WILEY | Hindawi

Research Article

Identifying Big Five Personality Traits through Controller Area Network Bus Data

Yameng Wang (ID),[1,2] **Nan Zhao** (ID),[1] **Xiaoqian Liu** (ID),[1] **Sinan Karaburun** (ID),[3] **Mario Chen** (ID),[4] **and Tingshao Zhu** (ID)[1]

[1]*CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China*
[2]*School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing, China*
[3]*BMW China Automotive Trading Ltd., Beijing, China*
[4]*BMW China Services Ltd., Beijing, China*

Correspondence should be addressed to Tingshao Zhu; tszhu@psych.ac.cn

As adapting vehicles to drivers' preferences has become an important focus point in the automotive sector, a more convenient, objective, real-time method for identifying drivers' personality traits is increasingly important. Only recently has increased availability of driving signals obtained via controller area network (CAN) bus provided new perspectives for investigating personality differences. This study proposes a new methodology for identifying drivers' Big Five personality traits through driving signals, specifically accelerator pedal angle, frontal acceleration, steering wheel angle, lateral acceleration, and speed. Data were collected from 92 participants who were asked to drive a car along a pre-defined 15 km route. Using statistical methods and the discrete Fourier transform, some time-frequency features related to driving were extracted to establish models for identifying participants' Big Five personality traits. For these five personality trait dimensions, the coefficients of determination of effective predictive models were between 0.19 and 0.74, the root mean squared errors were between 2.47 and 4.23, and the correlations between predicted scores and self-reported questionnaire scores were considered medium to strong (0.56–0.88). The results showed that personality traits can be revealed through driving signals, and time-frequency features extracted from driving signals are effective in characterizing and identifying Big Five personality traits. This approach could be of potential value in the development of in-car integration or driver assistance systems and indicates a possible direction for further research on convenient psychometric methods.

1. Introduction

It has been shown that personality traits can be used to explore individuals' potential needs in different contexts, such as driving. Several studies have demonstrated that risky driving behaviors are positively associated with neuroticism and extraversion [1–3], but negatively associated with agreeableness, openness, and conscientiousness [1, 4, 5]. Furthermore, Shen et al. [6] found that positive driving behaviors are negatively correlated with neuroticism and positively correlated with openness, conscientiousness, extraversion, and agreeableness. Currently, there is a need for individualization of vehicles in the automotive industry with

the aim of improving driving experiences [7]. Thus, it has become an important focal point [8] to adapt the vehicle to drivers' preferences (e.g., personality).

In the traditional method of measuring personality traits, self-report questionnaires, such as the 44-item Big Five Inventory (BFI-44), are used [9]. Although personality traits are relatively stable individual psychological variables that do not need to be measured frequently over a short period of time [10], relying on self-report questionnaires limits its potential to improve driving experiences in some scenarios. For instance, for nonfixed drivers (e.g., taxis, rental cars, and family cars) filling out a questionnaire before every time they drive not only does not meet a driver's need for vehicle

adaptation but also takes considerable time and concentration, which limits the availability and effectiveness of self-reported personality traits. Therefore, a more convenient, objective, real-time method to identify drivers' personality traits has become increasingly important.

Sensors and electronic control units (ECUs) have only recently become increasingly common in the automotive industry, because they not only guarantee optimal engine function, but also provide a large amount of almost real-time data about the car, driver, and surrounding environment. Various tools, including sensors, driving simulators, and controller area network (CAN) bus data logger, have been applied to conduct studies and many meaningful conclusions have been achieved. For instance, regarding the results of vehicle acceleration and steering behaviour analysis as indicators of driving safety, Wu et al. [11] attempted to determine the optimum design of pavement marking to reduce the rutting on asphalt pavements. Besides, for safety consideration, driving simulators have been used in studies where field operating tests cannot be carried out, such as the study investigating the safety of trucks under crosswind of tunnel and bridge sections [12]. Moreover, the CAN data have been used for the communication among ECUs mounted to a car [13], the tuning problem of digital proportional-integral-derivative parameters for a DC motor [14], and integrated motor-transmission powertrain systems [15]. Additionally, as one of the five protocols used in OBD-II vehicle standards, CAN technology has become the standard for automotive embedded systems [16]. The increased availability of rich driving data has provided new perspectives for investigating individual behaviors and psychological indicators.

With the advantage of high quality and fine data granularity of driving signals provided by in-vehicle sensors, many studies have been conducted based on these data. It has been demonstrated that driving signals can be used to recognize drunk driving behaviors [17], identify drivers [18], and detect anomalous driving [19]. Additionally, the capability of recording real-time driving information is soon used in other applications with the help of machine learning technology [20]. Furthermore, Wan et al. [21] attempted to detect anger states while driving based on multiple sensor signals using a least square support vector machine model (82.20% accuracy rate).

In summary, although there has been evidence that driving signals can reveal personality traits, the method of identifying personality traits based on driving signals has not been established in previous studies. It motivates our efforts to intensively explore the possibility of a solution for real-time identification of personality traits through driving signals. In this work, we aimed to construct feature sets from raw driving signals provided by in-vehicle sensors using CAN bus and identify Big Five personality traits based on these features using a machine learning approach.

2. Materials and Methods

In this section, a methodology with the aim of identifying personality traits through CAN bus data is proposed. Using statistical methods and the discrete Fourier transform, the features related to personality traits are extracted from raw driving signals provided by in-vehicle sensors using CAN bus in the time and frequency domains, respectively. These features will be then used to identify Big Five personality traits automatically by the linear regression, support vector regression, etc. In the study, a four-step procedure was conducted: (1) Data collection, (2) Data preprocessing, (3) Feature extraction and selection, and (4) Model training, as shown in Figure 1.

2.1. Data Collection

2.1.1. Experimental Settings. A BMW i3 test vehicle was equipped with a data logger to record the signals on the CAN bus at the sampling frequency of 10 data points per second for this study. We collected data from 92 participants (52 males and 40 females) who were recruited using convenience sampling from BMW China. All of the participants were asked to drive the BMW i3 test vehicle on a pre-defined route as shown in Figure 2. The pre-defined car route was 15 km and included traffic lights stop signs, surface streets etc. With this user consistent driving task, we wanted to eliminate interference information, so as to explore deeper insights between driving behavior and personality traits. To facilitate data analysis, we divided the route into different sub-routes according to road conditions, and an instructor sitting in the copilot recorded the time that the car passed through different sub-routes during the experiment.

Once the procedure of driving signals collection was done, each participant was required to complete the BFI-44 to measure their Big Five personality traits. The questionnaire consists 44 items and five subscales: openness (10 items); conscientiousness (9 items); extraversion (8 items); agreeableness (9 items); and neuroticism (8 items). Each item of BFI-44 is assessed on a Likert 5-point scale, ranging from 1 ("disagree strongly") to 5 ("agree strongly"). In this study, the Chinese version of the questionnaire was implemented. Its validity and reliability has been proved [22].

2.1.2. Signals Selection. Among the signals transmitted on the CAN bus, the analyses of this study focused on five signals recorded at the sampling frequency of 10 data points per second: accelerator pedal angle, frontal acceleration, steering wheel angle, lateral acceleration, and speed. Compared with other signals, these signals are not only more stable and easy to obtain on different types or models of vehicles, but also can reflect drivers' driving behavior from different aspects. For instance, accelerator pedal and steering wheel signals are the direct output of drivers that directly reflect the interaction between the driver and the vehicle [23]; speed and accelerations are measures of drivers' driving style [24] that can reflect drivers' specific driving preferences and habits, e.g., harsh accelerations or speeding. An example of these signals is shown in Figure 3.
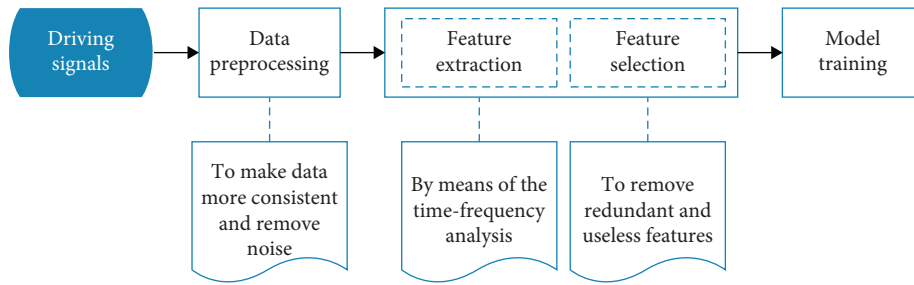
FIGURE 1: The procedure for identifying Big Five personality traits from driving signals.

FIGURE 2: Pre-defined route during data collection (image blurred for anonymity purposes).
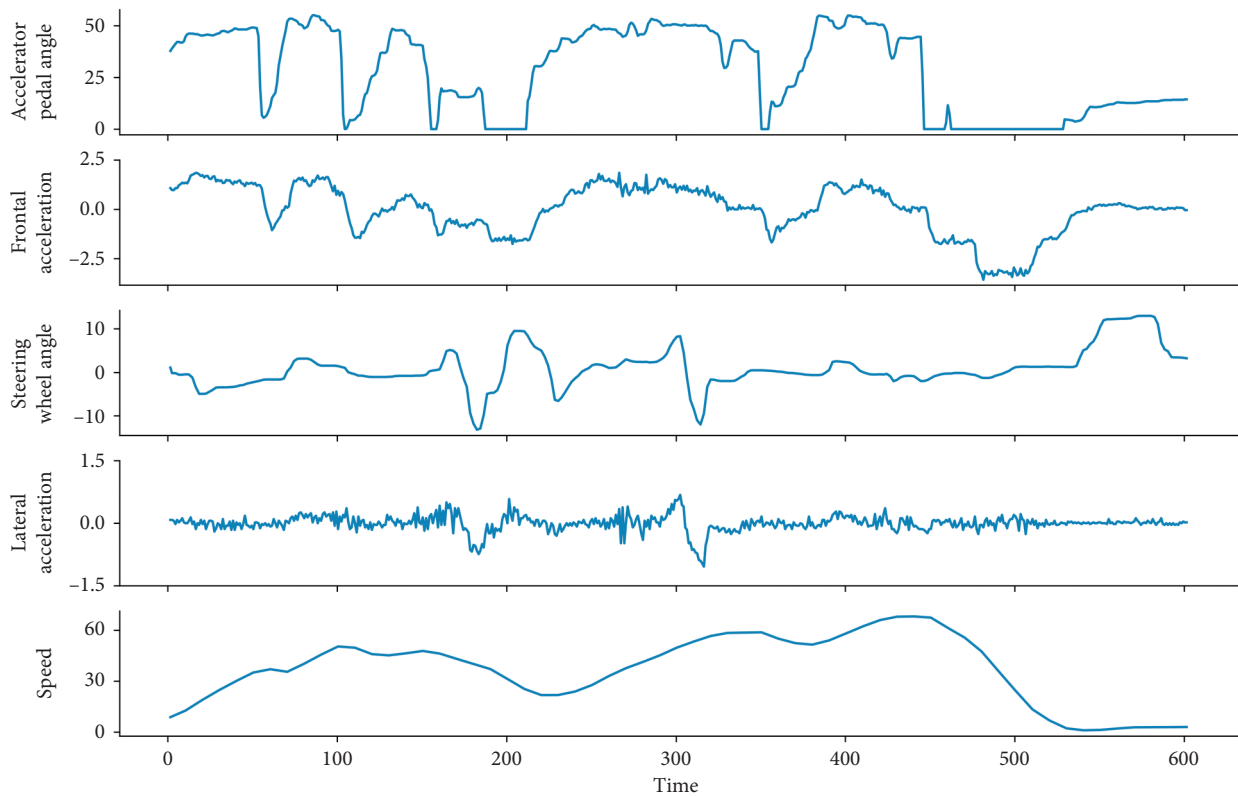

FIGURE 3: Example of signals acquired from the CAN bus.

2.2. Data Preprocessing.

Raw driving signals with noisy and redundant information may bring more redundancy and complexity for models training and affect the performance of recognition models. Therefore, we need to preprocess the raw driving signals, which includes two steps: (1) data segmentation and (2) low-pass filtering.

2.2.1. Data Segmentation.

Since driving under the same road conditions can be regarded as repetitive behaviors, large amounts of repetitive data may lead to low computational efficiency and data redundancy. In addition, it is difficult to guarantee the consistency of road conditions such as corners or curved roads in the actual driving environment. Then recognition models trained based on data obtained under such road conditions may have a poor generalization ability in practice. In this work, we analyzed driving signals of a straight sub-route from point A to point B (as shown in Figure 1). On average, participants took 26.03 minutes (SD = 7.48) to complete the course. For the consistency of driving data, we used driving signals for the first 9600 data points (16 minutes).

2.2.2. Low-Pass Filtering.

As unexpected jolts or vibrations might cause some noise or high-frequency components in data collection, we should do the job of filtering on the raw driving signals as the signal processing. Gaussian filter is a low-pass filter, attenuating noises and high-frequency components in signal data [25]. We computed the convolution of each driving signal and the Gaussian filter, whose window length is 5, and whose coefficients are $g = (1/16)[1, 4, 6, 4, 1]$. The procedure of filtering is defined as

$$y(n) = \sum_{t=-\infty}^{\infty} x(t)g(n-t) = x(n) * g(n), \quad (1)$$

where x is the driving signal, $*$ stands for convolution, and g denotes the Gaussian filter. We take a fragment of the frontal acceleration as an example. After low-pass filtering, the filtered data (See Figure 4(b)) are smoother compared to the raw data (See Figure 4(a)). And many little fluctuations and burrs shown in the red circle in Figure 4(a) are removed.

2.3. Feature Extraction and Selection.

After data preprocessing, we then need to extract and select features from driving signals that can effectively characterize the Big Five personality traits. Specifically, using the time-frequency analysis method, we first extracted features in the time and frequency domains, respectively. And then we find and remove redundant information from these features by dimensionality reduction and feature selection.

2.3.1. Temporal Domain Features Extraction.

Temporal domain information related to the statistical value of driving signals (e.g., mean value, median value, and standard deviation value) was used to characterize drivers' behavior patterns. Since the global statistical value of signals cannot reflect the details of driving behavior, this information was integrated into a given sliding temporal window. Specifically, in a temporal window of width w, we defined the set of data $U_{j \in I_i} x_j$, $I_i = \{i + 1, i + 2, \cdots, i + w\}$ and the following features:

(1) Moving median: the median value of the set.

(2) Moving mean: the mean value of the set

(3) Moving standard deviation: the standard deviation value of the set.

To exam linear dependence of a signal, we estimated autocorrelation and partial autocorrelation of different lags. Specifically, autocorrelation is the correlation of a signal with a delayed copy of itself [26], which is defined as

$$r_k = \frac{\sum_{t=1}^{n-k} (x_t - \mu)(x_{t+k} - \mu)}{\sum_{t=1}^{N} (x_t - \mu)^2}, \quad (2)$$

where n refers to the length of the signal, μ refers to mean of the signal, and k refers to the lag. Partial autocorrelation gives the partial correlation of a stationary time series with its own lagged values [26], which is defined as

$$p_k = \frac{\text{cov}(x_t, x_{t-k}|x_{t-1}, \ldots, x_{t-k+1})}{\sqrt{\text{var}(x_t|x_{t-1}, \ldots, x_{t-k+1})\text{var}(x_{t-k}|x_{t-1}, \ldots, x_{t-k+1})}}, \quad (3)$$

where cov refers to the covariance and var refers to the variance and k refers to the lag.

For each signal, we obtained 45 statistical values through a temporal window of 2 minutes with an overlap ratio of 50%. By setting different delays from 2 seconds to 20 seconds in steps of 2 seconds ($k = 20, 40, \cdots, 200$), we extracted 20 linear dependence features. Finally, we obtained a total of $(45 + 20) * 5 = 325$ time domain features.

2.3.2. Frequency Domain Features Extraction.

In addition to temporal domain features extracted using statistical methods, we conducted. The discrete Fourier transform to convert data from temporal domain to frequency domain [27]. The formula is defined as

$$F_k = \sum_{j=0}^{n-1} x_j e^{-i2\pi k(i/n)}, \quad (4)$$

where n refers to the length of the signal, i is the sign of complex number.

For each signal, we chose the first 100 amplitudes and phases, respectively. Finally, we obtained a total of $(100 + 100) * 5 = 1000$ frequency domain features.

2.3.3. Dimensionality Reduction.

It must be emphasized that driving signals may be interrelated. For instance, the average Pearson correlation coefficient between different signals is shown in Figure 5. Therefore, some of the $325 + 1000 = 1325$ features may be closely related. This redundant information
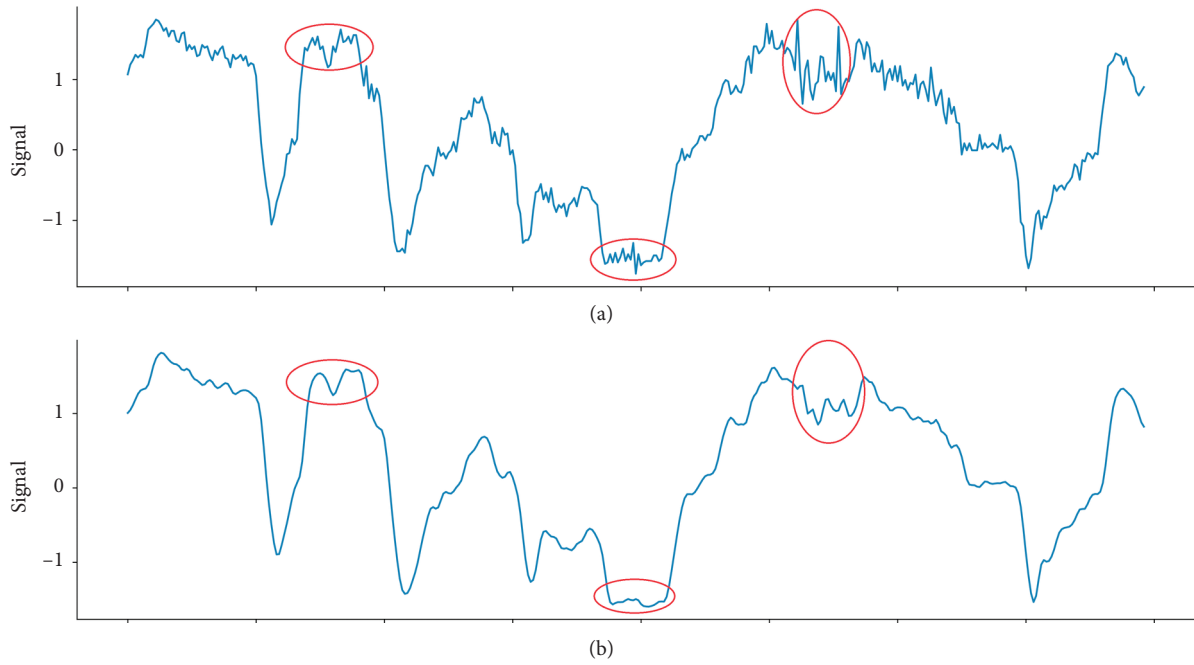
(a)

(b)

FIGURE 4: Example of signals before (a) and after (b) filtering.
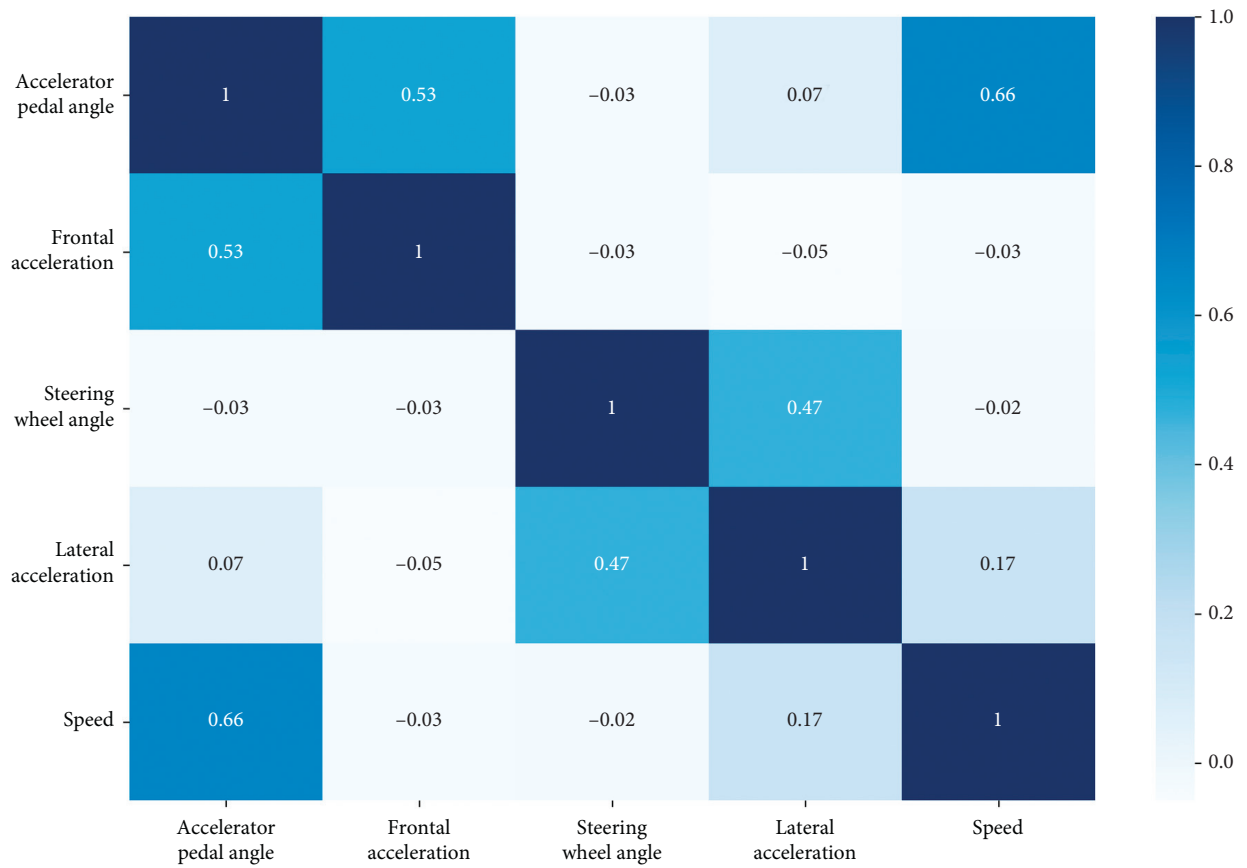

FIGURE 5: A matrix showing the correlation between the five signals acquired from the CAN bus.

may impact the performance of recognition models, so we need to reduce the redundancy of the feature set.

Since the values of different signals were measured on different scales, in case some important features extracted from signals with small values might be ignored, all features were firstly processed by Z-score normalization. Principal Component Analysis (PCA) was then utilized to reduce the feature dimensions, as it has been demonstrated that PCA could perform much better than other techniques on training sets with small size [28]. To make reconstruction error less than 5%, we reserved 77 principal components as features after dimensionality reduction.

2.3.4. Feature Selection. To get the optimal performance of recognition models, we should find and remove useless features from the above 77 features. In this study, we used the sequential backward selection (SBS) to find the best subset of features that reduced the feature dimension while minimizing the performance loss of recognition model [29], and Algorithm 1 describes the whole process. SBS is a greedy search algorithm that starts from the whole feature set X and sequentially discards the feature x' so as to improve (or minimally worsens) the evaluation measure J. And it stops when the evaluation measure J is not increased or the subset X' is an empty set, which means that all remaining features are useful for the recognition model.

2.4. Model Training. We trained regression models for the recognition of Big Five personality traits. Since there is no evidence showing that a certain machine learning algorithm is the most suitable for identifying personality traits, we investigated the state-of-the-art regression models in this study: linear regression (LR) [30], support vector regression (SVR) [31], and Gaussian process regression (GPR) [32].

LR is a parameter model, whose parameters are estimated by minimizing the mean square error, and makes predictions requires simple matrix multiplication [30]. SVR is an extension of support vector classification, which first maps feature vectors to a higher-dimensional feature space using kernel trick and then makes predictions based only on support vectors [31]. In contrast to the above described algorithms, GPR is a nonparametric kernel-based probabilistic model, with the advantage of automatic tuning of the kernel parameters from the training data by maximizing log marginal likelihood [32].

In this study, we took the linear kernel function for SVR, and the kernel function of the dot-product kernel plus the white kernel for GPR. To evaluate the predictive performance of the models, we considered the root mean squared error (RMSE), the coefficient of determination (R^2), and the Pearson correlation coefficient (r) between predicted scores and self-reported scores of the respective personality traits. Denote C, Γ as a regression function and its corresponding parameters set and $f_i, i = 1, 2, \cdots, n$ as the ith sample's feature set. The RMSE, R^2 and r can be written as

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(C(f_i, \Gamma) - L_i \right)^2},$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(C(f_i, \Gamma) - L_i \right)^2}{\sum_{i=1}^{n} \left(C(f_i, \Gamma) - \overline{L} \right)^2},$$

$$r = \frac{\sum_{i=1}^{n} \left(C(f_i, \Gamma) - \overline{C(f, \Gamma)} \right) \left(L_i - \overline{L} \right)}{\sqrt{\sum_{i=1}^{n} \left(C(f_i, \Gamma) - \overline{C(f, \Gamma)} \right)^2} \sqrt{\sum_{i=1}^{n} \left(L_i - \overline{L} \right)^2}},$$

$$(5)$$

$C(f_i, \Gamma)$ outputs the predicted score from features f_i and L_i refers the true score of the ith sample. In this work, we applied 10-fold cross validation and averaged performance measures across all folds within a single prediction model.

3. Results

3.1. Demographics and Questionnaire Scores of BFI-44. Of these 92 participants (52 males and 40 females), their ages ranged from 21 to 56 years (mean = 31.84, SD = 7.03), and their driving experience ranged from 0.5 to 33 years (mean = 7.84, SD = 5.96). In terms of education level, the participants reported having the following levels: below university diploma, 2.17% ($n = 2$); university diploma, 51.09% ($n = 47$); and above university diploma, 46.74% ($n = 43$). Descriptive statistics of self-reported personality traits are provided in Table 1. Of the 92 participants who formed the study sample, the personality traits scores between two genders showed no significant difference (openness: $t = 0.62$, $p = 0.53$; conscientiousness: $t = 1.14$, $p = 0.26$; extraversion: $t = -1.16$, $p = 0.25$; agreeableness: $t = -0.08$, $p = 0.45$; neuroticism: $t = -1.38$, $p = 0.17$), which means that gender was not a factor which affects the performance of the Big Five personality traits recognition models in our date set.

3.2. The Recognition of Big Five Personality Traits. After feature selection, the remaining features were different according to regression algorithms. The number of remaining features for LR, SVR, and GPR was shown in Table 2.

The performance of the regression models is presented in Figure 6 and Table 3. The results showed that personality traits can be revealed through driving signals. Specifically, for the five dimensions of personality traits, the best performance occurred with SVR predicting openness (RMSE = 2.47, $R^2 = 0.74$, $r = 0.88$), followed by SVR predicting conscientiousness (RMSE = 2.94, $R^2 = 0.54$, $r = 0.79$), SVR predicting extraversion (RMSE=3.33, $R^2 = 0.45$, $r = 0.75$), SVR predicting agreeableness (RMSE = 3.48, $R^2 = 0.38$, $r = 0.73$), and LR predicting neuroticism (RMSE = 4.23, $R^2 = 0.19$, $r = 0.57$). Furthermore, our results indicated that the performances of different models varied. The results showed that the average

```
Input:
    X: The whole feature set
    J: Evaluation measure.
Output:
    X': The best subset of features.
X' = X,
repeat
    x' = arg max_{x∈X}{J(X' − x)},
    X' = X' − x',
until not improvement in J OR X' = ∅
```

ALGORITHM 1: Sequential backward selection.

TABLE 1: Descriptive statistics of personality variables.

	All		Male		Female	
	M	SD	M	SD	M	SD
Openness	36.61	5.50	36.92	5.31	36.20	5.64
Conscientiousness	34.35	4.81	34.85	4.47	33.70	5.09
Extraversion	27.77	5.29	27.21	5.25	28.50	5.18
Agreeableness	35.16	5.61	34.77	6.29	35.67	4.45
Neuroticism	19.21	5.33	18.54	5.14	20.08	5.38

TABLE 2: Number of remaining features after feature selection.

	LR	SVR	GPR
Openness	39	39	47
Conscientiousness	23	21	30
Extraversion	31	34	36
Agreeableness	37	45	32
Neuroticism	24	16	27

performance of the SVR model is better than the LR model and GPR model.

4. Discussion

We collected driving signals provided by in-vehicle sensors using CAN bus and trained machine learning models for identifying an individual's Big Five personality traits. Using the time-frequency analysis method, we extracted features from driving signals in the time and frequency domains, respectively, which were used to build personality traits recognition models. For the five personality trait dimensions, the coefficients of determination of the different models were between 0.19 and 0.74, the root mean squared errors were between 2.47 and 4.23, and the correlations between self-reported questionnaire scores and predicted scores were considered medium to strong (0.56–0.88). Our findings demonstrated that driving signals can be used to automatically identify individual personality traits in real-time.

Our results shown the driving signals are a convenient and objective source for measuring individual personality traits. As can be seen from our work, participants only need to drive for less than 10 km before their personality traits can be identified quite precisely. These results were consistent with previous studies showing an association between personality traits and driving behavior [2, 4, 6]. It is worth noting that the effective machine learning models in this current study were built based on low-level features in the time and frequency domains. The high-level features of driving behaviors in this field (e.g., lane switching, tailgating, overtaking, and speeding) are often based on subjective qualitative evaluations [33, 34], which limits the effectiveness of integrating these features into one machine learning model in practice. Although, time-frequency features may not provide much intuitive understanding of individual driving behaviors, they could provide more comprehensive information about driver's personality reflected in driving. Our results demonstrated the validity of building machine learning models to identify self-reported personality traits based on low-level features extracted using the time-frequency analysis.

Modern cars have recently become equipped with several hundred sensors and ECUs, which means we can easily obtain driving signals at minimal cost. Thus, this method to identify personality traits based on driving signals is suitable for the development in-car integration and single-chip embedded systems. Additionally, personalization in the automobile sector is a relatively recent trend to ensure optimal user experience in recent years [35]. Although personalization can be explicitly implemented by providing drivers with system parameters that can be manual tune, the
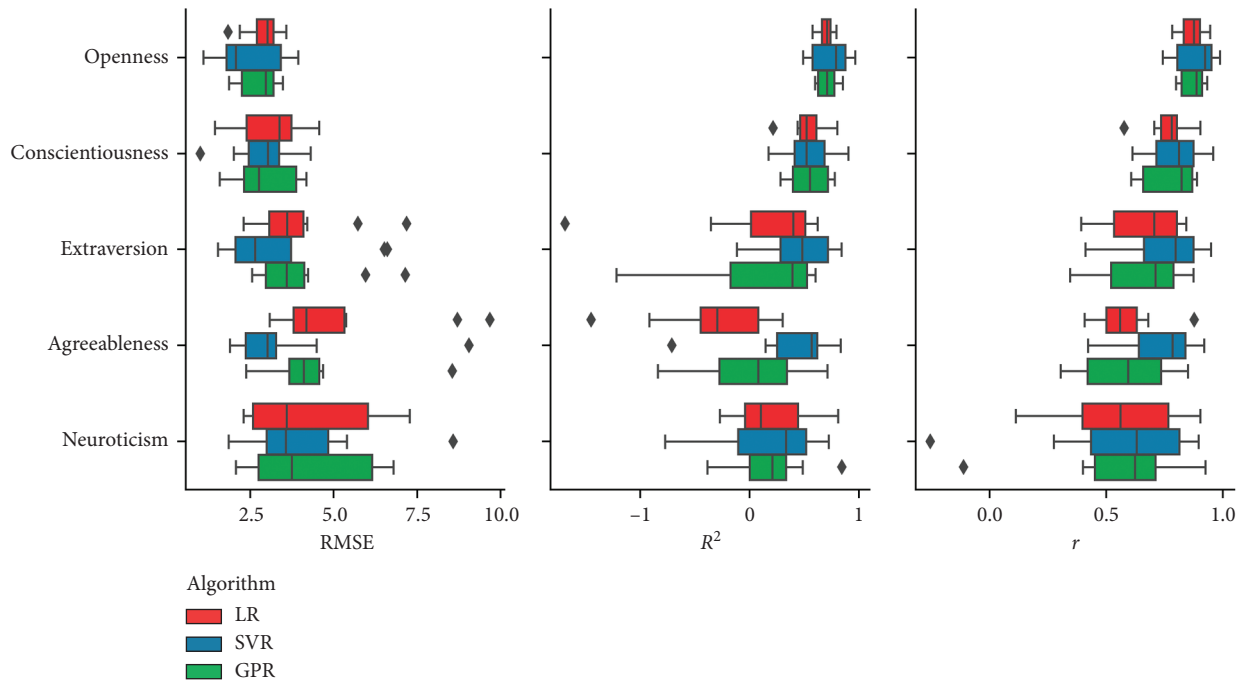
Figure 6: Box and whisker plot of prediction performance measures from 10-fold cross validation.

Table 3: The performance of the regression models.

	LR			SVR			GPR		
	RMSE	R^2	r	RMSE	R^2	r	RMSE	R^2	r
Openness	2.87	0.69	0.87	**2.47**	**0.74**	**0.88**	2.78	0.71	0.87
Conscientiousness	3.07	0.53	0.76	**2.94**	**0.54**	**0.79**	2.93	0.55	0.77
Extraversion	3.95	0.11	0.66	**3.33**	**0.45**	**0.75**	3.96	0.13	0.65
Agreeableness	5.16	−0.32	0.57	**3.48**	**0.38**	**0.73**	4.31	0.02	0.58
Neuroticism	**4.23**	**0.19**	**0.57**	4.12	0.19	0.56	4.26	0.20	0.56

implicit mode that estimates drivers preferences based on observing their behavior not only reduces the tedious and error-prone task of manual tuning, but also satisfies drivers' need for vehicle adaptation through fine-tuning [36]. For example, the "Intelligent Personal Assistant" (IPA) in vehicles is an important feature which offers a way for drivers to interact with their vehicles using their voice [37]. Identifying driver's personality traits by driving behavior and personalizing the IPA dynamically to the current driver will increase the customer experience. Therefore, this method may have potential value of the development of human-centered intelligent driving environments.

As a pilot study, it is appropriate to highlight several limitations. First, in this study personality traits were measured using self-report questionnaires. Although the validity of the questionnaires in accessing personality traits has been well supported in the literature [22], more criteria could be included in future studies. Second, this study's sample population comprised white-collar workers and was not sufficiently large. Therefore, the validity of our model in identifying self-reported personality traits cannot be equated with the effectiveness in populations of individuals with

different occupations, education levels, and cultures. Third, the current study built recognition models based on low-level features extracted using the time-frequency analysis, which cannot provide a clear understanding for the relationship between driving behavior and personality. Further research based on intuitively visible high-level features is necessary. Fourth, although our results showed the validity of identifying personality traits using this model, why the performance of models of personality traits in different dimensions is varied remains unclear. The disparity of the accuracies in identifying different dimensions implied that not all the personality-relevant could be equally reflected in driving. For a better understanding of how driving behavior reflects individual personality traits, more future works need to continue from two aspects: first, conducting more experiments, such as driving simulator experiments using fMRI technology [38]; second, explore the relationship between driving behavior and personality traits using more in-depth analysis, such as factor analysis.

Despite those limitations due to the exploratory nature of the study, it suggests the potential in future research on data-driven psychological measurement. Driving signals

have the advantages of being real-time, continuous, non-intrusive, and reliable [39], while requiring him/her finishing a questionnaire frequently and repeatedly is often not acceptable in practice; therefore, this method can measure personality traits in real-time and objectively, which cannot be achieved by a questionnaire. So our recognition model may show advantages in some cases, such as the driver is nonfixed but has a high demand for vehicle adaptation. Moreover, future research can transfer this method to the recognition of other psychological indicators in driving environment, because this method can monitor the continuous change of driver's psychological indicators. Additionally, although technological progress enables increasing automation in vehicles, the current general assumption for designing driving systems, such as driving assistance systems, is that drivers prefer to use systems that adopt a similar driving style to their own [8]. However, there is little empirical evidence to support this assumption. Thus, this method provides a new direction for the research on designing driving assistance systems.

5. Conclusions

This study moved one step forward toward a low-cost, nonintrusive solution for real-time identification of Big Five personality traits, which could be of potential value in the development of in-car integration. Our experiment demonstrated that driving signals provided by in-vehicle sensors using CAN bus can be an objective data source for measuring personality traits, and the predictive machine learning models showed effectiveness in identifying self-reported personality traits. Furthermore, this pilot study indicated a possible direction for further investigation on convenient psychometric methods and provided new perspectives for the development of intelligent driving environments from a human-centered perspective.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] L. Mallia, L. Lazuras, C. Violani, and F. Lucidi, "Crash risk and aberrant driving behaviors among bus drivers: the role of personality and attitudes towards traffic safety," *Accident Analysis & Prevention*, vol. 79, pp. 145–151, 2015.

[2] N. J. Starkey and R. B. Isler, "The role of executive function, personality and attitudes to risks in explaining self-reported driving behaviour in adolescent and adult male drivers," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 38, pp. 127–136, 2016.

[3] E. R. Dahlen, R. C. Martin, K. Ragan, and M. M. Kuhlman, "Driving anger, sensation seeking, impulsiveness, and boredom proneness in the prediction of unsafe driving," *Accident Analysis & Prevention*, vol. 37, no. 2, pp. 341–348, 2005.

[4] J. Yang, F. Du, W. Qu, Z. Gong, and X. Sun, "Effects of personality on risky driving behavior and accident involvement for Chinese drivers," *Traffic Injury Prevention*, vol. 14, no. 6, pp. 565–571, 2013.

[5] W. Arthur Jr and D. Doverspike, "Predicting motor vehicle crash involvement from a personality measure and a driving knowledge test," *Journal of Prevention & Intervention in the Community*, vol. 22, no. 1, pp. 35–42, 2001.

[6] B. Shen, W. Qu, Y. Ge, X. Sun, and K. Zhang, "The relationship between personalities and self-report positive driving behavior in a Chinese sample," *PLoS One*, vol. 13, no. 1, 2018.

[7] K. Bengler, K. Dietmayer, B. Farber, M. Maurer, C. Stiller, and H. Winner, "Three decades of driver assistance systems: review and future perspectives," *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 4, pp. 6–22, 2014.

[8] M. Hasenjager, M. Heckmann, and H. Wersing, "A survey of personalization for advanced driver assistance systems," *IEEE Transactions on Intelligent Vehicles*, vol. 5, 2019.

[9] O. P. John and S. Srivastava, "The Big Five trait taxonomy: history, measurement, and theoretical perspectives," *Handbook of Personality: Theory and Research*, vol. 2, pp. 102–138, 1999.

[10] S. Soldz and G. E. Vaillant, "The Big Five personality traits and the life course: a 45-year longitudinal study," *Journal of Research in Personality*, vol. 33, no. 2, pp. 208–232, 1999.

[11] G. Wu, F. Chen, X. Pan, M. Xu, and X. Zhu, "Using the visual intervention influence of pavement markings for rutting mitigation-part I: preliminary experiments and field tests," *International Journal of Pavement Engineering*, vol. 20, no. 6, pp. 734–746, 2019.

[12] F. Chen, H. Peng, X. Ma, J. Liang, W. Hao, and X. J. T. Pan, "Examining the safety of trucks under crosswind at bridge-tunnel section: a driving simulator study," *Tunnelling and Underground Space Technology*, vol. 92, Article ID 103034, 2019.

[13] U. Kiencke, S. Dais, and M. Litschel, "Automotive serial controller area network," *SAE Transactions*, vol. 95, no. 2, pp. 823–828, 1986.

[14] Z. Qi, Q. Shi, and H. Zhang, "Tuning of digital PID controllers using particle swarm optimization algorithm for a CAN-based DC motor subject to stochastic delays," vol. 67, no. 7, pp. 5637–5646, 2019.

[15] K. Jiang, H. Zhang, H. R. Karimi, J. Lin, and C. Systems, "Simultaneous input and state estimation for integrated motor-transmission systems in a controller area network environment via an adaptive unscented kalman filter," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, 2018.

[16] On Board Diagnostic Net, "Building adapter for vehicle onboard diagnostic," vol. 9, 2009, https://www.obddiag.net/adapter.html.

[17] A. E. B. El Masri, H. Artail, and H. Akkary, "Toward self-policing: detecting drunk driving behaviors through sampling CAN bus data," in *Proceedings of the 2017 International*

Conference on Electrical and Computing Technologies and Applications, pp. 1–5, Ras Al Khaimah, UAE, January 2018.

[18] A. B. Makar, K. E. McMartin, M. Palese, and T. R. Tephly, "Formate assay in body fluids: application in methanol poisoning," *Biochemical Medicine*, vol. 13, no. 2, pp. 117–126, 1975.

[19] V. Sadhu, T. Misu, and D. Pompili, "Deep multi-task learning for anomalous driving detection using CAN bus scalar sensor data," 2019, http://arxiv.org/abs/1907.00749.

[20] Z. E. A. Elassad, H. Mousannif, H. Al Moatassime, and A. Karkouch, "The application of machine learning techniques for driving behavior analysis: a conceptual framework and a systematic literature review," *Engineering Applications of Artificial Intelligence*, vol. 87, p. 103312, 2020.

[21] P. Wan, C. Wu, Y. Lin, and X. Ma, "Driving anger states detection based on incremental association markov blanket and least square support vector machine," *Discrete Dynamics in Nature and Society*, vol. 2019, Article ID 2745381, 2019.

[22] R. Carciofo, J. Yang, N. Song, F. Du, and K. Zhang, "Psychometric evaluation of Chinese-language 44-item and 10-item big five personality inventories, including correlations with chronotype, mindfulness and mind wandering," *PloS One*, vol. 11, no. 2, 2016.

[23] M. Enev, A. Takakuwa, K. Koscher, and T. Kohno, "Automobile driver fingerprinting," *Proceedings on Privacy Enhancing Technologies*, vol. 16, no. 2, pp. 34–50, 2016.

[24] T. Lajunen, J. Karola, and H. Summala, "Speed and acceleration as measures of driving style in young male drivers," *Perceptual and Motor Skills*, vol. 85, no. 1, pp. 3–16, 1997.

[25] B. D. Anderson and J. B. Moore, *Optimal Filtering*, Courier Corporation, North Chelmsford, MA, USA, 2012.

[26] G. E. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control, John Wiley and Sons*, NJ, USA, Hoboken, 2015.

[27] S. Weinstein and P. Ebert, "Data transmission by frequency-division multiplexing using the discrete Fourier transform," *IEEE Transactions on Communication Technology*, vol. 19, no. 5, pp. 628–634, 1971.

[28] A. M. Martínez and A. C. Kak, "Pca versus lda," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, 2001.

[29] I. A. Gheyas and L. S. Smith, "Feature subset selection in large dimensionality domains," *Pattern Recognition*, vol. 43, no. 1, pp. 5–13, 2010.

[30] G. A. Seber and A. J. Lee, *Linear Regression Analysis*, John Wiley & Sons, Hoboken, NJ, USA, 2012.

[31] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.

[32] C. E. Rasmussen, *Gaussian Processes in Machine Learning*, pp. 63–71, MIT Press, Cambridge, MA, USA, 2005.

[33] C. Atombo, C. Wu, M. Zhong, and H. Zhang, "Investigating the motivational factors influencing drivers intentions to unsafe driving behaviours: speeding and overtaking violations," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 43, pp. 104–121, 2016.

[34] R. Fernandes, J. Hatfield, and R. F. Soames Job, "A systematic investigation of the differential predictors for speeding, drink-driving, driving while fatigued, and not wearing a seat belt, among young drivers," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 13, no. 3, pp. 179–196, 2010.

[35] M. Hasenjäger, M. Heckmann, and H. Wersing, "A survey of personalization for advanced driver assistance systems," *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 2, pp. 335–344, 2020.

[36] H. Fan and M. S. Poole, "What is personalization? perspectives on the design and implementation of personalization in information systems," *Journal of Organizational Computing and Electronic Commerce*, vol. 16, no. 3-4, pp. 179–202, 2006.

[37] N. Horn, "Hey BMW, now we're talking! bmws are about to get a personality with the company's intelligent personal assistant," 2018, https://www.press.bmwgroup.com/global/article/attachment/T0284429EN/413869.

[38] P. Chen, F. Chen, L. Zhang, X. Ma, and X. J. T. Pan, "Examining the influence of decorated sidewaall in road tunnels using fMRI technology," *Tunnelling and Underground Space Technology*, vol. 99, p. 103362, 2020.

[39] Z. Li, S. Li, R. Li, B. Cheng, and J. Shi, "Online detection of driver fatigue using steering wheel angles for real driving conditions," *Sensors*, vol. 17, no. 3, p. 495, 2017.