Driver Drowsiness Detection based on Non-intrusive Metrics Considering Individual Specifics

Xuesong Wang, PhD
Professor
School of Transportation Engineering,
Tongji University
4800 Cao'an Road, Jiading District
Shanghai, 201804, China
Phone: +86-21-69583946
Fax: +86-21-69582897
Email: wangxs@tongji.edu.cn

Chuan Xu*
PhD student
School of Transportation Engineering
Tongji University
Room 338, Building of School of Transportation Engineering
4800 Cao'an Road, Shanghai, 201804, P. R. of China
Tel: +86 15216706076
Email: xuchuan7@gmail.com
(Corresponding Author)

Xiaohong Chen, PhD
Professor
School of Transportation Engineering
Tongji University
4800 Cao'an Road, Shanghai, 201804, P. R. of China
Tel) +86-21-65989270 Fax) +86-21-65982897
Email: chenxh@tongji.edu.cn

Word Count: 4732 (Text) +6 Figures+3 Tables=6982words

July 2014

Submitted for Presentation at the TRB Annual Meeting, and Publication in the Journal of the Transportation Research Board

ABSTRACT

Drowsy driving is a serious highway safety problem. If drivers could be warned before they became too drowsy to drive safely, some of these crashes could be prevented. Being able to reliably detect drowsiness depends on the presentation of timely warnings of drowsiness. To date, the effectiveness of drowsiness detection methods has been limited by their failure to consider individual differences. The present study sought to develop a drowsiness detection model that accommodates the varying effects of drowsiness on individual driving performance. Nineteen driving behavior variables and four eye feature variables were measured as participants drove a fixed road course in a high fidelity motion-based driving simulator after having worked an 8-hour night shift. During the test, participants were asked to report their drowsiness level using the Karolinska Sleepiness Scale (KSS) at the midpoint of each of the six rounds through the road course. A multilevel Ordered Logit model (MOL), an Ordered Logit model (OL), and an Artificial Neural Network model (ANN) were used to determine drowsiness. The MOL had the highest drowsiness detection accuracy, and this finding shows that consideration of individual differences improves the models' ability to detect drowsiness. According to the results of the models, percentage of eyelid closure, average pupil diameter, standard deviation of lateral position and steering wheel reversals were most important in the 23 variables.

Keywords: Drowsiness Detection, Multilevel Ordered Logit Model, Non-Intrusive, Driving Behavior, Eye Feature, Driving Simulator

1 INTRODUCTION

Drowsy driving is a serious threat to road safety. In 2009, over 800 fatalities and 30,000 injuries from car crashes were attributed to drowsy driving in the United States (*1*). In Europe, an estimated 20% of traffic crashes are caused by drowsy driving (*2*). The problem seems more serious in China where, in 2007, 1768 fatalities were attributed to drowsy driving (*3*). Considering that as of 2012, China had over 85,000 kilometers of expressways (*4*), and the annual growth in kilometers exceeded 14% during 2002-2011, the problem requires immediate attention. A recent study using a naturalistic driving method, estimated the increase in crash risk associated with drowsy driving to be four to six times greater than when driving while alert (*5*). Besides increasing crash risk, drowsy driving crashes are often more severe than other crashes because they frequently occur on high speed expressways, and are frequently run-off-the-road crashes with no braking prior to impact.

Unlike drinking and driving, drowsy driving does not provide an objective measure of its occurrence, and therefore enforcement cannot be used to counter this problem (*6*). An alternative would be to notify the driver if he becomes too drowsy to drive safely. This requires the reliable detection of drowsy driving—a problem that has been extensively researched. Based on the type of data used, drowsiness detection can be conveniently separated into the two categories of intrusive and non-intrusive methods. Intrusive methods, such electroencephalograms (EEGs) (*7*) or electrocardiograms (EKGs) (*8*), show good detection accuracy, however, are limited to the research laboratory. In contrast, methods based on non-intrusive measures detect drowsiness by measuring driving behavior and sometimes eye features, and so are useful for real world driving situations.

To date, non-intrusive methods have been less reliable than intrusive methods, partly because individual differences in non-intrusive measures have prevented the identification of the point at which drowsiness impairs driving (*9; 10*). In the previous research, individual differences of non-intrusive methods were frequently mentioned, for both driving behavior and eye features. For driving behavior, Ingre et al. (*9*) report the individual difference of standard deviation of lateral position (SDLP), an important driving behavior measure to drowsiness. In this research, for the same drowsiness level, different drivers have different SDLPs. Another study finds the individual difference of driver's lane departure behavior (*11*). Thiffault and Bergeron (*12*) reported the individual difference of the standard deviation of steering wheel movements. Regarding eye features, individual differences of blink duration (*9; 13*), percentage of eyelid closure (PERCLOS) (*14*) were also observed in many studies. However, most drowsiness detection methods such as decision trees, logistic regression, Bayesian networks (*15*), artificial neural networks (*8*), and support vector machines (*16*) have not properly handled the problem of differences in the manifestation of drowsiness among individuals. Ignoring such differences reduces the accuracy and reliability of these models, especially for non-intrusive measures.

To address and increase the accuracy for drowsiness detection model based on non-intrusive measures, a multilevel logit model was built based on both driving behavior measures and eye features, which detect drowsiness by using individual-specific criterions. To compare the detection accuracy, two non-individual specific models (i.e., ordered logit model and artificial neural network) were established. All the data in this research is based on a high fidelity driving simulator experiment, in which driving behavior, eye features, and subjective drowsiness were scaled and collected.

2 METHOD

2.1 Participants

Sixteen male participants aged 24-40 (mean 32.8, SD, 5.0) with valid Chinese drivers licenses were recruited from students and staff at Tongji University. They were required to be in good health, have no sleep related disorders, and not to have taken any pharmaceuticals within one month prior to entering the study. Subjects who had a history of motion sickness were screened out. All subjects provided written consent and were paid about 200 RMB Yuan, depending on the total time in the laboratory. During the experiment, one subject’s eye movement data was lost because of a technical problem. One subject fell asleep before completing the task, however, his data up to that point was used.

2.2 Apparatus

The Tongji University driving simulator is shown in FIGURE 1. This simulator, currently the most advanced in China, incorporates a fully instrumented Renault Megane III vehicle cab in a dome mounted on an 8 degree-of-freedom motion system with an X-Y range of 20 × 5 meters. An immersive 5 projector system provides a front image vi Tongji advanced driving simulator ew of 250 °× 40 °at 1000 × 1050 resolution refreshed at 60 Hz. LCD monitors provide rear views at the central and side mirror positions. SCANeR^{TM} studio software presented the simulated roadway and controlled a force feedback system that acquired data from the steering wheel, pedals and gear shift lever.

FIGURE 1 Tongji advanced driving simulator

Eye movement data were recorded using a Smarteye® eye tracking system. The cameras and interface of this system are shown in FIGURE 2. The system uses four cameras located in the front of the vehicle to record the driver's eye movements at 60 Hz sampling rate.

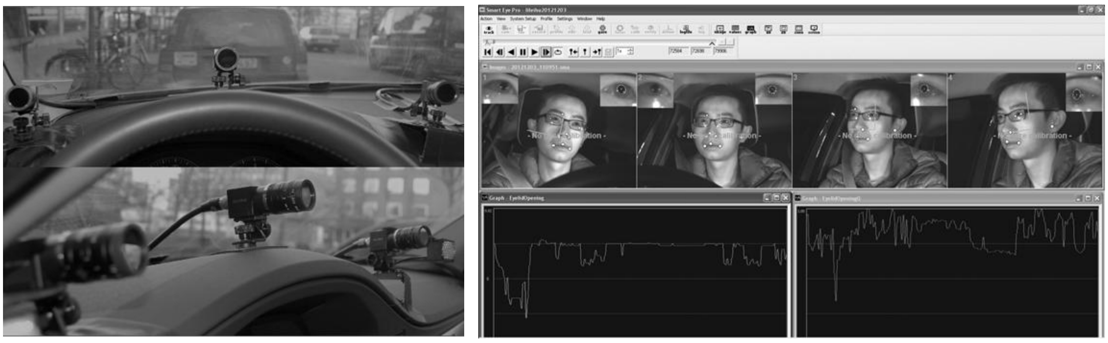

FIGURE 2 Smarteye® eye tracking system hardware and software interface

2.3 Procedure

2.3.1 Experiment design

All drivers received were presented with the same conditions in the same order. The driving course, diagrammed in FIGURE 3, simulated a 20km ring shaped 6-lane rural highway;, 3.75m in width composed of the straight segments numbered 1, 3, 5, 9, 11, 13; two circle curves each with a 700 meter radius, (segments numbered 7, 15); and several transition curves (numbered 2,4,6,9,10,12,14,16). Only the straight line road segments were used for the analysis. The length of both straight line and curve segments was a total of 2 km. Grass and trees were placed beside the highway, as well as a few small villages along the straight segments.

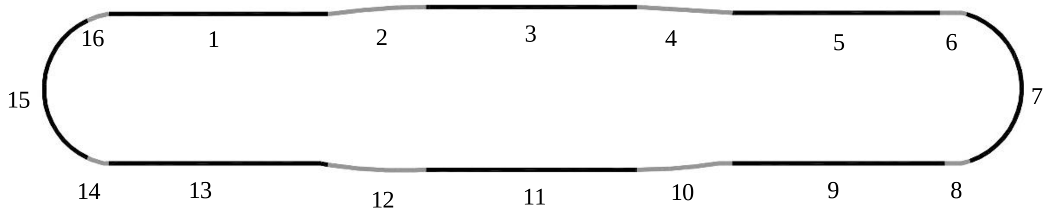

FIGURE 3 Ring shaped highway and segment ID numbers

To assess driving performance at high levels of drowsiness, night shift workers were selected and tested just after shift completion around 8:00 a.m. Upon arrival at the simulator facility, participants were asked to complete questionnaires on their basic driver information and current levels of drowsiness, after which they spent five minutes familiarizing themselves with driving the simulator. At about 8:30 a.m., they received the main test, in which each subject was asked to drive and respect road rules for one hour.

In order to induce drowsiness,

- The driving task was reduced to one lane, eliminating the need to change lanes
- Drivers were required to drive at a constant speed (120 km/h), eliminating the need for manual gear changes
- No radio or music was played
- No environmental disturbances (e.g. crosswinds) were introduced

- All driving was during daytime periods and no tunnel or weather changes occurred eliminating the need to adjust headlights
- Only occasional and uneventful traffic was present.

After the main test, participants were asked to complete post experiment questionnaires on their levels of drowsiness.

2.3.2 Measurement of drowsiness—participants' assessments

To track drivers' drowsiness changes during the one hour driving task, the participants were asked to report their Karolinska Sleepiness Scale (KSS) level at the middle point of pass through the driving course. KSS is on a 9 point ordinal scale, however, it's not necessary to distinguish all nine levels when using the KSS. It is, however necessary to identify the drowsiness level with a high crash risk. Several studies (*15; 17*) suggested that serious behavioral and physiological changes do not occur until KSS>=7. In addition, the description of KSS 7 is "sleep, but no effort to keep alert," but KSS 8, 9 are described as "need some or great effort to keep alert." Therefore, drowsiness in this study is categorized into three levels as follows :

- Level 1 (DL=1): KSS range from 1 to 6, no drowsiness or low-level of drowsiness;
- Level 2 (DL=2): KSS is 7, moderate-level of drowsiness;
- Level 3 (DL=3): KSS range from 8 to 9, high-level of drowsiness.

2.3.3 Measurement of the effects of drowsiness

Vehicle-based signals measuring driving behavior, with 10Hz sample frequency, were obtained from the system, including vehicle speed, lateral position and steering wheel angle. Based on these signals, several drowsiness related behavior indicators were extracted. Eye movement signals including eyelid opening and pupil diameter were also recorded, with 60Hz sample frequency. Using the Smarteye Pro software, eye blink was identified. Eye activity indicators including percentage of eye closure (PERCLOS), average pupil diameter, blink frequency, and average blink duration were calculated. The measures are summarized in TABLE 1 below.

TABLE 1 Driving Behavior and Eye Feature Metrics

Metrics	Description of the Variables	Mean	S.D.
Driving Behavior			
LP_stdev	Standard deviation of lateral position (m)	0.306	0.134
LP_avg	Average of lateral position (m)	0.214	0.269
LD_Area	Sum of lane departure time-space area (m•s)	1.627	4.207
LD_TArea	Sum of lane departure time-space area weighted by lane crossing time (m•s)	6.129	35.383
LD_Frequency	Lane departure frequency	0.660	1.118
LD_Speed	Lane departure lateral speed (m/s)	0.046	0.099
LD_Tc	Time percentage of lane crossing of the vehicle center	0.002	0.017
LD_Te	Time percentage of lane crossing of the vehicle edge	0.021	0.045
SW_Speed_stdev	Standard deviation of steering angular speed (degree/s)	0.012	0.008
SW_Area_MA	Area surrounded by steering angle and its moving average	0.440	0.257
SWM_Re	Steering wheel reversals	190.485	29.522
SW_Range_1	Percentage of steering speed in 0-2.5 degree/s	0.876	0.085
SW_Range_2	Percentage of steering speed in 2.5-5 degree/s	0.077	0.037
SW_Range_3	Percentage of steering speed in 5-7.5 degree/s	0.024	0.020
SW_Range_4	Percentage of steering speed in 7.5-10 degree/s	0.010	0.012
SW_Range_5	Percentage of steering speed exceeding 10 degree/s	0.013	0.024
Speed	Average speed (km/h)	117.424	6.650
Speed_stdev	Standard deviation of speed (km/h)	2.866	2.860
Speeding_T	Time percentage of speed exceeding the limit speed 120 km/h	0.311	0.372
Eye Features			
Blink_Frequency	Average blink frequency per second	0.504	0.318
Blink_duration	Average blink duration (second)	0.402	0.054
PERCLOS	Percentage of eyelid closure	0.132	0.099
Pupil	Average pupil diameter (mm)	3.807	0.894

2.3.4 Data analysis

Using DL as the dependent variable, and the driving behavior and eye feature metrics described in TABLE 1 as the independent variables, three models, individual-specific model multilevel ordered logit model, and two non-individual-specific models, an ordered logit model and neural network model.

The data set was divided into a training set and a validation set. The data in the training set was used to build the models, and the data in the validation set was used to test the models. To ensure each of the two data sets contained the data from every subject at every drowsiness level, the training set was generated by randomly selecting 70% of the data for each subject at each drowsiness level, and the rest of the data was assigned to the validation set. The individual specific multilevel ordered logit model was established first, and then the non-individual specific ordered logit model and neural network model were constructed using the same variables as the multilevel ordered logit model.

- Multilevel Ordered Logit Model

Multilevel ordered logit models are often presented as cumulative logit models. Suppose an ordered DL_{ij} is the drowsiness level for i_{th} subject on the j_{th} road segment. A latent continuous variable DL_{ij}^* is established as the unobserved measure of DL_{ij}. DL_{ij}^* is related to DL_{ij} by a series of latent thresholds. Differing from the ordered logit model, the multilevel model accounts for each subject's individual performance by using a set of variable thresholds specific to each subject: γ_{ki} (k=1, 2), see formula (2).

$$\mathrm{DL}_{ij} = \begin{cases} 1 & if \quad DL_{ij}^{*} < \gamma_{1i} \\ 2 & if \ \ \gamma_{1i} < DL_{ij}^{*} < \gamma_{2i} \\ 3 & if \ \ \gamma_{2i} < DL_{ij}^{*} \end{cases} \tag{1}$$

The DL_{ij}^{*} can be written in the same form as the regular linear regression model.

$$DL_{ij}^{*} = \theta_{ij} + \varepsilon_{ij} \quad \text{and} \quad \theta_{ij} = \sum_{p=1}^{P} \beta_p x_{pij} \tag{2}$$

Where x_{pij} is the explanatory variable for i_{th} subject on j_{th} segment. ε_{ij} is the disturbance term, which is assumed as a logistic distribution as the cumulative density function. Thus, the cumulative response probabilities of the ordinal DL may be denoted as:

$$P_{ij(k)} = \Pr\left(DL_{ij}^{*} \leq k\right) = F\left(\gamma_{ki} - \theta_{ij}\right) = \frac{\exp(\gamma_{ki} - \theta_{ij})}{1 + \exp(\gamma_{ki} - \theta_{ij})}, k = 1,2 \tag{3}$$

$$\mathrm{Logit}\left(P_{ij(k)}\right) = \log\left[\frac{P_{ij(k)}}{1 - P_{ij(k)}}\right] = \log\left[\frac{\Pr\left(DL_{ij}^{*} \leq k\right)}{\Pr\left(DL_{ij}^{*} \geq k\right)}\right] = \gamma_{ki} - \theta_{ij}, \mathrm{k} = 1,2 \tag{4}$$

In order to accommodate differences among subjects, the thresholds γ_{ki} were specified as random effects.

$$\gamma_{ki} = \gamma_k + b_i, \mathrm{k} = 1, 2 \tag{5}$$

Where the intercept γ_k represents a constant component for thresholds for all subjects. A random effect component b_i is formulated to accommodate the between-subject heterogeneities.

An Intra-class Correlation Coefficient (ICC) is normally defined to examine the proportion of specific subject-level variance:

$$\mathrm{ICC} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2} \tag{6}$$

Where σ_w^2 is within group variance and σ_b^2 is between group variance. A value of ICC close to zero indicates there is a very small variation between the different subjects, and a model without multilevel structure is adequate for the data. Otherwise, a multilevel model would be preferred.

(2) Artificial Neural Network Model

Artificial Neural Networks (ANN), a popular class of computational intelligence models, have been widely applied to drowsiness detection, partly because of their ability to work with massive amounts of multi-dimensional data, their modeling flexibility, and their generally good predictive ability.

In this study, we built a feed-forward neural network with one hidden layer consisting of the interconnection of neurons only between two adjacent layers. A back propagation training method was used. Before modeling, the following standardization procedure was carried out for each metric:

$$x_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{7}$$

A basic computational element is called a node. Each node receives input from an external source or from other nodes. Each input has an associated weight (w_{ij}), which can be modified to model synaptic learning by the process of training. The input of the j_{th} node in the hidden layer is calculated as follows:

$$Z_j = \sum_i w_{ij}\left(x_i + b_j\right) \qquad i = 1,2,\dots,N \quad j = 1,2,\dots,M \tag{8}$$

Where Z_j is the input to j^{th} node in the hidden layer, w_{ij} is the weight of i^{th} node in the input layer to j^{th} node in the hidden layer, x_j is the value of i^{th} node in the input layer, b_j is bias value for the j^{th} node in the hidden layer, N is the number of nodes in the input layer, and M is the number of nodes in the hidden layer.

The output of a node is decided by its input as well as the activation function. Different activation functions such as sigmoid functions, hyperbolic tangent functions, and logistic functions can be used. A hyperbolic tangent function was used as the activation function of the hidden layer in our study. It was calculated as follows:

$$H_j = f\left(Z_j\right) = \tanh\left(Z_j\right) = \frac{e^{2z_j} - 1}{e^{2z_j} + 1} \tag{9}$$

Where H_j is the output of j^{th} node in hidden layer.

In our study, we want the outputs of ANN to be interpretable as probabilities for a categorical target variable (DL), for those outputs to lie between 0 and 1, and to have a sum of 1. Therefore, a Softmax activation function is used for the output layer, which is written as follows:

$$O_k = \frac{\exp(Z_k)}{\sum_{m=1}^{c} \exp(Z_m)} \tag{10}$$

Where O_k is the output of k^{th} node in the output layer, c is the number of categories for the target variable.

In the training process, the network output, in general, may not be equal to the desired output. Therefore, the output error is calculated as the difference between the network output and the desired output. If the output error does not satisfy the tolerance level, the network modifies the connection weights (w_{ij}) according to the value of the output error; then, training data is inputted again to the network and the network output is calculated. The training cycle is continued until the network achieves the desired tolerance level.

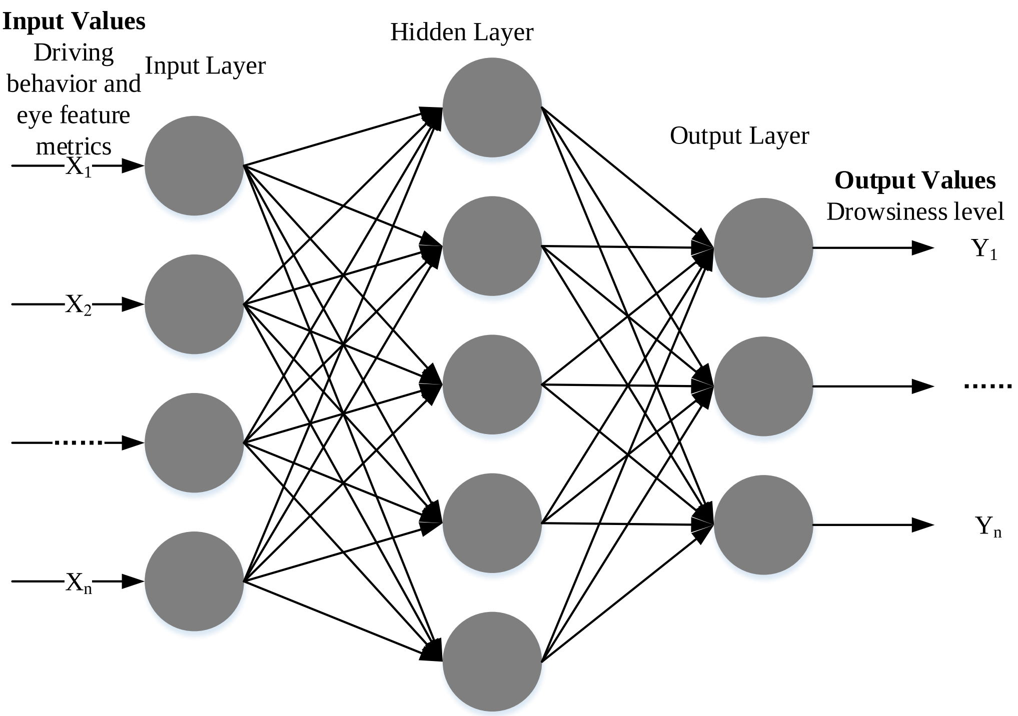

FIGURE 4 Configuration of multilayer ANN

3 RESULTS

In the MOL model development, beginning with the 23 variables, each variable was tested for the statistical significance and the insignificant ones were eliminated. Among the variables studied, five variables were identified as significant, as judged by 95% credible interval (CI): PERCLOS, Pupil, Blink duration, SWM_Re, LP_stdev. Then, among the significant variables, the Pearson correlation coefficients were examined. Blink durations were highly correlated with PERCLOS, but PERCLOS are more significantly related to DL. Therefore, in the final model, Blink duration was also eliminated.

The results of the MOL model are shown in TABLE 2. Two eye feature metrics, one steering variability metric and one lane variability metric are used as the explanatory variables in the final model. Among these explanatory variables, the fixed effects of PERCLOS, LP_stdev and SWM_Re are positive, while Pupil was negative. The threshold γ_1 is significant (t = -2.255, sig.=0.025) and γ_2 is insignificant (t = -0.427, sig.=0.669). For the random effects, ICC of the data set (training set) is 0.402, which shows a large between-group heterogeneity and within-group homogeneity. Therefore, it can be inferred that if an ordered logit model was implemented without considering the random effects between subjects, the results may be biased and inaccurate. It is also implied that the drowsiness detection algorithm shouldn't stay the same for different subjects.

The results of the OL model using the same explanatory variables with the MOL model are also shown in TABLE 2. All the explanatory variables are significant in 95% CI, and the coefficient for each variable shows different values but the same sign as that in MOL model. The threshold γ_2 is significant (t = 2.742, sig. = 0.006) and γ_1 is insignificant (t = -0.786, sig. = 0.432).

TABLE 2 MOL & OL Model Estimated Results

Parameters	Effect Estimate		t	sig	95% Confidence Level	
	Mean	S.D.			Lower	Upper
MOL						
Threshold						
γ_1	-3.778	1.675	-2.255	0.025	-7.070	-0.486
γ_2	-0.712	1.666	-0.427	0.669	-3.986	2.562
Fixed effects						
PERCLOS	5.226	2.129	2.455	0.014	1.043	9.410
Pupil	-1.780	0.318	-5.60	0.000	-2.403	-1.156
SWM_Re	0.010	0.005	2.182	0.030	0.001	0.020
LP_stdev	5.227	1.007	5.193	0.000	3.249	7.206
Random effects						
Between-subject variance	3.496					
Within-subject variance	5.195					
ICC	0.402					
OL						
Threshold						
γ_1	-0.545	0.694	-0.786	0.432	-1.907	0.816
γ_2	1.914	0.698	2.742	0.006	0.544	3.283
Fixed effects						
PERCLOS	9.891	1.441	6.865	0.000	7.063	12.719
Pupil	-1.065	0.121	-8.779	0.000	-1.303	-0.827
SWM_Re	0.011	0.003	3.212	0.001	0.004	0.017
LP_stdev	3.606	0.833	4.329	0.000	1.971	5.240

For ANN, PERCLOS, and Pupil, SWM_Re were standardized before input into the model. Three neurons are used in the hidden layer. Based on the ANN model, the importance of each variable were also calculated. The normalized importance of each variable is PERCLOS (100%), Pupil (74.5%), LP_stdev (65.2%), and SWM_Re (41.1%). This implies that eye feature metrics performed better than driving behavior metrics in drowsiness detection. The receiver operating characteristic (ROC) curve of ANN is also formed. The area under ROC curve for DL=3 (0.887) is larger than DL=1 (0.779) and DL=2 (0.647).

The summary of drowsiness detection accuracy is shown in TABLE 3. For each data set (train set and test set), the overall accuracy of the MOL model is the highest among the three models. However, the detection accuracy difference between the train set and test set for MOL is also the largest among the three models, which implies the MOL also has the largest generalization error. For the other two models, ANN performed better than the OL model in both the train set and test set. For the test set, the detection accuracy of MOL varies in a very small range across drowsiness levels (63.33% ~ 65.32%), while the other two models vary greatly (OL: 50.00%~55.95%; ANN: 47.92%~65.83%). The top three possible detection errors for the three models were the same, which are mistaking 1 for 2, mistaking 3 for 2 and mistaking 2 for 1. Mistaking 1 for 3 and mistaking 3 for 1 have the smallest possibility.

TABLE 3 Models Accuracy Summary

Model & Dataset	Observed	Predicted			Correct	Overall Accuracy
		1	2	3		
MOL (Train set)	1	67.90%	31.00%	1.20%	67.90%	68.40%
	2	22.80%	68.90%	8.30%	68.90%	
	3	1.90%	29.60%	68.50%	68.50%	
MOL (Test set)	1	64.17%	30.28%	5.56%	64.17%	64.15%
	2	27.13%	63.33%	9.54%	63.33%	
	3	5.00%	29.68%	65.32%	65.32%	
OL (Train set)	1	59.52%	39.29%	1.19%	59.52%	54.80%
	2	40.00%	51.67%	8.33%	51.67%	
	3	3.70%	43.52%	52.78%	52.78%	
OL (Test set)	1	50.00%	50.00%	0.00%	50.00%	52.70%
	2	30.83%	52.50%	16.67%	52.50%	
	3	1.19%	42.86%	55.95%	55.95%	
ANN (Train set)	1	54.17%	40.28%	5.56%	54.17%	57.80%
	2	26.67%	63.33%	10.00%	63.33%	
	3	9.09%	36.36%	54.55%	54.55%	
ANN (Test set)	1	47.92%	52.08%	0.00%	47.92%	56.04%
	2	30.00%	65.83%	4.17%	65.83%	
	3	12.20%	36.59%	51.22%	51.22%	

4 DISCUSSION

In order to find out a group of suitable indicators, twenty-three non-intrusive indicators were developed in this study. Eight metrics are based on lane lateral position, eight metrics are based on steering wheel angel, three metrics are based on vehicle speed and four metrics are based on eye features. The MOL results indicate that the indicators group formed by PERCLOS, Pupil, SWM_Re, and LP_stdev are appropriate to detect drowsiness. To test whether the variables selected in MOL are acceptable, all twenty-three variables were input into an ANN model with 10 neurons in hidden layers. The importance of each variable is shown in FIGURE 5. In order of importance, PERCLOS, Pupil, Blink_duration, SWM_Re, and LP_Stdev were the most important variables to predict the likelihood of DL. With the exception of Blink_duration, which was eliminated because of the correlation with PERCLOS, these variables are also significant in the MOL model.

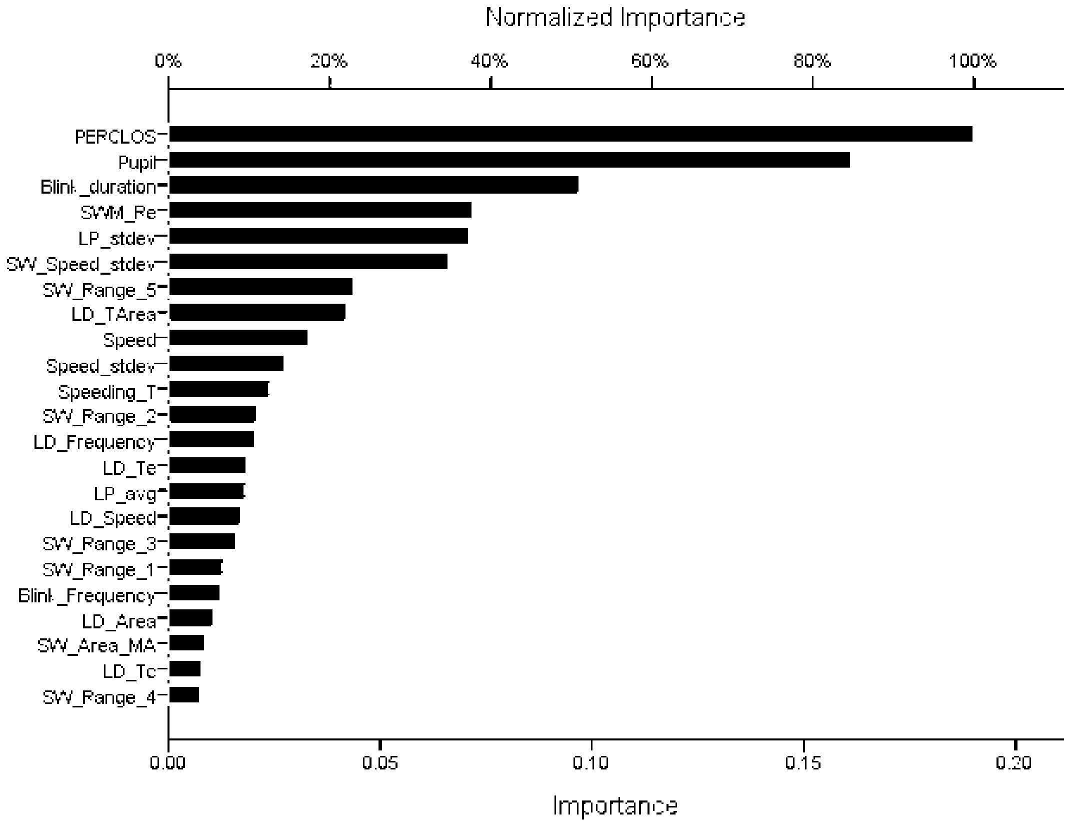

FIGURE 5 Variables importance in ANN with all the metrics

According to FIGURE 5, the top three important variables are all eye feature metrics. These results can be interpreted that eye feature metrics perform better than driving behavior metrics in drowsiness detection. The following test also verified this inference. After removing eye feature metrics, the detection accuracy of the ANN model for the test set was reduced to 45.8%. Meanwhile, keeping only eye feature metrics, the detection accuracy of the ANN model was reduced to 49.3%. Because of this, some drowsiness detection studies have used only eye feature metrics to detect drowsiness (*10; 16*). However, the detection accuracy of the ANN model using both metrics has the highest detection accuracy. Use of driving behavior metrics is also needed. Moreover, the eye features are often measured by cameral and image processing, which may not be reliable. The driving behavior metrics can be a supplement to increase the reliability of the detection system.

For driving behavior metrics, both lane related and steering related metrics are important in the models. Among lane related metrics, the lane variability measure (LP_stdev) performed better than other lane departure measures in drowsiness detection models. The possible reason is the lane departure metrics only measure the features of lane departure events and the lane variability information is missed for the non-departure parts. SWM_Re is the most important variable among steering related metrics, which measures steering variability. Rapid steering wheel movement is suggested to be a drowsiness measurement (*18*). In this study, it is measured by SWM_Rang_5 and is not significant in the MOL model. Moreover, some studies (*19*) concluded the lane variability is highly correlated with steering variability. We calculated the Pearson correlation of SWM_Re and LP_stdev at 0.089 (sig. = 0.059), which indicates the correlation between these variables is small.

In addition, previous research found there might be a curvilinear relationship between KSS and drowsiness metrics (9), with a stronger change at high KSS levels when compared with low KSS levels. In this study, among the four significant variables in the MOL model, the mean change between DL 3 and DL 2 are larger than that between DL 2 and DL 1 (see FIGURE 6). PERCLOS is a typical example and it is inferred to be sensitive to a high drowsiness level. This also might explain why it is the variable with the highest importance in the ANN model. Due to the larger change of metrics on higher DL, it can be inferred that higher drowsiness is easier to be detected, which is also verified by the detection accuracy of MOL and OL models. The correct detection rate for DL=3 is the highest.

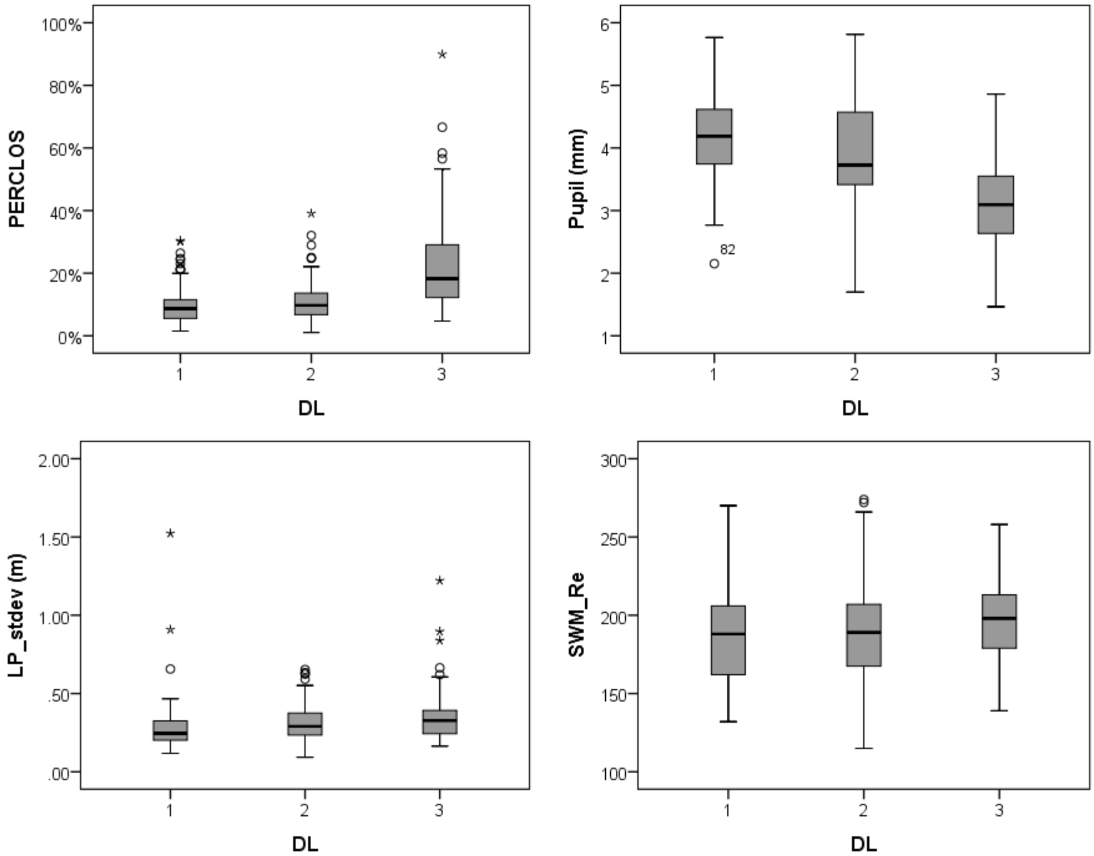

FIGURE 6 Boxplot of four important variables

In this study, the OL model can be considered as the basic model, while MOL and ANN models can be viewed as two improvements on the OL model. The MOL model's improvement is considering individual difference, while the ANN's improvement is insuring higher adaptability for the data. The results show MOL has the highest detection accuracy among the three models. This was attributed to using a series of individual specific thresholds, which were achieved by adding a random intercept in the subject level of MOL. That means the improvement by using an individual specific logit model is larger than by using a more complex algorithm with higher flexibility. However, in MOL, this random intercept for each driver is hard to predict. It can be decided only by a procedure like "training." Driving experience, gender, age, and other characteristic variables of the driver were attempted as explanations of individual differences, but no strong results were found. Yet the individual difference is still the main obstacle to increasing the drowsiness detection accuracy.

Some research that has classified drowsiness in only two levels (alert or drowsy) has achieved high detection accuracy. But two levels of drowsiness aren't enough to support warning the driver before the crash risk becomes critical. Therefore, three levels of drowsiness were used in this study. However, the detection accuracy is highly related with lower numbers of drowsiness levels. If we used two levels of drowsiness (Level 1: KSS 1~7; Level 2: 8~9), the detection accuracy increases from 64.15% to 88.6% for the MOL model with the same variables, and the detection accuracy increases from 56.04 to 83.3% for ANN. It can be inferred that the more drowsiness levels to be classified, the lower detection accuracy we would get in the models. Therefore, when comparing the detection accuracy among detection models, the way to classify drowsiness levels should also be considered.

5 CONCLUSION AND RECOMMENDATIONS

Twenty-three non-intrusive metrics including driving behavior and eye feature metrics were evaluated in a simulated shift work study with a motion based high-fidelity driving simulator in a controlled laboratory environment. Based on these metrics, MOL was developed to detect three-level drowsiness. For comparison, two non-individual specific models, OL and ANN, were also established.

MOL has the highest detection accuracy. This may be attributed to using a series of individual specific criteria, which was achieved by adding a random intercept at the subject level. Among the twenty-three variables, PERCLOS, Pupil, LP_stdev and SWM_Re were significant in MOL and OL, and were also confirmed in ANN. Metrics of eye features performed better (showed higher importance) in the drowsiness detection models than other metrics, which was also verified using ANN by comparing the detection accuracy between eye features only and driving behaviors only. We also found higher DL is more easily detected because of higher heterogeneity between adjacent DLs.

Based on the analysis, using a user-specific method to detect driver drowsiness is recommended in order to address the inaccuracies caused by individual difference. In this study, the group characteristics (like age, gender) of participants were controlled. However, if group characteristics exist, we can build group-specific models that would simplify the model training process by group determinants. Therefore group characteristics are recommended for study. Also, because establishment of drowsiness level in this research was subjective, more accurate measures should be applied, for example, using EEG to determine drowsiness level.

ACKNOWLEDGEMENTS

Supported by Road and Traffic Engineering Key Laboratory of Tongji University, Ministry of Education.

REFERENCES

1. NHTSA's National Center for Statistics and Analysis. *Traffic Safety Facts: A Brief*

Statistical Summary. NHTSA, U.S. Department of Transportation. DOT HS 811 449, 2011.

2. Maycock, G. Sleepiness and Driving: the Experience of UK Car Drivers. *Accident Analysis & Prevention*, Vol.29, No.4, 1997, No.453-462.
3. The Ministry of Public Security of the People's Republic of China, Road and Transport Authority. *Road Traffic Crash Statistics 2001-2008*, 2009.
4. Central Intelligence Agency, United States. *The world factbook*. 2010. Retrieved 2013.
5. Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., and Ramsey, D. J. *The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data*. No. HS-810 594, 2006.
6. Radun, I., Ohisalo, J., Radun, J., Rajalin, S., 2012. Law Defining the Critical Level of Driver Fatigue in Terms of Hours without Sleep: Criminal Justice Professionals' Opinions and Fatal Accident Data. *International Journal of Law, Crime and Justice*, Vol.40, No.3, 2012, pp.172-178.
7. Li, W., He, Q. C., Fan, X. M., and Fei, Z. M. Evaluation of Driver Fatigue on Two Channels of EEG Data. *Neuroscience Letters*, Vol.506, No.2, 2012, pp.235-239.
8. Patel, M., Lal, S. K. L., Kavanagh, D., and Rossiter, P. Applying Neural Network Analysis on Heart Rate Variability Data to Assess Driver Fatigue. *Expert Systems with Applications*, Vol.38, No.6, 2011, pp.7235-7242.
9. Ingre, M., Åkerstedt, T., Peters, B., Anund, A., Kecklund, G. Subjective Sleepiness, Simulated Driving Performance and Blink Duration: Examining Individual Differences. *Journal of Sleep Research*, Vol.15, No.1, 2006, pp.47-53.
10. Jo, J., Lee, S. J., Park, K. R., Kim, I. J., and Kim, J. Detecting Driver Drowsiness Using Feature-Level Fusion and User-Specific Classification. *Expert Systems with Applications*, Vol.41, No.4, 2014, pp.1139-1152.
11. Ingre, M., ÅKerstedt, T., Peters, B., Anund, A., Kecklund, G., and Pickles, A. Subjective Sleepiness and Accident Risk Avoiding the Ecological Fallacy. *Journal of sleep research*, Vol.15, No.2, 2006, pp.142-148.
12. Thiffault P. and Bergeron J. Fatigue and Individual Differences in Monotonous Simulated Driving. *Personality and Individual Differences,* Vol.34, No.1, 2003, pp.159-176.
13. Hamada, T., Ito, T., Adachi, K., Nakano, T., and Yamamoto, S. Detecting Method for Drivers' Drowsiness Applicable to Individual Features. *In proceeding of Intelligent Transportation Systems*, Vol.2, 2003, pp.1405-1410.
14. Wierwille, W. W., Wreggit, S. S., Kirn, C. L., Ellsworth, L. A., and Fairbanks, R. J. *Research on Vehicle-Based Driver Status/Performance Monitoring; Development, Validation, and Refinement of Algorithms for Detection of Driver Drowsiness*. DOT HS 808 247, 1994.
15. Yang, G., Lin, Y., and Bhattacharya, P. A Driver Fatigue Recognition Model Based on Information Fusion and Dynamic Bayesian Network. *Information Sciences*, Vol.180, No.10, 2010, pp.1942-1954.
16. Hu, S. and Zheng, G. Driver Drowsiness Detection with Eyelid Related Parameters by Support Vector Machine. *Expert Systems with Applications*, Vol.36, No.4, 2009,

pp.7651-7658.
17. Åkerstedt, T., and Gillberg, M. Subjective and Objective Sleepiness in the Active Individual. *International Journal of Neuroscience*, Vol.52, No.1-2, 1990, pp.29-37.
18. Sandberg, D. and Wahde, M. Particle Swarm Optimization of Feed Forward Neural Networks for the Detection of Drowsy Driving. *In Neural Networks*. IEEE International Joint Conference, 2008, pp.788-793.
19. Forsman, P. M., Vila, B. J., Short, R. A., Mott, C. G., and Van Dongen, H. Efficient Driver Drowsiness Detection at Moderate Levels of Drowsiness. *Accident Analysis & Prevention*, Vol.50, 2013, pp.341-350.