. https://doi.org/10.1109/TITS.2024.3416382 (2024).
42. Misra, A., Samuel, S., Cao, S. & Shariatmadari, K. Detection of driver cognitive distraction using machine learning methods. *IEEE Access.* **11**, 18000–18012. https://doi.org/10.1109/ACCESS.2023.3245122 (2023).
43. Dehzangi, O., Sahu, V., Taherisadr, M. & Galster, S. Multi-modal system to detect on-the-road driver distraction. In *Proceedings of 21st International Conference on Intelligent Transportation Systems (ITSC)* (2018). https://doi.org/10.1109/ITSC.2018.8569893
44. Papakostas, M., Das, K., Abouelenien, M., Mihalcea, R. & Burzo, M. Distracted and drowsy driving modeling using deep physiological representations and multitask learning. *Appl. Sci.* **11** (1), 88. https://doi.org/10.3390/app11010088 (2020).
45. Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29** (5), 1189–1232 (2001).
46. Lundberg, S. M., Erion, G. G. & Lee, S. I. Consistent individualized feature attribution for tree ensembles. *ArXiv* https://doi.org/10.48550/ArXiv.1802.03888 (2018). 1802.03888.
47. Nohara, Y., Matsumoto, K., Soejima, H. & Nakashima, N. Explanation of machine learning models using Shapley additive explanation and application for real data in hospital. *Comput. Methods Programs Biomed.* **214**, 106584. https://doi.org/10.1016/j.cmpb.2021.106584 (2022).
48. Zhou, F. et al. Predicting driver fatigue in monotonous automated driving with explanation using GPBoost and SHAP. *Int. J. Hum. Comput. Interact.* **38** (8), 719–729. https://doi.org/10.1080/10447318.2021.1965774 (2022).
49. Zhou, X., Ma, L. & Zhang, W. Event-Related driver stress detection with smartphones among young novice drivers. *Ergonomics* **65** (8), 1154–1172. https://doi.org/10.1080/00140139.2021.2020342 (2022).
50. Strle, G., Košir, A., Sodnik, J. & Pečečnik, K. S. Physiological signals as predictors of cognitive load induced by the type of automotive head-up display. *IEEE Access.* **11**, 87835–87848. https://doi.org/10.1109/ACCESS.2023.3305383 (2023).
51. Huang, J., Huang, X., Peng, Y. & Hu, L. Driver state recognition with physiological signals: Based on deep feature fusion and feature selection techniques. *Biomed. Signal. Process. Control.* **93**, 106204. https://doi.org/10.1016/j.bspc.2024.106204 (2024).
52. Fasanmade, A. *A context aware classification system for monitoring drivers' distraction levels* (Montfort University, 2021).
53. Alreshidi, I., Bisandu, D. & Moulitsas, I. Illuminating the neural landscape of pilot mental States: A convolutional neural network approach with Shapley additive explanations interpretability. *Sensors* **23** (22), 9052. https://doi.org/10.3390/s23229052 (2023).
54. Villani, V., Righi, M., Sabattini, L. & Secchi, C. Wearable devices for the assessment of cognitive effort for human-robot interaction. *IEEE Sens. J.* **20** (21), 13047–13056. https://doi.org/10.1109/JSEN.2020.3001635 (2020).
55. Corcoba Magana, V. et al. The effects of the Driver's mental state and passenger compartment conditions on driving performance and driving stress. *Sensors* **20** (18), 5274. https://doi.org/10.3390/s20185274 (2020).
56. Fu, R., Zhou, Y., Yuan, W. & Han, T. Effects of cognitive distraction on speed control in curve negotiation. *Traffic Inj Prev.* **20** (4), 431–435. https://doi.org/10.1080/15389588.2019.1602769 (2019).
57. He, B., Li, P., Merat, N. & Li, Y. Time-Dependency-Aware driver distraction detection using linear-chain conditional random fields. In *Proceedings of 23rd IEEE International Conference on Intelligent Transportation Systems (ITSC)*, Electr Network (2020).
58. Lithfous, S., Després, O., Devanne, J., Pebayle, T. & Dufour, A. Preserved distraction analgesia but greater impact of pain on task performance in older adults compared with younger subjects. *Pain Med.* **24**, 818–828. https://doi.org/10.1093/pm/pnac207 (2022).
59. Saboul, D., Pialoux, V. & Hautier, C. The breathing effect of the LF/HF ratio in the heart rate variability measurements of athletes. *Eur. J. Sport Sci.* **14** (1), S282–288. https://doi.org/10.1080/17461391.2012.691116 (2014).
60. Akselrod, S. et al. Power spectrum analysis of heart rate fluctuation: A quantitative probe of beat-to-beat cardiovascular control. *Science* **213** (4504), 220–222. https://doi.org/10.1126/science.6166045 (1981).
61. Yang, S., Gu, L., Li, X., Jiang, T. & Ren, R. Crop classification method based on optimal feature selection and hybrid cnn-rf networks for multi-temporal remote sensing imagery. *Remote Sens.* **12** (19), 3119. https://doi.org/10.3390/rs12193119 (2020).
62. Park, J. et al. Assessing workload in using electromyography (EMG)-based prostheses. *Ergonomics* **67** (2), 257–273. https://doi.org/10.1080/00140139.2023.2221413 (2024).
63. Jiang, X., Zhang, Y., Li, Y. & Zhang, B. A. Forecast and analysis of aircraft passenger satisfaction based on RF-RFE-LR model. *Sci. Rep.* **12** (1), 11174. https://doi.org/10.1038/s41598-022-14566-3 (2022).
64. Gao, W., Wang, W., Dimitrov, D. & Wang, Y. Nano properties analysis via fourth multiplicative ABC indicator calculating. *Arab. J. Chem.* **11** (6), 793–801. https://doi.org/10.1016/j.arabjc.2017.12.024 (2018).
65. Ogungbire, A. & Pulugurtha, S. S. Effectiveness of data imbalance treatment in weather-related crash severity analysis. *Transp. Res. Rec.* **03611981241239962** https://doi.org/10.1177/03611981241239962 (2024).
66. Sogabe, M. et al. The influence of various distractions prior to upper gastrointestinal endoscopy: A prospective randomized controlled study. *BMC Gastroenterol.* **18** (1), 132. https://doi.org/10.1186/s12876-018-0859-y (2018).
67. Zhang, Y. et al. Comparison of prediction models for acute kidney injury among patients with hepatobiliary malignancies based on XGBoost and LASSO-logistic algorithms. *Int. J. Gen. Med.* **14**, 1325–1335. https://doi.org/10.2147/IJGM.S302795 (2021).
68. Hidalgo-Muñoz, A. R. et al. Respiration and heart rate modulation due to competing cognitive tasks while driving. *Front. Hum. Neurosci.* **12**, 525. https://doi.org/10.3389/fnhum.2018.00525 (2019).
69. Tervonen, J., Pettersson, K. & Mäntyjärvi, J. Ultra-short window length and feature importance analysis for cognitive load detection from wearable sensors. *Electronics*. 10 (5): 613 (2021). https://doi.org/10.3390/electronics10050613
70. Neumann, S. A., Waldstein, S. R., Sellers, J. J. 3, Thayer, J. F., Sorkin, J. D. & rd, & Hostility and distraction have differential influences on cardiovascular recovery from anger recall in women. *Health Psychol.* **23**, 631–640. https://doi.org/10.1037/0278-6133.23.6.631 (2004).
71. Ohkura, M., Takahashi, N., Seki, H., Saito, Y. & Hasegawa, K. Study on driver distraction indexes using biological signals applying assessment of information display. In *Proceedings of 15th International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (2013).
72. Koskelo, J. et al. Cardiac autonomic responses in relation to cognitive workload during simulated military flight. *Appl. Ergon.* **121**, 104370. https://doi.org/10.1016/j.apergo.2024.104370 (2024).

Author contributions

Liu. H wrote the main manuscript text. All authors reviewed the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-92248-6.

Correspondence and requests for materials should be addressed to C.J.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.