

A Machine Learning Based Personalized System for Driving State Recognition

Dewei Yi^a, Jinya Su^{b,*}, Cunjia Liu^a, Mohammed Quddus^c, Wen-Hua Chen^a

^aDepartment of Aeronautical and Automotive Engineering, Loughborough University, Loughborough, LE11 3TU, U.K.

^bSchool of Computer Science and Electronic Engineering, University of Essex, Colchester, CO4 3SQ, U.K.

^cSchool of Architecture, Building and Civil Engineering, Loughborough University, Loughborough, LE11 3TU, U.K.

Abstract

Reliable driving state recognition (e.g. normal, drowsy, and aggressive) plays a significant role in improving road safety, driving experience and fuel efficiency. It lays the foundation for a number of advanced functions such as driver safety monitoring systems and adaptive driving assistance systems. In these applications, state recognition accuracy is of paramount importance to guarantee user acceptance. This paper is mainly focused on developing a personalized driving state recognition system by learning from non-intrusive, easily accessible vehicle related measurements and its validation using real-world driving data. Compared to conventional approaches, this paper first highlights the necessities of adopting a personalized system by analysing feature distribution of individual driver's data and all drivers' data via advanced data visualization and statistical analysis. If significant differences are identified, a dedicated personalized model is learnt to predict the driver's driving state. Spearman distance is also drawn to evaluate the differences between individual driver's data and all drivers' data in a quantitative manner. In addition, five categories of classifiers are tested and compared to identify a suitable one for classification, where random forest with Bayesian parameter optimization outperforms others and therefore is adopted in this paper. A recently collected dataset from real-world driving experiments is adopted to evaluate the proposed system. Comparative experimental results indicate that the personalized learning system with road information significantly outperforms conventional approaches without considering personalized characteristics or road information, where the overall accuracy increases from 81.3% to 91.6%. It is believed that the newly developed personalized learning system can find a wide range of applications where diverse behaviours exist.

Keywords: GPS, Driver state monitoring, Anomalous driving detection, Road safety

1. Introduction

Recent research interest in intelligent vehicles becomes more concentrated on how to enhance driving safety, convenience, and improve fuel efficiency (St-Aubin et al., 2015; Birrell et al., 2015; Jamson et al., 2015). To this end, a number of advanced functions have been developed such as Advanced Driver Assistance Systems (ADAS) (e.g. adaptive cruise control, collision avoidance systems), In-Vehicle Information Systems (IVIS) (e.g. navigation and entertainment systems) (Ziegler et al., 2011) and “economic driving” systems (Vaiana et al., 2014). Although these functions are to improve driving experience, it should be also noted that inappropriate use of these advanced functions may create problems such as increasing driver's distraction and consequently annoying the driver or even increasing the risk of traffic accidents (Ziegler et al., 2011; Yi et al., 2019a).

To mitigate the adverse effects so that these functions are more acceptable and user-friendly, there is a trend towards developing “adaptive aiding” systems (Zhu et al., 2019; Ziegler et al., 2011;

*Corresponding author

Email addresses: d.yi@lboro.ac.uk (Dewei Yi), sujinya1990@163.com (Jinya Su), c.liu5@lboro.ac.uk (Cunjia Liu), m.a.quddus@lboro.ac.uk (Mohammed Quddus), w.chen@lboro.ac.uk (Wen-Hua Chen)

Hajek et al., 2013). In this approach, necessary aid is supplied to the driver at the right time and in an appropriate manner; this is achieved by monitoring the driver’s driving states (e.g. normal, drowsy or aggressive) in real time and adapting ADAS and IVIS systems accordingly (Hu et al., 2017; Meiring and Myburgh, 2015). For instance, an earlier and more clear collision warning signal could be generated to alert a driver under drowsy state compared to normal one in order to enhance driving safety. An incoming phone call could be diverted to voice mail when driver workload is identified to be higher than usual. Moreover, driving behaviours are closely related to fuel efficiency as it is reported in (Xiong et al., 2019; Vaiana et al., 2014) that the difference in fuel consumption between normal and aggressive driving (e.g. sudden acceleration, rapid braking and harsh cornering) is estimated to be as high as 40%. Aggressive driving can be warned in “economic driving” systems to improve fuel efficiency (Vaiana et al., 2014). In all these applications, it is evident that driving state recognition performance is of paramount importance.

A large number of studies is available on driving state recognition (readers are referred to recent survey papers (Meiring and Myburgh, 2015; Wahlström et al., 2017) among others). In these algorithms, driving state is recognized by analysing large volume of driving data collected by various sensors including exteroceptive sensors (e.g. Global Navigation Satellite Systems (GNSS), magnetometers, cameras), proprioceptive sensors (e.g. accelerometers, gyroscopes) and complementary sensors such as vehicle’s On-Board Diagnostics (OBD). Based on the algorithms adopted, they can be broadly divided into two categories namely heuristic rule based approaches (Castignani et al., 2015; Saiprasert et al., 2017; Joubert et al., 2016) and machine learning based approaches (Hong et al., 2014; Chandaka et al., 2009; Woo and Kulić, 2016; Wu et al., 2016). In (Saiprasert et al., 2017), hard rules and fuzzy logic are adopted to recognize different driving behaviours (e.g. hard acceleration/braking and aggressive steering) by using yaw rate, jerk, speed and bearing signals respectively. This approach is relatively intuitive, however, it is generally difficult to choose an appropriate threshold. Recent work focuses more on different machine learning algorithms such as Support Vector Machines (SVM) (Chandaka et al., 2009; Woo and Kulić, 2016), decision tree (Wu et al., 2016) and neural network (Zhang et al., 2014). In these approaches, rather than using heuristic rules, driving state recognition is achieved by learning from training dataset via classification algorithms.

1.1. Related work

Drivers may exhibit different driving states (or styles) under various driving situations such as normal (or safe), drowsy, aggressive or even drunk driving (Hu et al., 2017; Meiring and Myburgh, 2015; Masala and Grosso, 2014). Normal driving is a reference driving state from which deviation (such as drowsy or aggressive driving) can be identified, and to which a driver should adapt his behaviour. Drowsy driving appears when drivers suffer from fatigue, which is usually followed by delayed reactions and responses (Dong et al., 2011). It was reported in (Hailin et al., 2010) that 15%–20% of all vehicle accidents are estimated to be related to sleepiness and that fatigue driving increases the risk of an accident by four to six times. Aggressive driving is usually caused by driver impatience, annoyance, or an attempt to minimise travelling time, which is usually associated with irregular and abrupt changes in vehicle speed, quick changes in vehicle lateral position, and excessive acceleration or deceleration. This type of driving behaviour is also closely related to road accident, and also results in unnecessary fuel consumption than normal driving (Vaiana et al., 2014).

Real-time monitoring driver’s behaviours and providing adaptive assistance (e.g. driver assistance systems) or timely feedback (e.g. economic driving systems) is an effective solution to improve road safety, enhance driving experience and increase fuel efficiency (Kantarachos et al., 2018; Meiring and Myburgh, 2015). The commonly used approach for driving state recognition is to analyse the Vehicle Related Measurements (VRM), which includes speed, three-axis acceleration and orientation angles via GPS and IMU, throttle position and airflow rate via OBD system (Kantarachos et al., 2018; Sun et al., 2016). This is due to the fact that this approach is non-intrusive, easily accessible and has a low requirement on working condition (Yi et al., 2019a,b) in comparison with physiological sensors

based approaches (Zhang et al., 2015) or computer vision (via cameras) based approaches (You et al., 2012). In the past decade, studies on VRM based monitoring system have been conducted due to its engineering importance. These systems can be broadly divided into in-vehicle data recorder (e.g. OBD system) based and smartphone based according to different sensors adopted for data acquisition (Saiprasert et al., 2017). They can also be categorised into heuristic rule based (Castignani et al., 2015; Saiprasert et al., 2017; Joubert et al., 2016) and machine learning based approaches (Hong et al., 2014; Chandaka et al., 2009; Woo and Kulić, 2016; Wu et al., 2016) according to different algorithms adopted for classification. As discussed in Introduction, machine learning based approaches are drawing increasing attention due to less expert knowledge involved and high performance compared to rule based approaches (Saiprasert et al., 2017). The flowchart for common VRM based driving state recognition systems is depicted in Fig 1.

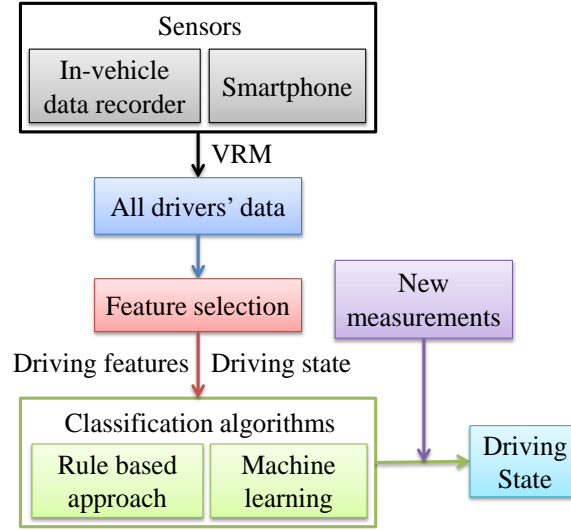


Figure 1: Flowchart for conventional generic driving state recognition system.

As shown in Fig 1, the driving state recognition system starts with driving data acquisition by using different sensors. In conventional driving state recognition systems, different drivers’ driving data is aggregated indiscriminately. On this basis, features are selected based on either expert knowledge or different feature selection algorithms (Guyon and Elisseeff, 2003). In supervised learning, the feature data for the purpose of algorithm training should be labelled appropriately, which corresponds to particular driving states in this paper. With given training data (i.e. feature data with class labels), different models can be applied to achieve the task of classification by using various algorithms. Then with the arrival of new feature measurements, the driving state can be identified via the trained classifier.

It should be highlighted that the aforementioned approaches train a generic (or average) model using the driving data of existing drivers indiscriminately, and consequently may overlook the personalized driving characteristics of individual drivers (Yi et al., 2019a; Butakov and Ioannou, 2015; Schnelle et al., 2017). In practice, different drivers may have distinct driving characteristics and preferences, and as a result may experience different driving states even for a similar driving situation. **Therefore, it is not surprising that conventional generic/average state recognition models may lead to a degraded performance for a particular driver (Zhang et al., 2019; Yi et al., 2019a; Hu et al., 2017).** To overcome this issue, this paper proposes a “personalized” driver state recognition system accommodating personalized driving characteristics (*hidden in the driving data*) of individual drivers and so leading to a personalized driving state recognition system dedicated to a particular driver. Moreover, contextual information regarding the driver, the vehicle, and the environment may also affect the driving behaviours (Ji et al., 2004; Al-Sultan et al., 2013). Following this line of thought and without

loss of generality, the road type information available in the experimental dataset is further considered as an illustrating example to further improve the recognition performance.

The proposed personalized learning system is evaluated by using the existing real-world driving dataset collected in (Romera et al., 2016), where five drivers of various background are required to drive under three different behaviours (i.e. normal, drowsy and aggressive) on two types of roads (i.e. motorway and secondary road) for groundtruth data collection. Naturalistic driving data of about 430 minutes including speed, orientation and three-axis acceleration are collected by using a cellphone application entitled “DriveSafe” (Romera et al., 2016). In this work, comparative experiments are conducted on different algorithms, where the results indicate that incorporating personalized characteristics and road type information can substantially improve the algorithm performance. To the best of the authors’ knowledge, this work is the first attempt to apply the concept of “personalized learning” into driver state recognition along with real-world experimental data validation. To be more exact, the contributions are summarized as follows:

- (i) A personalized learning system is proposed for driving state recognition, where both individual driver’s characteristics and road information are accommodated;
- (ii) State-of-the-art feature analysis algorithms such as t-Distributed Stochastic Neighbour Embedding (t-SNE) are adopted to demonstrate the distinct driving characteristics of different drivers, necessitating the “personalized learning”;
- (iii) Five categories of classification algorithms are tested and compared to identify the most suitable one for driving state recognition, where random forest with Bayesian hyperparameter optimization outperforms others;
- (iv) The newly proposed framework is validated by using an existing dataset from real-world driving experiments; comparative experimental results demonstrate that incorporating personalized driving characteristics and road information substantially improves performance with overall accuracy increasing from 81.3% to 91.6%.

The remainder of this paper is organized as follows. Section 2 formulates a problem for driving state recognition. Section 3 provides theoretical background of this study, where methods of data visualization, different classifiers, the approach of parameters optimization, and models are introduced. In Section 4, comparative experiments are conducted to validate the proposed framework by using real-world naturalistic driving dataset (i.e. UAH-DriveSet dataset). It demonstrates the construction of the personalized driving state recognition system, results of data visualization, classifiers comparison, and the personalized model with considering road information. In Section. 5, the primary outcomes and limitations of the study are discussion. Section 6 concludes the paper.

2. Problem formulation

The primary objective of this paper is to build a personalized driving state recognition system that considers both personalized driving characteristics and road information. The driver state is considered as the output of the system. The measurements from different sensors (e.g. GPS, IMU, gyroscope) are synchronized once per second, with discrete time instants $t \in 1, \dots, T$, into feature vector $X_t = [x_t^1, x_t^2, x_t^3, x_t^4, x_t^5, x_t^6, x_t^7, x_t^8]^T$, where.

- x_t^1 and x_t^2 are obtained by GPS, where x_t^1 is a discrete variable representing road types (motorway or secondary road in this paper). x_t^2 is a continuous variable representing vehicle speed [km/h].
- x_t^3, x_t^4 , and x_t^5 are obtained by IMU and all of them are continuous variables. In particular, x_t^3, x_t^4, x_t^5 denote the vertical, lateral and longitudinal acceleration [Gs], respectively.

- x_t^6 , x_t^7 , and x_t^8 are obtained by gyroscope and all of them are continuous variables. In particular, x_t^6 , x_t^7 and x_t^8 represent the roll angle, pitch angle, and yaw angle [degrees], respectively.

The system behaviours change according to the driving state of a specific driver. A driver usually has several different states (e.g. normal, drowsy, and aggressive) during driving. Thus, the problem of driving state recognition can be transformed into a classification problem. However, different drivers have different driving characteristics and driving preferences. Moreover, the driving state is also affected by driving condition (e.g. road type). With a consideration of the above problems, the task of accurate driving state recognition is formulated as follows:

- Analyse feature distribution between all drivers' and a specific driver's data. When the distribution of all drivers' data is significantly different from a specific driver's data, a personalized model is needed and very likely to outperform a conventional generic model.
- Select an appropriate classifier to recognize driving state. Classifier selection is very challenging, therefore, several types of classifiers should be compared to identify the most suitable one.
- It is also tricky to tune the hyperparameters of classifiers. In order to address this problem, the parameters of classifiers should be optimized in an automatic manner.
- After the most suitable classifier is selected, a personalized model without road information, a generic model with road information (i.e. road-aware model) and a personalized model with road information should be built based on the selected classifier. Moreover, the three models are compared against a generic model without road information to validate their performance.

3. Theoretical background

In this paper, theoretical background involves in four parts: method of data visualization, classifier selectin, parameters optimization, and models. 1) Different methods of data visualization are applied to identify the difference between individual driver's feature distribution and all drivers' feature distribution. It is to determine the necessity of constructing the personalized model. 2) It is challenging to determine an appropriate classifier for a specific application. To this end, several categories of classifiers and their variants are compared to identify a suitable classifier for driving state recognition. 3) The parameters of classifiers should be optimized before having a comparison. 4) Based on the most suitable optimized classifier, the proposed personalized model and personalized model with road information are compared with generic model and road-aware model.

3.1. Methods of data visualization

To analyse the driving characteristics hidden in individual driving data, data visualization and analysis is first conducted to investigate the distinct driving behaviours among different drivers so as to demonstrate the necessities of personalized state recognition. Without the loss of generality, the distribution of typical features and the distribution of entire features are visualised and analysed by using histograms along with statistical analysis. **It is generally difficult to analyze high-dimensional feature data (with a dimension of seven in this work) by using histograms.** Instead, t-SNE algorithm is used to reduce the dimensionality of the entire feature distribution so that the entire feature distribution can be visualised and analysed by using histograms along with statistical analysis.

For **typical feature visualization**, data analysis between different drivers is conducted on individual features such as speed, acceleration. For one-dimensional random variable, histograms are an effective way to graphically represent the distribution of numeric data (i.e. probability density function). In this approach, the entire range of values are first divided into a series of intervals and then the number of values (or "relative" frequencies) that fall into each interval is counted. The differences

between histograms can visually be described by their patterns such as symmetric, skewed left/right, unimodal/bimodal/multimodal. Statistics can also be drawn accordingly. In this work, typical feature histograms are plotted under various driving states for different drivers. Individual feature analysis is effective, however, may overlook the correlation between different features. Therefore, the analysis of entire feature distribution is also necessary.

For **entire feature visualization**, t-SNE (Maaten and Hinton, 2008) is adopted to make an intuitive understanding. t-SNE is a powerful dimensionality reduction algorithm, which is used for high-dimensional feature visualization. t-SNE can keep the relative distance of samples. It is achieved by minimizing the Kullback-Leibler divergence between the distribution of measuring pairwise similarities of the input objects and the distribution of measuring pairwise similarities of the corresponding low-dimensional points in the embedding. For the same of completeness, t-SNE is briefly introduced as follows.

Given high-dimensional input objects $D = \{x_1, \dots, x_N\}$ and a distance function $d(x_i, x_j)$, it is to learn an s -dimensional embedding where every object is a point. An s -dimensional embedding set is represented by $\varepsilon = \{y_1, \dots, y_N\}$ with $y_i \in \mathbb{R}^s$. Therefore, the joint probabilities of t-SNE is defined by p_{ij} , which measure the pairwise similarity between objects x_i and x_j , given by

$$p_{j|i} = \frac{e^{-\frac{d(x_i, x_j)^2}{2\sigma_i^2}}}{\sum_{k \neq i}^N e^{-\frac{d(x_i, x_k)^2}{2\sigma_i^2}}}, \quad p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N},$$

where $p_{i|i} = 0$, σ_i is the bandwidth of Gaussian kernels, which is set so that the perplexity of the conditional distribution P_i equals to a predefined perplexity u . In the s -dimensional embedding ε , the similarities between y_i and y_j are measured using a normalized heavy-tailed kernel. A normalized Student-t kernel with a single degree of freedom is drawn to calculate the embedding similarity q_{ij} with $q_{ii} = 0$ between y_i and y_j , which is defined by

$$q_{i|j} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq i}^N (1 + \|y_k - y_i\|^2)^{-1}}.$$

Dissimilar input objects x_i and x_j can be modelled by low-dimensional counterparts y_i and y_j due to the heavy tails of the normalized Student-t kernel. The locations of the embedding points y_i are determined by minimizing the Kullback-Leibler divergence between the joint distributions P and Q , given by

$$C(\varepsilon) = KL(P \parallel Q) = \sum_{j \neq i}^N p_{ij} \log \frac{p_{ij}}{q_{ij}}.$$

3.2. Classifier selection

In machine learning applications, it is generally not easy to select an appropriate algorithm for the task of interest. Therefore, in this paper, various categories of classification algorithms are first evaluated and compared by using all driver's data so that the most suitable one for the task of driving state recognition is identified. The algorithms tested in this paper contain five categories: discriminant analysis, decision tree, k-nearest-neighbour, support vector machine and ensemble learning; since these algorithms have been previously applied to similar classification tasks (Meiring and Myburgh, 2015; Wahlström et al., 2017).

3.2.1. Discriminant Analysis

Discriminant analysis assumes classes generated from different Gaussian distributions (Ba et al., 2017). Based on this assumption, data are classified into different classes. **There are two variants of**

discriminant analysis including linear discriminant analysis and quadratic discriminant analysis. In linear discriminant analysis, the means of different Gaussian distributions are vary and the variances of different Gaussian distribution are the same. In quadratic discriminant analysis, the means and variances of different Gaussian distributions are both different.

3.2.2. Decision Tree

Decision tree is nonparameteric method, which does not need to presume the distribution of a given dataset (Pitombo et al., 2017). In decision tree, the most effective feature is selected to partition data from root node until the certain conditions are met. According to the differences of tree structure, decision tree can be further divided into simple, medium, and complex decision trees.

3.2.3. K-Nearest-Neighbours

K-nearest-neighbours (KNN) is a model-free classification algorithm, which does not need to pre-determine model specification (Yu et al., 2011). KNN determines the category of a given sample by using the closest observations in the feature space. According to different ways for calculating distance between a given sample and its neighbours in feature space, several variants of KNN are proposed such as fine, quadratic, coarse, cosine, cubic, and weighted KNN algorithms.

3.2.4. Support Vector Machines

Support Vector Machine (SVM) implements classification by discovering the maximum margin boundary between the positive data and negative data and then this boundary is used to decide the category of data (Yuan and Cheu, 2003). To reduce the computation load, training vector (termed ‘‘Support Vectors’’) are extracted to construct the classification boundary. The training vectors are the data lying closest to the classification boundary. To fit into different applications, differnt kernel functions should be embedded in the SVM such as linear, quadratic, cubic, fine Gaussian, medium Gaussian, coarse Gaussian kernel.

3.2.5. Ensemble Learning

Ensemble learning is to combine several classifiers together to make a group decision (Chen et al., 2017). Each classifier votes for the prediction result. The final decision relies on the vote result and weight of each classifier. This learning mechanism can improve accuracy, generalizability, and robustness. Several popular ensemble learning models are implemented and tested in this paper including boosted trees, random forest, subspace discriminant, subspace KNN, and RUSBoot trees.

3.3. Parameters optimization

Tuning the parameters of a classifier is a tricky job. In order to find out the suitable values of parameters, Bayesian optimization algorithm is used to achieve this goal. The purpose of Bayesian optimization is to minimize an objective function $f(x)$, which is the cross-validation loss of a classifier in this work. To achieve a minimum classification loss, three key elements are involved: a Gaussian process model of $f(x)$, a Bayesian update strategy and an acquisition function $a(x)$. A Gaussian process model is used to generate prior distribution of $f(x)$ and then it is updated at each new evaluation. In order to obtain a prior distribution of $f(x)$, kernel function of Gaussian process model should be pre-defined. Here, *Matern 5/2 kernel* is adopted due to its good performance in spatial statistics, machine learning and image analysis (Snoek et al., 2012) and *Matern 5/2 kernel* is given by:

$$k(x_i, x_j) = \sigma_f^2 \left(1 + \frac{\sqrt{5}r}{\sigma_l} + \frac{5r^2}{3\sigma_l^2} \right) e^{-\frac{\sqrt{5}r}{\sigma_l}}$$

where $r = \sqrt{(x_i - x_j)^T(x_i - x_j)}$ is the Euclidean distance between x_i and x_j .

Bayesian update strategy is to revise the Gaussian process model according to new evaluation of $f(x)$. Then, the posterior distribution over $gp(f|x_i, y_i) \ i \in \mathbb{N}^+$ is obtained. While acquisition function is used to determine the next evaluation point x based on a posterior distribution, denoted by:

$$a(x) = E[\max(0, \mu(x_{best}) - f(x))]$$

where x_{best} is the point achieving the minimum posterior mean and $\mu(x_{best})$ is the minimum value of the posterior mean. Readers are referred to (Gelbart et al., 2014; Bull, 2011) for more details.

3.4. Models

After a suitable classifier with parameter optimization is identified, four models are introduced including the generic model without road information, the proposed personalized model, road-aware model (i.e. the generic model with road information), and the proposed personalized model with road information. To be more clear, the four models in this work are detailed as follow.

3.4.1. Generic model

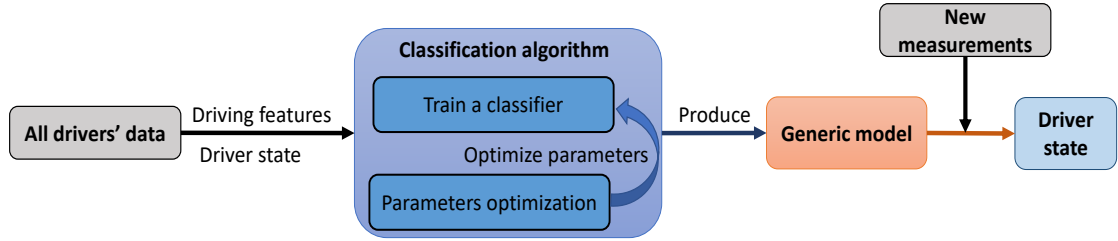


Figure 2: Generic model for driving state recognition.

Generic/average model refers to the algorithm that aggregates and learns from all drivers' data indiscriminately and is directly applied to all drivers. Generic model is designed based on generic driver behaviours, which provides reasonable performance for all drivers but may lead to a degraded performance for a particular driver. Fig. 2 describes the procedure of training a generic model.

3.4.2. Personalized model

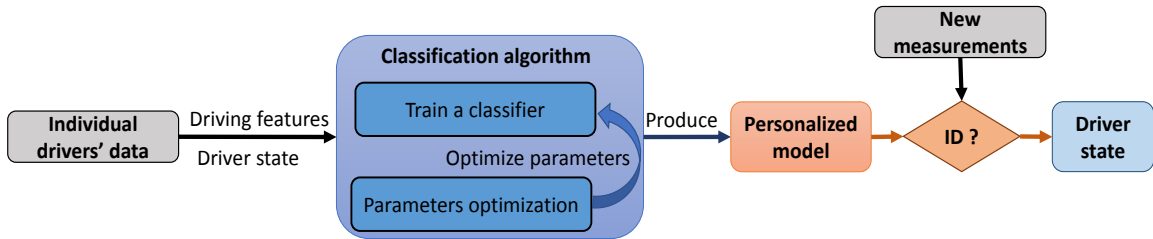


Figure 3: Personalized model for driving state recognition (where, ID is to identify a specific driver).

Personalized model refers to the algorithm dedicated to a particular driver accommodating the personalized driving characteristics. Personalized model focuses on individual driving behaviours rather than generic driving behaviours and is trained by using personal data. For each driver, a model is trained and then the corresponding model is used to recognize his/her driving state. The procedure of building personalized model is presented in Fig. 3.

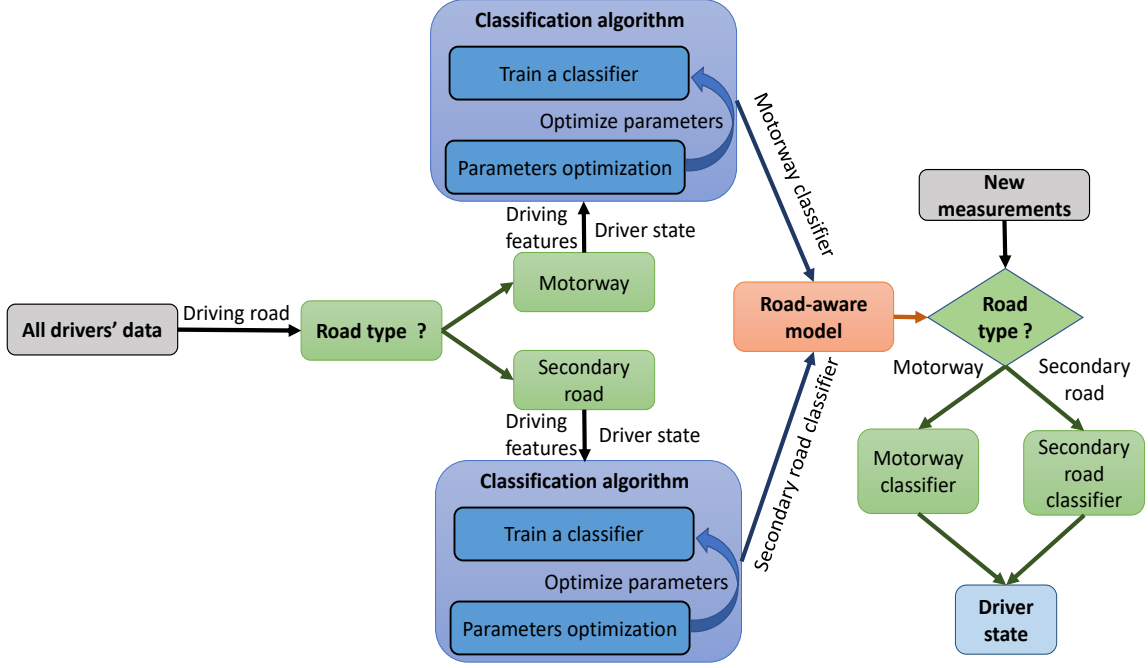


Figure 4: Road-aware model for driving state recognition.

3.4.3. Road-aware model

Contextual information regarding the driver, the vehicle, and the driving environment also affects the driving performance (Ji et al., 2004, 2006; Al-Sultan et al., 2013). For example, the driving data under daytime and nighttime or rainy and shiny condition usually have significant differences (Konstantopoulos et al., 2010). Conventional approaches accommodating this type of information usually treat the context as new features or conduct statistical reasoning using Bayesian network (Ji et al., 2004; Al-Sultan et al., 2013). The substantially increased features may require a large amount of dataset or rely on prior probability information, which is inhibitive in our scenario.

Considering that this research is mainly focused on personalized driving state recognition, only road type information available in the UAH-DriveSet data is fused into the personalized learning framework to further improve its performance. This is achieved by dividing the training data into different categories, so that road-aware classifiers are trained. It should be noted that although this approach is simple, it is effective in the task of driving state recognition (see, Section 4.4.2 for comparative results against generic model without contextual information). Fig. 4 presents the generic model with road information in this paper.

3.4.4. Personalized model with road information

Personalized model with road information is to consider personalized driving characteristics and road information simultaneously. This model is built hierarchically because even for the same driver, the driver reflects diverse driving behaviours in different traffic condition (e.g. road type). In this model, two models are trained to recognize driver state and both of them are based on personalized driving preferences. The first one is used to recognize driver state when driving on the motorway. The second one is applied to recognize driver state when driving on the secondary road. The procedure of deriving personalized model with road information is demonstrated in Fig. 5.

4. Proposed personalized learning system

In this section, the development process of the proposed personalized learning system for driving state recognition is elaborated. Vehicle-related measurements of drivers are used in classifier selection

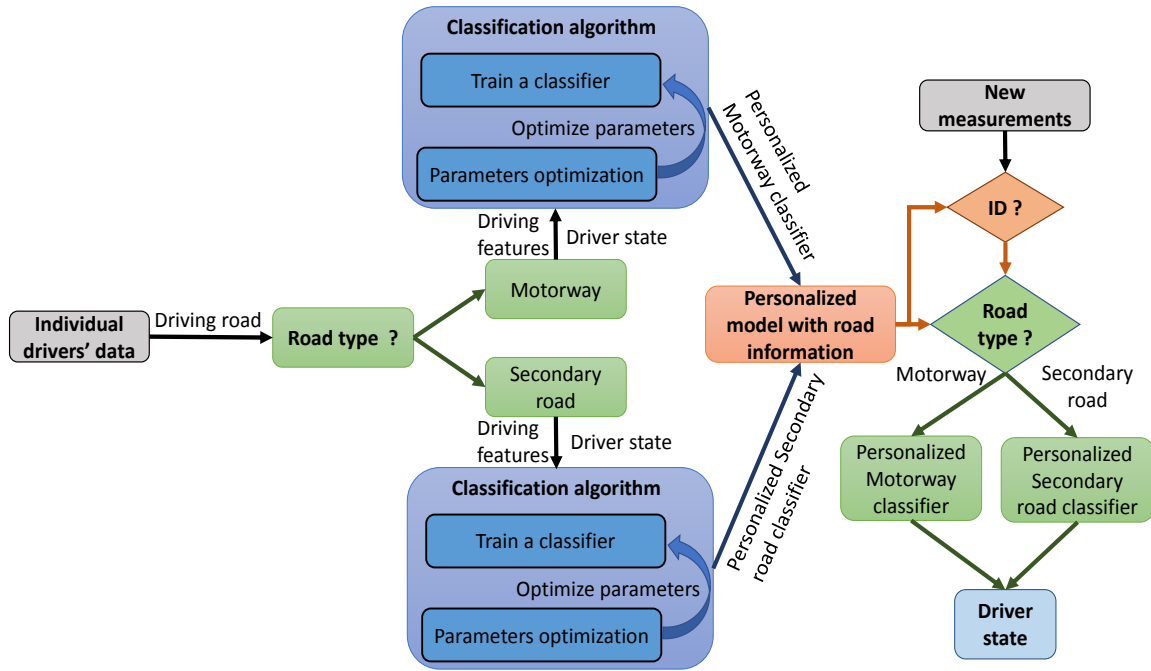


Figure 5: Personalized model with road information for driving state recognition (ID is to identify a specific driver).

and visualization analysis. In classifier selection, several different categories of classifiers and their variants are implemented, where their hyperparameters are tuned by Bayesian optimization algorithm. As a result, the most suitable classifier can be identified for driving state recognition. In visualization analysis, two data visualization methods are adopted to analyse the differences between individual driver's and all drivers' data. The first method is to analyse a typical feature such as speed or yaw angle by using a histogram. The second method is to analyse the entire features. Considering that the entire features are of a high-dimension (seven-dimensional), t-SNE is used to reduce dimensionality from seven to two so that their differences can be identified visually and by statistical measures. Then the similarity between a specific driver's and all drivers' data is evaluated in a quantitative manner. The visualization analysis and similarity computation are to demonstrate the necessity of constructing a personalized model. The philosophy is that if remarkable differences are identified (i.e. similarity is higher than the threshold), a personalized model is very likely to outperform a generic model. Moreover, road information is accommodated into the personalized model by considering that a driver's behaviours are affected by driving condition (e.g. road type). After a personalized model with road information is constructed, the model can be used to recognize the driving state for a specific driver based on incoming new measurements. To highlight the differences from conventional generic systems, a flowchart is presented in Fig 6, where the blocks in orange with dashed lines are the new blocks compared to a generic system as shown in Fig 1. It should be noted that the block of visualization and analysis is to demonstrate the necessities of personalized state recognition; in practical applications, this step can be automatically processed by the system without the involvement of human' visual inspection. The main differences between conventional generic learning system and the proposed personalized learning system with road information are summarized below.

- Personalized driving characteristic analysis: state-of-the-art data analysis is conducted to investigate the distinct driving behaviours among different drivers. If significant differences are identified in terms of feature distribution (e.g. mean, variation), personalized model is adopted where a dedicated model is trained for the particular driver so that personalized driving characteristics hidden in the personalized driving data are accommodated. Otherwise, the conventional generic model for all drivers is retained;

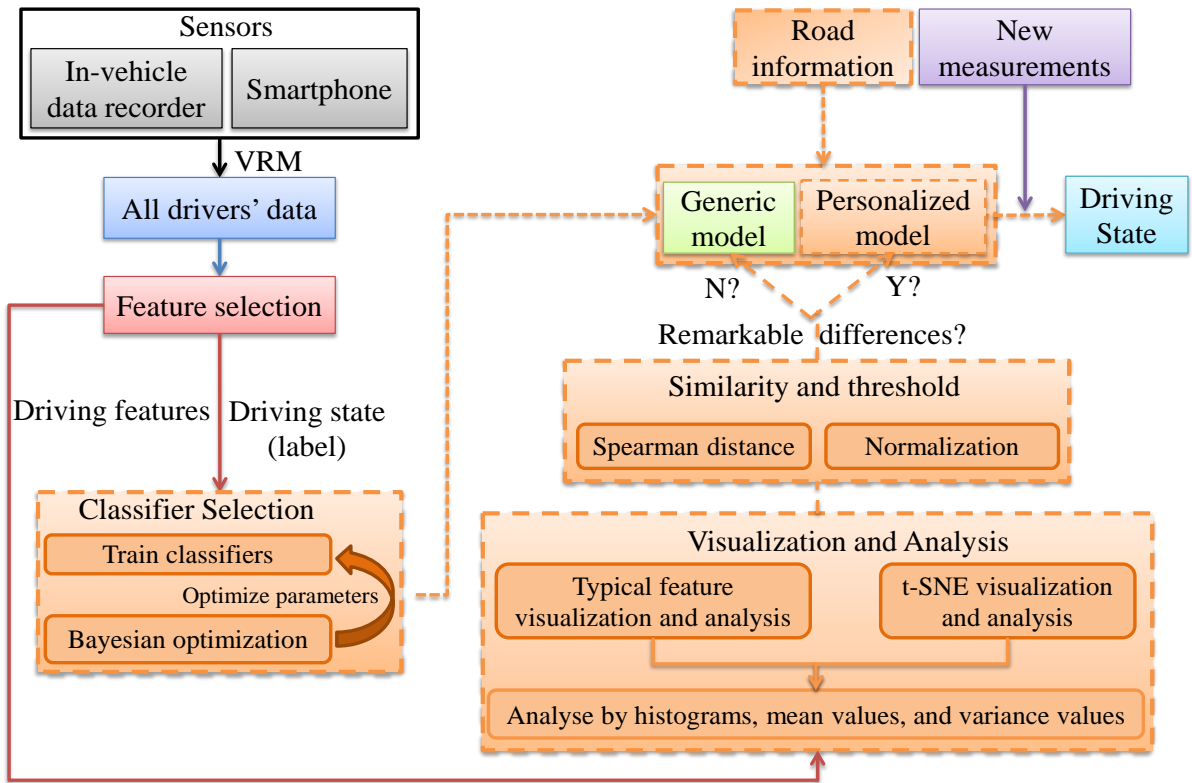


Figure 6: Flowchart of the proposed personalized learning system with road information, where the blocks in orange with dashed lines are new components and conducted offline compared to the conventional one in Fig 1.

- Road information available in the experimental dataset is fused into the framework to further enhance the performance, since it can provide useful background information regarding the driving environment.
- Five categories of classification algorithms are implemented and compared to identify the most suitable one for the task of driver state recognition.

In the following parts, different components of the proposed system are elaborated.

4.1. Real-world driving dataset

The existing real-world UAH-DriveSet dataset is adopted to validate the algorithms in this paper, which was collected in 2016 using a smartphone app DriveSafe for the purpose of deep driving behaviour analysis (Romera et al., 2016). In this dataset, six different drivers are involved where each driver is required to perform three different behaviours (i.e. normal, drowsy and aggressive) on two types of roads (i.e. motorway and secondary road) as shown in Fig 7.

Under the careful experiment setup in (Romera et al., 2016), the datasets were automatically labelled (i.e. groundtruth data). Regarding data labelling, each participant was required by the instructors to repeat pre-designated routes under normal, drowsy, and aggressive driving states, and therefore the data was automatically labelled. More than 420 minutes of naturalistic driving data along with video footage of the trips are collected. After an inspection it is discovered that the data of the sixth driver is incomplete, where aggressive driving data on secondary road is missing. Therefore, data for the first five drivers are adopted, which approximately contains 420 minutes of naturalistic driving. The raw dataset was processed by the authors in (Romera et al., 2016) so that all variables are synchronized into the frequency of 1 Hz (i.e. the frequency of GPS for speed measurement). It should be noted, however, that the proposed method is applicable to data with different frequencies.



Figure 7: Routes in the motorway test (a) and secondary road tests (b) (Romera et al., 2016).

The 420 minutes are the total length taken by five drivers under three different driving states on two types of roads. Each driver took six trips in total. For each driving state (i.e. normal, drowsy, and aggressive), a driver took two trips. One is driving on the motorway, and the other is on the secondary road. The driving duration and distance of each driver under different driving states are summarized in Table. 1.

Table 1: Driving duration and distance for five driver under different driving state

State	Normal		Drowsy		Aggressive	
	Motorway	Secondary Road	Motorway	Secondary Road	Motorway	Secondary Road
Trip Info.	Duration (Distance)					
D1	14 min. (25 km)	10 min. (16 km)	15 min. (25 km)	8 min. (13 km)	12 min. (24 km)	8 min. (16 km)
D2	15 min. (26 km)	10 min. (16 km)	15 min. (25 km)	10 min. (16 km)	14 min. (26 km)	10 min. (16 km)
D3	15 min. (26 km)	11 min. (16 km)	16 min. (26 km)	10 min. (17 km)	13 min. (26 km)	11 min. (16 km)
D4	16 min. (25 km)	11 min. (16 km)	17 min. (25 km)	11 min. (17 km)	15 min. (25 km)	10 min. (16 km)
D5	15 min. (25 km)	11 min. (16 km)	18 min. (25 km)	11 min. (16 km)	13 min. (25 km)	7 min. (12 km)

4.2. Driving state model specification

It is noted that the existing dataset in (Romera et al., 2016) (see, Section 4.1) is drawn to validate the algorithms in this work. In this dataset, smartphone is preferred for data collection due to the fact that it is easily accessible, widely available and of low cost (Wahlström et al., 2017; Saiprasert et al., 2017). In a modern smartphone, different sensors are available such as cellular, GPS. Three sensors of smartphone are considered in the dataset including GPS (i.e. speed), three-axis accelerometer (i.e. lateral, longitudinal and vertical accelerations) and gyroscope (i.e. yaw, pitch and roll angle).

With raw sensory data, data preprocessing is first conducted to unify the data length via data resampling (due to different sampling frequencies of various sensors) and attenuate the effect of sensor noise (see, Section 4.1 for more information). Speed, three-axis accelerations and orientation angles are chosen as the feature vector in this work, which are as defined below

$$\text{Feature} = [\text{Speed}, 3\text{-axis Accelerations}, \text{Orientations}]. \quad (1)$$

Since these signals have been proved to be effective for modelling driving behaviours under various conditions (Meiring and Myburgh, 2015; Wahlström et al., 2017). For example, lateral acceleration captures the side to side movements representing driving events such as left/right turn and lane change, while longitudinal acceleration corresponds to vehicle braking and accelerating behaviours. It should be highlighted that other related signals can also be accommodated by the proposed system via augmenting the feature vector.

4.3. Results

Results of the study includes four parts. The first part is the results of data visualization and analysis. Typical feature distribution and entire features distribution are both visualized and analysed. The second part is about the comparisons of different classifier. As a result, random forest is identified as the most suitable classifier. The third part is the results of optimizing parameters for random forest. The last part is the results of comparing generic, road-aware, personalized model, and personalized model with road information.

4.3.1. Data visualization and analysis

To investigate the differences among different drivers' driving behaviours so that the rationale for personalized state recognition are provided, different data visualization and analysis approaches are adopted including typical features analysis using histogram, and data visualization using t-SNE algorithm along with statistical analysis.

Typical feature analysis: feature analysis under various driving states for different drivers is first conducted using histogram as discussed in Section 3.1. There are seven features for classification as in (1). For the sake of brevity, two typical features are compared including speed and yaw angle. Previous research has indicated that these two signals can better reflect driving behaviours than others (Meiring and Myburgh, 2015; Castignani et al., 2015), where speed indicates the longitudinal behaviour and yaw angle indicates the steering behaviour. The histograms are displayed in Figs 8 and 9, where their corresponding statistics are also calculated and given in Fig 10.

From Figs 8 and 9, it is evident how the distribution of speed and yaw angle varies by different driving states. First for a given driver D_i , the speed and yaw distributions under different driving states differ in terms of shape, mode number and range, which reflects the differences in various driving states. Take the speed histogram of driver No. 5 as an example, it is skewed right under normal state, symmetric under drowsy state and bimodal under aggressive state. More importantly, for a given driving state (i.e. a given column), the speed and yaw distributions for different drivers also differ substantially. Take the speed histogram under aggressive state as an example, it is skewed right for D1, symmetric for D2, multimodal for D3, skewed right for D4 and bimodal for D5. This is also obvious from statistics as shown in Fig 10. By comparing the bar of the same color, one can see the substantial differences (i.e. height and sign) between different drivers. For example, under aggressive driving state (red color), the mean value of yaw angle is positive for D2, D3, but is negative for D1, D4, D5. One can also see from Fig 10 the significant differences in the speed variance under aggressive driving state.

t-SNE visualization: it would be also of great interest to analyse the differences of the entire feature distributions for different drivers under various driving states. Considering that it is generally difficult to visualize high-dimensional data, advanced dimension reduction algorithm t-SNE in Section 3.1 is drawn to achieve this task. By using this approach, the original seven dimensional data is projected onto two dimensional space represented by principle component 1 and principle component 2. For the sake of brevity only the data under aggressive driving state is displayed in Fig 11. The statistics of the principle components for different drivers (D1 to D5) and overall data under various driving states are also calculated and displayed in Fig 12.

It can be visually seen from Fig 11 that the data distributions of different drivers under aggressive driving state vary a lot, similar conclusion can be drawn for the data under normal and drowsy driving

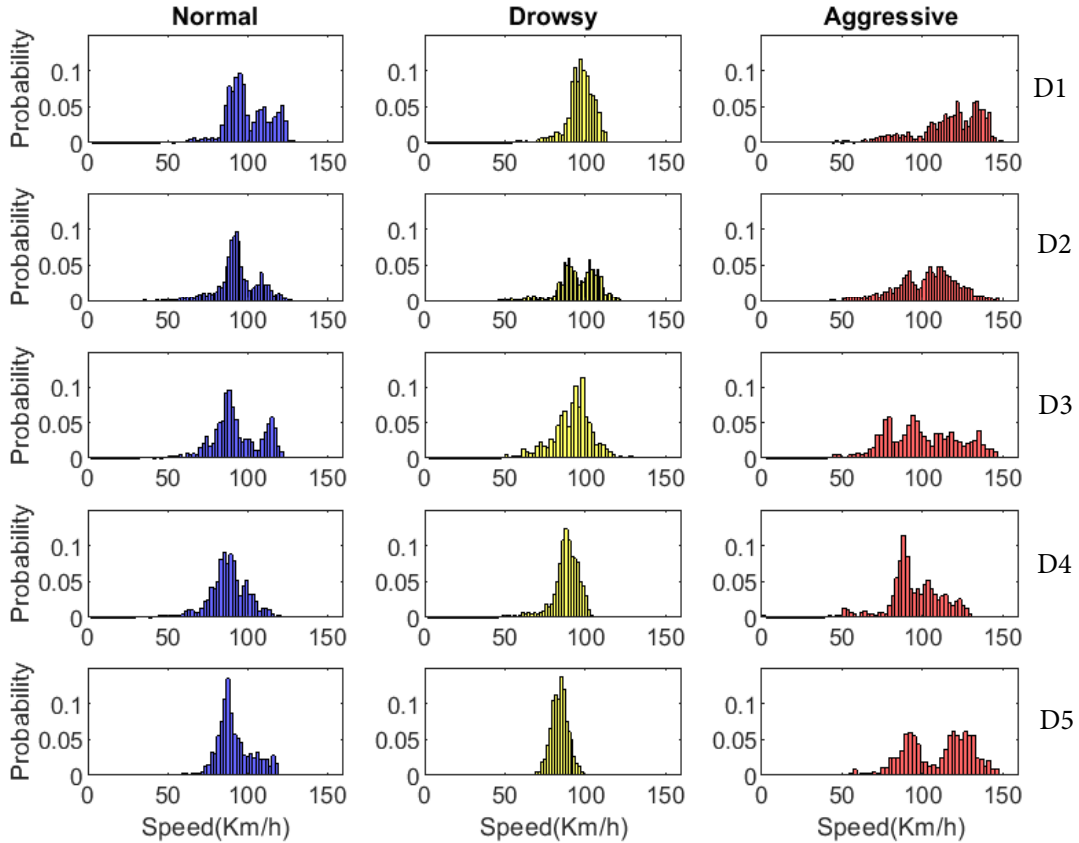


Figure 8: Speed histogram under various driving states (i.e. normal, drowsy, aggressive) for five drivers (D1 to D5 from top to bottom).

states. For example, there are few samples near the centre for driver 1, which is quite different from other drivers. Fig 12 investigates the differences quantitatively through descriptive statistics. One can see the substantial differences (i.e. high and sign) between different drivers. Take the mean of PC 1 as an example, under aggressive driving state (red color), the value for D1, D2, D3 and D4 is negative, but the value for D4 is positive; besides, the value of D2 under normal driving state is quite similar to that of D3 under drowsy driving state. Similar observations can be drawn for the data of PC 2. As a result, if a generic (average) model is trained by using all driver’s data indiscriminately, the personalized driving characteristics of individual driver will be overlooked and consequently results in very limited performance for a particular driver, particularly when the drivers’ behaviours vary a lot.

4.3.2. Similarity and Threshold

The proposed system is able to demonstrate the necessities of adopting a personalized system by analysing the feature distributions of a specific driver’s and all drivers’ data. This is significantly important when a user just starts to use this system and therefore no enough labelled individual data is available. The logic is that if the feature distributions of a specific driver’s and all drivers’ data are very close, a generic model is used to predict his/her driving state, which may avoid overfitting issue. While if the feature distributions are significantly different, a personalized model is trained by using labelled personalized data.

In order to evaluate the similarity between a specific driver’s and all drivers’ data in a quantitative manner, Spearman distance is adopted. Spearman distance is to estimate the fitness of two variables expressed by a monotonic function (Xie et al., 2016). It is also able to compute the similarity between

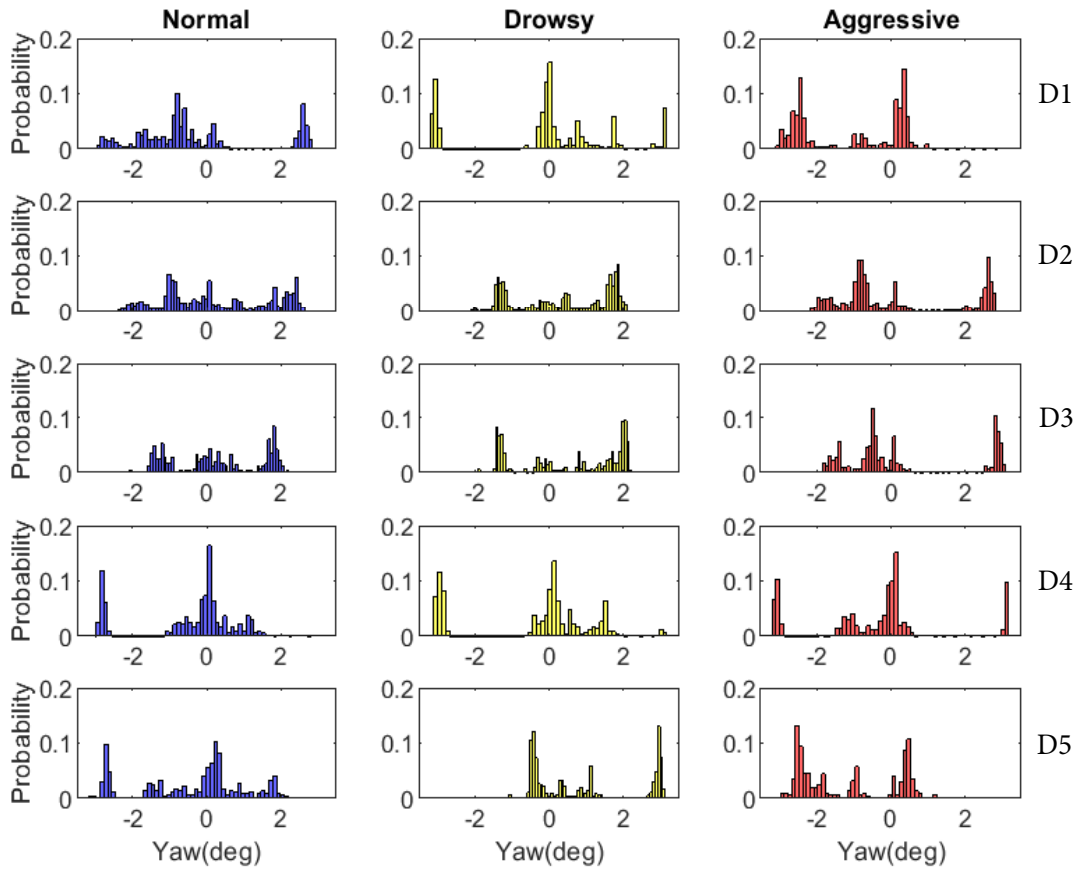


Figure 9: Yaw angle histogram under various driving states (i.e. normal, drowsy, aggressive) for five drivers (D1 to D5 from top to bottom).

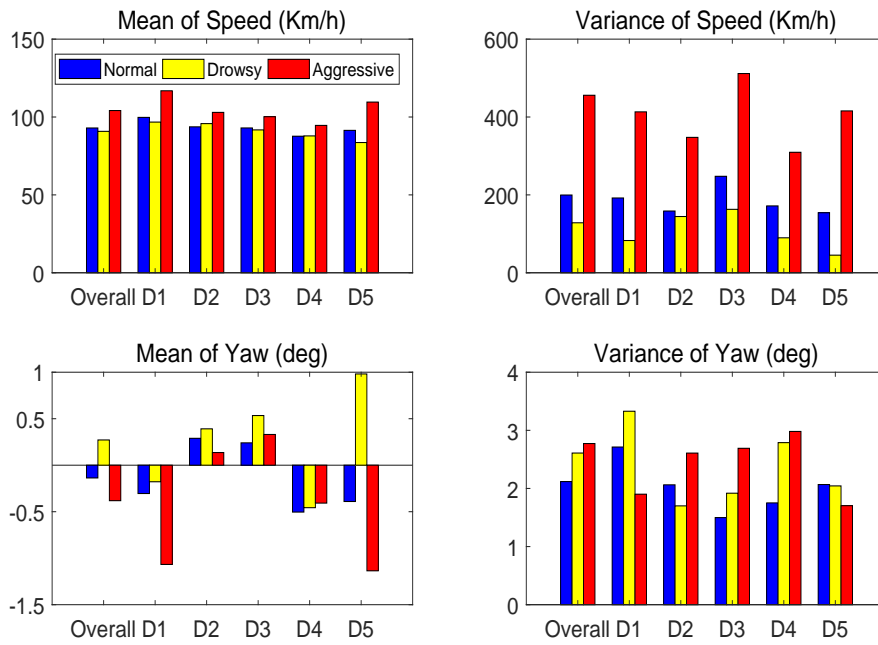


Figure 10: Statistics corresponding to the histograms for speed and yaw angle under various driving states for different drivers including overall data.

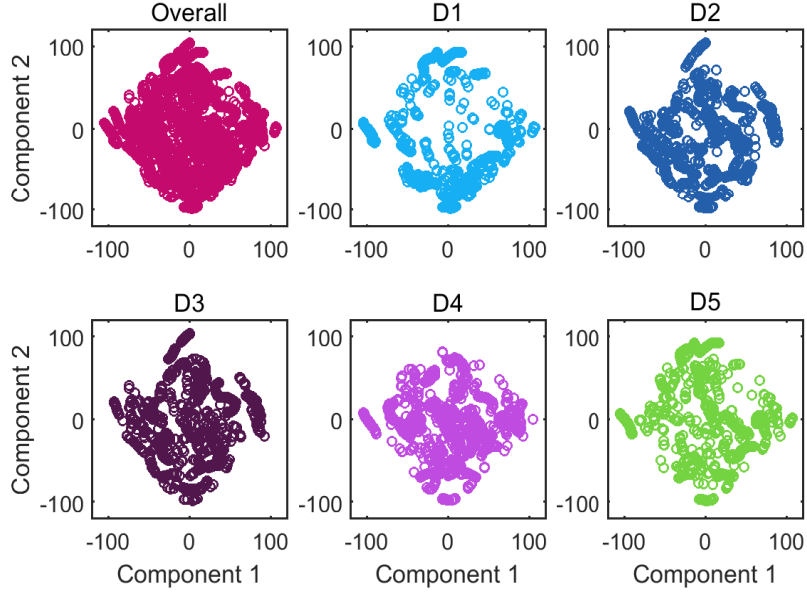


Figure 11: t-SNE visualization for different drivers including overall data under aggressive driving state.

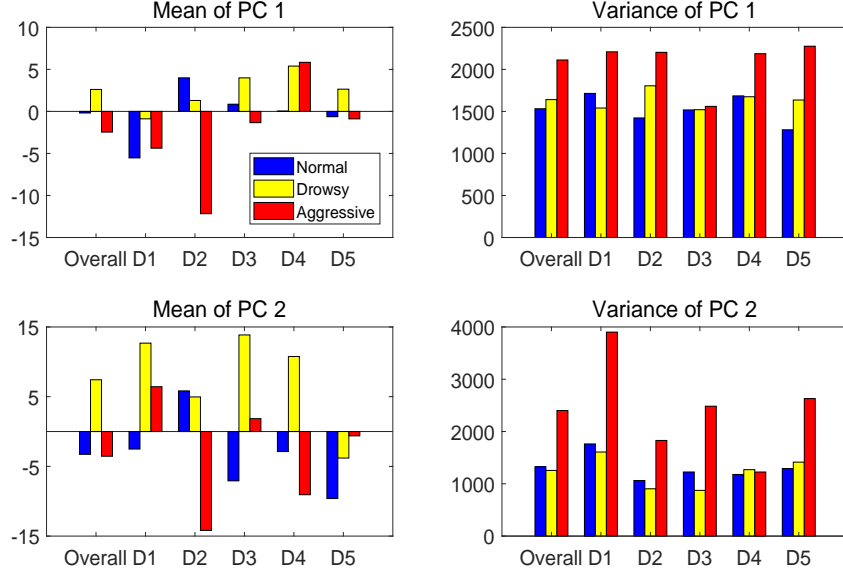


Figure 12: Statistics (mean and variance) comparisons of different principal components for different drivers and overall data under various driving states.

two different datasets with different sizes. The Spearman distance is defined as follows.

$$D = 1 - \frac{(r_s - \bar{r}_s)(r_t - \bar{r}_t)^T}{\sqrt{(r_s - \bar{r}_s)(r_s - \bar{r}_s)^T} \sqrt{(r_t - \bar{r}_t)(r_t - \bar{r}_t)^T}}$$

where r_{sj} is the rank of x_{sj} taken over $x_{1j}, \dots, x_{sj}, x_{mj}$ and r_{tj} is the rank of y_{tj} taken over $y_{1j}, \dots, y_{tj}, y_{mj}$, computed by tired rank method. In the tired rank method, the values are ranked from smallest to largest. Moreover, it returns the average ranks when several x_{sj} or y_{tj} have the same values. In this paper, x and y are two different datasets from a specific driver and all drivers. r_{sj} is the rank of j -th feature with length m for dataset x and r_{tj} is the rank of j -th feature with length m for dataset y . Moreover, r_s and r_t are the coordinate-wise rank vectors of x_s and y_t . x_s and y_t represent a specific item of dataset x or y (a measurement includes all measured features). Here, n is the number of

features. \bar{r}_s and \bar{r}_t are given by

$$\bar{r}_s = \frac{1}{n} \sum_{j=1}^n r_{sj}, \quad \bar{r}_t = \frac{1}{n} \sum_{j=1}^n r_{tj}$$

In order to normalize the difference within $[0, 1]$, (2) is applied to convert Spearman distance into similarity. After obtaining the similarity, an appropriate threshold needs to be selected to determine constructing whether a personalized model should be constructed or not. The threshold is computed by comparing all drivers’ data with the data randomly picked from all drivers’ data. The number of random picking is 5288 (5288 seconds driving data) which is the average data number for five drivers. The aforementioned procedure is automatically implemented by the proposed system.

$$S_i = \frac{D_B - |D_B - D_i|}{D_B}, \quad (2)$$

where D_B is the Spearman distance between the data of all five drivers and the data of randomly picked average data number for five drivers (termed “baseline similarity”). D_i is the Spearman distance between the data of i -th driver and all five drivers. $|D_B - D_i|$ is the absolute difference between D_B and D_i and S_i is the similarity between all five drivers and i -th driver. The baseline similarity and the similarity of five drivers are summarized in Table. 2.

Table 2: The similarity of five drivers

Driver	D1	D2	D3	D4	D5	Baseline
Similarity	0.9783	0.9787	0.9698	0.9822	0.9867	0.9992

According to the experimental results, the baseline similarity is 0.9992. When the similarity of a specific driver is lower than the baseline similarity, a personalized model is constructed for driving state recognition. One can see from Table 2 that the similarity values of five drivers are all lower than the baseline similarity. This is because drivers’ diversity was considered in collecting the driving data in Romera et al. (2016), where the participants were from different genders, different age groups, different vehicle models, and different fuel types. The background information for five test drivers are also summarized in Table. 3 of Romera et al. (2016). Therefore, personalized models are adopted for all five drivers.

Table 3: Background information for different drivers

No.	Gender	Age	Vehicle model	Fuel type
D1	Male	40-50	Audi Q5 (2014)	Diesel
D2	Male	20-30	Mercedes B180 (2013)	Diesel
D3	Male	20-30	Citröen C4 (2015)	Diesel
D4	Female	30-40	Kia Picanto (2004)	Gasoline
D5	Male	30-40	Opel Astra (2007)	Gasoline

4.3.3. Comparisons of different classifiers

In this section, different classifiers are implemented and compared. The algorithms contain five categories: discriminant analysis, decision tree, KNN, SVM and ensemble learning. It should be noted that the random forest with Bayesian parameter optimization in Section 4.3.4 is adopted. These algorithms have been previously applied to different classification applications (Meiring and Myburgh, 2015; Dong et al., 2011) such as driving mode classification, but it is unclear which one is most suitable for driving state recognition. To this end, all drivers’ driving data is aggregated to derive

a generic driving state recognition model, where 5-fold cross validation (Witten et al., 2016) is adopted to evaluate the performance. It is also noted that cross validation is also an effective way to assess the problem of overfitting, since if the trained model is over-fitted its cross validation performance would be poor. The classification results of different approaches are summarized in Table 4, where results are obtained via Bayesian optimization.

Table 4: Comparisons of performance by Classifiers

Classifier	Type	Accuracy
Discriminant Analysis	Linear	51.0%
	Quadratic	51.1%
Decision Tree	Complex	63.1%
	Medium	58.6%
	Simple	52.7%
KNN	Fine	63.1%
	Quadratic	51.1%
	Coarse	61.7%
	Cosine	58.2%
	Cubic	64.1%
	Weighted	66.8%
SVM	Linear	51.0%
	Quadratic	56.4%
	Cubic	24.9%
	Fine Gaussian	67.4%
	Medium Gaussian	62.0%
	Coarse Gaussian	55.6%
Ensemble Learning	Boosted Trees	60.5%
	Random Forest	81.3%
	Subspace Discriminant	51.5%
	Subspace KNN	69.0%
	RUSBoost Trees	57.1%

One can see from Table 4 that random forest with Bayesian parameter optimization outperforms others significantly and consequently chosen as the classification algorithm for the task of driving state recognition. For the sake of completeness, it is briefly discussed in this part. Random forest is one of the most popular machine learning algorithms (Breiman, 2001). As one type of ensemble learnings (Bootstrap Aggregation or bagging), it combines a number of weak learners and works out the final output through voting. Each weak learner is trained by a bootstrap replica, which is obtained by randomly selecting N observations out of N with replacement (where N is the size of dataset). Different from conventional bagging methods, bootstrap replicas are generated randomly and the features for each weak learner is chosen randomly as well. As a result, it is more robust against noise and has less correlation with weak learners. In this paper, Classification And Regression Tree (CART) (Breiman, 2001) is chosen as the weak learner. Therefore, generating bootstrap replica of the dataset and establishing CART tree by using the bootstrap replica are two key steps in random forest in this paper.

Bootstrap replicas of dataset: a subset $S \subset D$ of training samples is randomly chosen with replacement, where D is the entire dataset. Then, the tree is grown using this subset and is not pruned.

Classification And Regression Tree: in this step, generated bootstrap replicas are used to produce CART trees. Given n training samples, namely, features $X = [x^1, \dots, x^p]$ with class label $Y = \{1, 2, \dots, k\}$, classification is to build a model so that the class label Y can be worked out for

new feature data X . The CART model is to divide X into k disjoint sets $\{C_1, C_2, \dots, C_k\}$ and so the predicted value of Y is j if X belongs to C_j . This is achieved by building a binary decision tree according to certain splitting rules (for node determination and threshold selection) based on the features variables. The detailed algorithm is referred to (Rutkowski et al., 2014), where the pseudocodes are given as follows for the sake of completeness.

Various impurity measures for variable selection are available such as Gini gain and Information gain, in this work, Gini index is chosen via trial & error testing (Rutkowski et al., 2014). Moreover, some stopping rules exist in CART algorithm to avoid its adverse effect such as over-fitting. In this paper, one of the stopping rule is considered, i.e., the minimum leaf size, where its determination is transformed into the minimization of the cross validation loss solved via Bayesian optimization (see, Section 4.3.4).

Pseudocodes for CART:

Inputs:

S is a sequence of examples; A is a set of discrete attributes

Output: CART model

1. Let CART be a tree with a single leaf A_0 (root)
 2. Let $A_0 = A$
 3. For each attribute $a^i \in A$, the set of attribute values A^i is partitioned into two disjoint subsets A_L^i and A_R^i , let X^i denote all possible A_L^i
 4. For each $A_L^i \in X^i$,
 5. Calculate the Gini gain and select the variable ($\tilde{A}_{L,q}^i = A_L^i$) which maximizes Gini gain ($\tilde{A}_{L,q}^i$ is a spilt);
 6. Send data $S(\tilde{A}_{L,q}^i)$ to the “leaf node”;
 7. Send data $S(\tilde{A}_{R,q}^i)$ to the “right node”;
 8. Recursively repeat the same process on these two “nodes” until stopping rules are satisfied
 9. Return the CART model
-

4.3.4. Random forest with Bayesian optimization

The parameters of random forest in Section 3.2 are optimized via Bayesian optimization, where the ensemble quality is also evaluated. The maximum number of iterations in Bayesian optimization is chosen 30 (Levitan and Herman, 1987). The number of learners (trees) and minimum leaf size are two key parameters. In this paper, the optimal minimum leaf size is one and minimum observed objective (i.e. classification error) is obtained when function evaluation equals to five. The optimal number of learners are determined by minimizing the estimated objective, which is 496.

Three different methods are adopted to evaluate the quality of ensembles including independent test set (note: the independent test set is not included in training set), out-of-bag error and cross validation. For the more than 26000 samples in this paper, 70% is chosen as training set and the remaining 30% is chosen as testing set. The classification error under the three methods are plotted in Fig 13. One can see that the optimal number of learners is close to 500, similar to the result by using Bayesian optimization. This validates the effectiveness of Bayesian parameter optimization.

4.4. Integrate optimized random forest into models

After discovering that random forest with Bayesian optimization outperforms others in driving state recognition, optimized random forest is integrated into different models including generic model, personalized model, generic model with road information (also termed road-aware model), and personalized model with road information. In this section, the performance of the proposed personalized driving state recognition with road information is evaluated. To investigate the effect of personalized driving characteristics and road information on classification performance, they are evaluated respectively leading to personalized system and road-aware system, respectively.

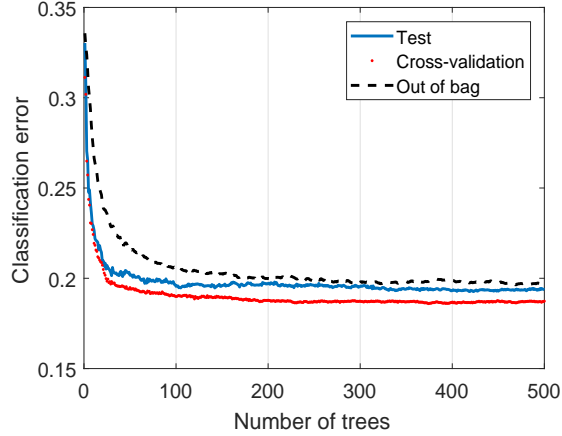


Figure 13: Different approaches to evaluate the ensemble quality.

4.4.1. Personalized model vs. Generic model

First the personalized model is compared to the generic model, where their comparative results for different drivers are displayed in Table 5. It can be seen from Table 5 that

Table 5: Comparisons between personalized model and generic model

Model	Accuracy		
	Personalized model	Generic model	Difference
D1	96.0%	86.9%	9.1%
D2	87.2%	78.3%	8.9%
D3	90.3%	82.0%	8.3%
D4	79.9%	70.9%	9.0%
D5	95.2%	87.4%	7.8%
Overall	89.5%	81.3%	8.2%

- (i) The result of D5 is the best for generic model, and the improvement for D5 using personalized model is the smallest; this is because the distribution of overall data is closest to that of D5, which can be seen from Fig 12.
- (ii) Incorporating personalized driving characteristics, the personalized model substantially outperforms the generic model for all test drivers. And the overall recognition accuracy increases from 81.3% to 89.5%.

4.4.2. Road-aware model vs. Generic Model

The effect of road information available in the experimental dataset is further investigated. Road type can provide important information such as maximum speed, congestion among others. The comparative results between road-aware model and generic model for different drivers are displayed in Table 6. One can also see the performance improvement from Table 6, where road-aware model outperforms generic model for all test drivers. Although the improvement is not significant, it is enough to demonstrate the benefits of incorporating road information into driving state recognition.

4.4.3. Personalized model with road information vs. Road-aware model

In this model, personalized driving characteristics and road information are considered simultaneously. Its comparative results against road-aware model are displayed in Table 7. It can be seen from Table 7 that for all drivers the personalized model outperforms the road-aware model for both

Table 6: Comparisons between road-aware model and generic model

Model	Accuracy		
	Road-aware model	Generic model	Difference
D1	88.9%	86.9%	2.0%
D2	80.8%	78.3%	2.5%
D3	84.2%	82.0%	2.2%
D4	76.7%	70.9%	5.8%
D5	88.6%	87.4%	1.2%
Overall	83.6%	81.3%	2.3%

Table 7: Comparisons between personalized model with road information and road-aware model

Model	Accuracy			
	Personalized model		Road-aware model	
Road Type	Motorway	Secondary	Motorway	Secondary
D1	98.0%	96.7%	89.4%	88.4%
D2	92.5%	86.3%	82.1%	79.4%
D3	96.0%	88.0%	88.7%	79.5%
D4	84.6%	82.9%	77.5%	75.9%
D5	98.0%	94.7%	93.8%	82.7%
Overall	93.6%	89.4%	86.2%	80.9%
D1	97.4%		88.9%	
D2	89.5%		80.8%	
D3	92.0%		84.2%	
D4	83.8%		76.7%	
D5	96.4%		88.6%	
Overall	91.6%		83.6%	

motorway and secondary road. As a consequence, the personalized model with road information outperforms the road-aware model, where the overall accuracy increases from 83.6% to 91.6%.

4.4.4. Detailed comparisons

To make a detailed comparison for four different models under various driving states, the confusion matrices are also plotted and displayed in Fig 14. In this figure, the diagonal cells in green show the number and percentage of correct classification for different driving states. Take the generic model as an example, 9563 samples are correctly classified as normal corresponding to 36.2% of all 26440 samples. The off-diagonal cells show where the mistakes come from. For example, in first row, 1110 of normal samples are incorrectly classified as drowsy corresponding to 4.2% of all 26440 samples. The rightmost column presents the accuracy for each predicted class, while the bottom row shows the accuracy for each true class. For example, out of $9563+1448+1210=12221$ normal prediction, 78.3% are correct and 21.7% are wrong; out of $9563+1110+625=11298$ normal samples, 84.6% are correctly predicted as normal and 15.4% are wrongly predicted where drowsy class takes 4.2% and aggressive class takes 2.4%. The cell at the right bottom shows the overall accuracy, which for the generic model is 81.3%.

Finally the comparative results for the overall accuracy of the four different approaches are summarized in Table 8. It can be concluded that both personalized driving characteristics and road information can improve driving state recognition performance. More importantly, the proposed solution accommodating personalized driving characteristics and road information simultaneously results in the best performance, where the overall accuracy increases from 81.3% to 91.6%.

In practice, the false classification of drowsy driving is more significant to avoid driving risks. To

Output Class	Normal	9563 36.2%	1448 5.5%	1210 4.6%	78.3% 21.7%
	Drowsy	1110 4.2%	6487 24.5%	345 1.3%	81.7% 18.3%
	Aggressive	625 2.4%	217 0.8%	5435 20.6%	86.6% 13.4%
	Recall	84.6% 15.4%	79.6% 20.4%	77.8% 22.2%	81.3% 18.7%
	Target Class	Normal	Drowsy	Aggressive	Precision

(a) Generic model

Output Class	Normal	10315 39.0%	900 3.4%	573 2.2%	87.5% 12.5%
	Drowsy	635 2.4%	7115 26.9%	194 0.7%	89.6% 10.4%
	Aggressive	348 1.3%	137 0.5%	6223 23.5%	92.8% 7.2%
	Recall	91.3% 8.7%	87.3% 12.7%	89.0% 11.0%	89.5% 10.5%
	Target Class	Normal	Drowsy	Aggressive	Precision

(b) Personalized model

Output Class	Normal	9721 36.8%	1242 4.7%	1024 3.9%	81.1% 18.9%
	Drowsy	1013 3.8%	6739 25.5%	309 1.2%	83.6% 16.4%
	Aggressive	564 2.1%	171 0.6%	5657 21.4%	88.5% 11.5%
	Recall	86.0% 14.0%	82.7% 17.3%	80.9% 19.1%	83.6% 16.4%
	Target Class	Normal	Drowsy	Aggressive	Precision

(c) Road-aware model

Output Class	Normal	10468 39.6%	741 2.8%	413 1.6%	90.1% 9.9%
	Drowsy	568 2.1%	7316 27.7%	141 0.5%	91.2% 8.8%
	Aggressive	262 1.0%	95 0.4%	6436 24.3%	94.7% 5.3%
	Recall	92.7% 7.3%	89.7% 10.3%	92.1% 7.9%	91.6% 8.4%
	Target Class	Normal	Drowsy	Aggressive	Precision

(d) Personalized model with road information

Figure 14: Confusion matrices for four different models: generic model, personalized model, road-aware model and personalized model with road information.

obtain the false classification rate of drowsy driving, the problem is formulated as a binary classification problem, where drowsy driving is defined as one kind of class and other driving states (normal driving and aggressive driving) are defined as the other class. The binary classification error tables can be derived from Fig. 14. The error tables are omitted for the sake of brevity. The false classification rate of drowsy driving can be calculated by false negatives (Type II error) over (true positives + false negatives). According to numeric experiments, the proposed personalized model with road information outperforms the other models, where the false classification rate of driving drowsy is only 10.3%. The false classification rate of generic model, road-aware model and personalized model are 20.4%, 17.3% and 12.7%, respectively. For Type I error, the false positive rate of the proposed model is only 8.8%, which is also better than generic model (18.3%), road-aware model (16.4%) and personalized model (10.4%). Therefore, the proposed personalized model with road information achieves the best performance, when drowsy driving recognition is concerned.

Table 8: Comparisons for four different models

Model	generic model	personalized model
Without road info	81.3%	89.5%
With road info	83.6%	91.6%

5. Discussion

The main aim of this study is to develop a personalized driving state recognition system by learning from non-intrusive, easily accessible vehicle related measurements. Previous studies are mainly focused on driving state based on generic driving behaviours of all drivers (Masala and Grosso, 2014). However, this generic model may overlook personalized driving characteristics and preferences, and therefore leads to a degraded performance for a specific driver.

As demonstrated in Section. 4.4, this aim is achieved and it is one of the key contributions of this paper. The proposed personalized driving state recognition system has advantages of (i) accommodating personalized driving characteristics and preferences; (ii) considering contextual information (e.g. road type). Due to the aforementioned advantages, the proposed personalized driving state recognition system significantly enhances the accuracy of driving state recognition for a specific driver.

Based on an intuitive feeling, a personalized system should be better than a generic model. However, both qualitative and quantitative analysis are required to demonstrate the necessities of constructing a personalized system especially when only a small personalized data is available. Using a small size data to train a model may lead to the problem of overfitting. To this end, two data visualization methods are applied to identify the differences between individual driver and all drivers in feature space. The first data visualization method is to analyse a typical driving feature such as speed, yaw angle (Birrell et al., 2015). The second data visualization method is t-SNE algorithm. Considering that the entire features are seven-dimensional, t-SNE algorithm is used to reduce dimensionality from seven to two so that the distribution of entire features can be visually analysed (Bejani and Ghatee, 2018). As shown in Section. 4.3.1, vehicle-related measurements of different drivers under different states are significantly different. As a result, a personalized model is more likely to outperform a generic model for driving state recognition.

Several classifiers are compared to select the most suitable one for driving state recognition. According to the experimental results in Section. 4.3.3, random forest with Bayesian optimization outperforms others. The possible reasons are summarized as follows. (i) Vehicle-related measurements from GPS, IMU, and gyroscope have been proved to be effective for modelling driving behaviours under various conditions (Meiring and Myburgh, 2015; Wahlström et al., 2017). (ii) Random forest makes a decision based on the prediction of all ensembles (classifiers), which is able to avoid over-fitting (Breiman, 2001). (iii) The ensembles of random forest are decision tree. When constructing a decision tree, the most useful features are always at the top layers, which makes random forest implement feature selection automatically (Pitombo et al., 2017).

In addition, it is generally difficult to tune the parameters of classifiers. In this study, the parameters of classifiers are optimized by Bayesian optimization algorithm, which can optimize the parameters of different classifiers in an automatic manner (Suzdaleva and Nagy, 2014).

There are also some limitations in the proposed personalized driving state recognition system, which should be further improved. The following aspects are identified.

- The false classification of drowsy driving can be reduced to further enhance driving safety. This can be achieved at decision level and at training level. At decision level, the threshold for determining drowsy driving can be tuned. The lower the threshold is, the more sensitive the system is. While at training level, drowsy driving data can be allocated a higher weight over other driving state data. As a result, the system will be more sensitive to drowsy driving.
- Online learning may be integrated into the proposed framework, since individual driver’s driving

behaviours may change with more driving experience being accumulated. The design of online learning algorithm can borrow the experiences from (Suzdaleva and Nagy, 2018).

- It is challenging to manually label personal data. To this end, a personalized learning framework can be developed, which can automatically learn personal driving characteristics and preference from raw data. The results in (Tang et al., 2018b) are very useful in this regard.
- It has been proved that travel time can be estimated more precisely by considering temporal information (Tang et al., 2018a). Therefore, temporal driving characteristics can be considered so that classification performance can be further improved;
- More related contextual information can be accommodated such as traffic situation, weather condition, continuous driving time; this will inevitably require a larger amount of driving data;

6. Conclusions

This paper addressed the problem of driving state recognition by learning from non-intrusive, easily accessible vehicle related measurements via a smartphone (speed via GPS, three-axis accelerations and orientation angles via IMU). To guarantee a satisfying performance so that adaptive assistance systems are more acceptable and user-friendly, a machine learning based personalized system with road information is proposed, where personalized driving characteristics of individual drivers and road information are accommodated cooperatively. The proposed system is validated by using the UAH-DriveSet dataset, which was collected in 2016 in real world environments.

Different classifiers are compared to identify a suitable one for the driving state recognition including Discriminant Analysis, Decision Tree, KNN, SVM, Ensemble Learning and their variations, where random forest with Bayesian parameter optimization outperform others. Adopting the random forest as the classifier, experimental comparisons are further made on different models under various frameworks. Experimental results demonstrate that the proposed machine learning based personalized system with road information significantly outperforms the conventional generic model because personalized driving characteristics and road information are taken into account, where the overall recognition accuracy increases from 81.3% to 91.6%. It should be noted that although the proposed personalized learning framework is proposed for the problem of driving state recognition, it can find a wide range of applications where diverse behaviours exist.

Acknowledgement

This work was supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) Autonomous and Intelligent Systems programme under the grant number EP/J011525/1 with BAE Systems as the leading industrial partner. Dewei Yi would also like to thank Chinese Scholarship Council for supporting his study in the UK.

References

- S. Al-Sultan, A. H. Al-Bayatti, and H. Zedan. Context-aware driver behavior detection system in intelligent transportation systems. *IEEE transactions on vehicular technology*, 62(9):4264–4275, 2013.
- Y. Ba, W. Zhang, Q. Wang, R. Zhou, and C. Ren. Crash prediction with behavioral and physiological features for advanced vehicle collision avoidance system. *Transportation Research Part C: Emerging Technologies*, 74: 22–33, 2017.
- M. M. Bejani and M. Ghatee. A context aware system for driving style evaluation by an ensemble learning on smartphone sensors data. *Transportation Research Part C: Emerging Technologies*, 89:303–320, 2018.

- S. A. Birrell, D. Wilson, C. P. Yang, G. Dhadyalla, and P. Jennings. How driver behaviour and parking alignment affects inductive charging systems for electric vehicles. *Transportation Research Part C: Emerging Technologies*, 58:721–731, 2015.
- L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- A. D. Bull. Convergence rates of efficient global optimization algorithms. *Journal of Machine Learning Research*, 12(Oct):2879–2904, 2011.
- V. A. Butakov and P. Ioannou. Personalized driver/vehicle lane change models for adas. *IEEE Transactions on Vehicular Technology*, 64(10):4422–4431, 2015.
- G. Castignani, T. Derrmann, R. Frank, and T. Engel. Driver behavior profiling using smartphones: A low-cost platform for driver monitoring. *IEEE Intelligent Transportation Systems Magazine*, 7(1):91–102, 2015.
- S. Chandaka, A. Chatterjee, and S. Munshi. Cross-correlation aided support vector machine classifier for classification of eeg signals. *Expert Systems with Applications*, 36(2):1329–1336, 2009.
- X. M. Chen, M. Zahiri, and S. Zhang. Understanding ridesplitting behavior of on-demand ride services: An ensemble learning approach. *Transportation Research Part C: Emerging Technologies*, 76:51–70, 2017.
- Y. Dong, Z. Hu, K. Uchimura, and N. Murayama. Driver inattention monitoring system for intelligent vehicles: A review. *IEEE transactions on intelligent transportation systems*, 12(2):596–614, 2011.
- M. A. Gelbart, J. Snoek, and R. P. Adams. Bayesian optimization with unknown constraints. *arXiv preprint arXiv:1403.5607*, 2014.
- I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182, 2003.
- W. Hailin, L. Hanhui, and S. Zhumei. Fatigue driving detection system design based on driving behavior. In *Optoelectronics and Image Processing (ICOIP), 2010 International Conference on*, volume 1, pages 549–552. IEEE, 2010.
- W. Hajek, I. Gaponova, K. Fleischer, and J. Krems. Workload-adaptive cruise control—a new generation of advanced driver assistance systems. *Transportation research part F: traffic psychology and behaviour*, 20: 108–120, 2013.
- J.-H. Hong, B. Margines, and A. K. Dey. A smartphone-based sensing platform to model aggressive driving behaviors. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 4047–4056. ACM, 2014.
- J. Hu, L. Xu, X. He, and W. Meng. Abnormal driving detection based on normalized driving behavior. *IEEE Transactions on Vehicular Technology*, 66(8):6645–6652, 2017.
- S. L. Jamson, D. L. Hibberd, and A. H. Jamson. Drivers ability to learn eco-driving skills; effects on fuel efficient and safe driving behaviour. *Transportation Research Part C: Emerging Technologies*, 58:657–668, 2015.
- Q. Ji, Z. Zhu, and P. Lan. Real-time nonintrusive monitoring and prediction of driver fatigue. *IEEE transactions on vehicular technology*, 53(4):1052–1068, 2004.
- Q. Ji, P. Lan, and C. Looney. A probabilistic framework for modeling and real-time monitoring human fatigue. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and humans*, 36(5):862–875, 2006.
- J. W. Joubert, D. de Beer, and N. de Koker. Combining accelerometer data and contextual variables to evaluate the risk of driver behaviour. *Transportation research part F: traffic psychology and behaviour*, 41:80–96, 2016.
- S. Kanarachos, S.-R. G. Christopoulos, and A. Chroneos. **Smartphones as an integrated platform for monitoring driver behaviour: The role of sensor fusion and connectivity.** *Transportation research part C: emerging technologies*, 95:867–882, 2018.

- P. Konstantopoulos, P. Chapman, and D. Crundall. Driver’s visual attention as a function of driving experience and visibility. using a driving simulator to explore drivers eye movements in day, night and rain driving. *Accident Analysis & Prevention*, 42(3):827–834, 2010.
- E. Levitan and G. T. Herman. A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography. *IEEE Transactions on Medical Imaging*, 6(3):185–192, 1987.
- L. v. d. Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov): 2579–2605, 2008.
- G. Masala and E. Grosso. Real time detection of driver attention: Emerging solutions based on robust iconic classifiers and dictionary of poses. *Transportation research part C: emerging technologies*, 49:32–42, 2014.
- G. A. M. Meiring and H. C. Myburgh. A review of intelligent driving style analysis systems and related artificial intelligence algorithms. *Sensors*, 15(12):30653–30682, 2015.
- C. S. Pitombo, A. D. de Souza, and A. Lindner. Comparing decision tree algorithms to estimate intercity trip distribution. *Transportation Research Part C: Emerging Technologies*, 77:16–32, 2017.
- E. Romera, L. M. Bergasa, and R. Arroyo. Need data for driver behaviour analysis? presenting the public uah-driveset. In *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*, pages 387–392. IEEE, 2016.
- L. Rutkowski, M. Jaworski, L. Pietruczuk, and P. Duda. The cart decision tree for mining data streams. *Information Sciences*, 266:1–15, 2014.
- C. Saiprasert, T. Pholprasit, and S. Thajchayapong. Detection of driving events using sensory data on smart-phone. *International Journal of Intelligent Transportation Systems Research*, 15(1):17–28, 2017.
- S. Schnelle, J. Wang, H. Su, and R. Jagacinski. A driver steering model with personalized desired path generation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(1):111–120, 2017.
- J. Snoek, H. Larochelle, and R. P. Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pages 2951–2959, 2012.
- P. St-Aubin, N. Saunier, and L. Miranda-Moreno. Large-scale automated proactive road safety analysis using video data. *Transportation Research Part C: Emerging Technologies*, 58:363–379, 2015.
- R. Sun, K. Han, J. Hu, Y. Wang, M. Hu, and W. Y. Ochieng. Integrated solution for anomalous driving detection based on beidou/gps/imu measurements. *Transportation research part C: emerging technologies*, 69:193–207, 2016.
- E. Suzdaleva and I. Nagy. [Data-based speed-limit-respecting eco-driving system](#). *Transportation Research Part C: Emerging Technologies*, 44:253–264, 2014.
- E. Suzdaleva and I. Nagy. An online estimation of driving style using data-dependent pointer model. *Transportation research part C: emerging technologies*, 86:23–36, 2018.
- K. Tang, S. Chen, and Z. Liu. Citywide spatial-temporal travel time estimation using big and sparse trajectories. *IEEE Transactions on Intelligent Transportation Systems*, 19(12):4023–4034, 2018a.
- K. Tang, S. Chen, Z. Liu, and A. J. Khattak. [A tensor-based Bayesian probabilistic model for citywide personalized travel time estimation](#). *Transportation Research Part C: Emerging Technologies*, 90:260–280, 2018b.
- R. Vaiana, T. Iuele, V. Astarita, M. V. Caruso, A. Tassitani, C. Zaffino, and V. P. Giofrè. Driving behavior and traffic safety: an acceleration-based safety evaluation procedure for smartphones. *Modern Applied Science*, 8 (1):88, 2014.
- J. Wahlström, I. Skog, and P. Händel. Smartphone-based vehicle telematics: A ten-year anniversary. *IEEE Transactions on Intelligent Transportation Systems*, 18(10):2802–2825, 2017.

- I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.
- C. Woo and D. Kulić. Manoeuvre segmentation using smartphone sensors. In *Intelligent Vehicles Symposium (IV), 2016 IEEE*, pages 572–577. IEEE, 2016.
- M. Wu, S. Zhang, and Y. Dong. A novel model-based driving behavior recognition system using motion sensors. *Sensors*, 16(10):1746, 2016.
- Y. Xie, Y. Wang, A. Nallanathan, and L. Wang. An improved k-nearest-neighbor indoor localization method based on spearman distance. *IEEE Signal Processing Letter*, 23(3):351–355, 2016.
- C. Xiong, M. Shahabi, J. Zhao, Y. Yin, X. Zhou, and L. Zhang. **An integrated and personalized traveler information and incentive scheme for energy efficient mobility systems**. *Transportation Research Part C: Emerging Technologies*, 2019.
- D. Yi, J. Su, C. Liu, and W.-H. Chen. Personalized driver workload inference by learning from vehicle related measurements. *IEEE Transactions on Systems, Man and Cybernetics: Systems*, 49(1):159 – 168, 2019a.
- D. Yi, J. Su, C. Liu, and W.-H. Chen. New driver workload prediction using clustering-aided approaches. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(1):64–70, 2019b.
- C.-W. You, M. Montes-de Oca, T. J. Bao, N. D. Lane, H. Lu, G. Cardone, L. Torresani, and A. T. Campbell. Carsafe: a driver safety app that detects dangerous driving behavior using dual-cameras on smartphones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 671–672. ACM, 2012.
- B. Yu, W. H. Lam, and M. L. Tam. Bus arrival time prediction at bus stop with multiple routes. *Transportation Research Part C: Emerging Technologies*, 19(6):1157–1170, 2011.
- F. Yuan and R. L. Cheu. Incident detection using support vector machines. *Transportation Research Part C: Emerging Technologies*, 11(3-4):309–328, 2003.
- C. Zhang, H. Wang, and R. Fu. Automated detection of driver fatigue based on entropy and complexity measures. *IEEE Transactions on Intelligent Transportation Systems*, 15(1):168–177, 2014.
- J. Zhang, Z. Yin, and R. Wang. Recognition of mental workload levels under complex human–machine collaboration by using physiological features and adaptive support vector machines. *IEEE Transactions on Human-Machine Systems*, 45(2):200–214, 2015.
- T. Zhang, D. Tao, X. Qu, X. Zhang, R. Lin, and W. Zhang. **The roles of initial trust and perceived risk in publics acceptance of automated vehicles**. *Transportation research part C: emerging technologies*, 98:207–220, 2019.
- B. Zhu, Y. Jiang, J. Zhao, R. He, N. Bian, and W. Deng. **Typical-driving-style-oriented personalized adaptive cruise control design based on human driving data**. *Transportation research part C: emerging technologies*, 100:274–288, 2019.
- J. Ziegler, T. Hussein, D. Münter, J. Hofmann, and T. Linder. Generating route instructions with varying levels of detail. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 31–38. ACM, 2011.