Expert Systems With Applications 85 (2017) 279–291

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers

Lan-lan Chen*, Yu Zhao, Peng-fei Ye, Jian Zhang, Jun-zhong Zou

Company: East China University of Science & Technology, Department: Department of Automation, B- 1102, Graduate Building, Meilong Road #130, Shanghai, 200237, China

ARTICLE INFO

Article history:
Received 25 October 2016
Revised 13 December 2016
Accepted 25 January 2017
Available online 1 March 2017

Keywords:
Driving stress
Physiological signals
Multimodal feature extraction
Feature selection
Kernel-based classifiers

ABSTRACT

Monitoring driving status has great potential in helping us decline the occurrence probability of traffic accidents and the aim of this research is to develop a novel system for driving stress detection based on multimodal feature analysis and kernel-based classifiers. Physiological signals such as electrocardiogram, galvanic skin response and respiration were record from fourteen drives executed in a prescribed route at real drive environments. Features were widely extracted from time, spectral and wavelet multi-domains. In order to search for the optimal feature sets, Sparse Bayesian Learning (SBL) and Principal Component Analysis (PCA) were combined and adopted. Kernel-based classifiers were employed to improve the accuracy of stress detection task. Analysis I used features from 10 s intervals of data which were recorded during well-defined rest, highway and city driving conditions to discriminate three levels of diving stress achieving an averaging accuracy over 99% at per-drive level and 89% in cross-drive validation. Analysis II made continuous stress evaluation throughout a complete driving test attaining a high coincidence with the true road situation especially at the switching interval of traffic conditions. Experimental results reveal that different levels of driving stress can be characterized by specific set of physiological measures. These physiological measures could be applied to in-vehicle intelligent systems in various approaches to help the drivers better manage their negative driving status. Our design scheme for driving stress detection could also facilitate the development of similar in-vehicle expert systems, such as driver's emotion management, driver's sleeping onset monitoring, and human-computer interaction (HCI).

1. Introduction

Some unfavorable mental states, e.g., fatigue (Cyganek & Gruszczyński, 2014; Lee & Chung, 2012; Liu, Lin, Wu, Chuang, & Lin, 2015; Yang, Lin, & Bhattacharya, 2010), sleepiness (Åkerstedt et al., 2013; Hallvig et al., 2013; Morris, Pilcher, & Switzer, 2015); stress (Healey & Picard, 2005; Tamrin et al., 2014; Zheng, Yamabe, Nakano, & Suda, 2015) and distraction (Craye, Rashwan, Kamel, & Karray, 2015; Horberry, Anderson, Regan, Triggs, & Brown, 2006) contribute to traffic accidents, leading to considerable number of vehicle crashes, injuries, and fatalities annually. In addition, physical discomfort such as extreme ambient temperature (Daanen, van de Vliert, & Huang, 2003; Pimenta & Assunção, 2015), uncomfortable driving position (Smith, Mansfield, Gyi, Pagett, & Bateman, 2015) and muscle fatigue in the neck/shoulder/back area (Hostens & Ramon, 2005) might impact driving behavior. Searching for the countermeasures to reduce the amount of traffic accidents and enhance public road safety has become an urgent issue for both governments and automakers. It is very crucial to develop an automatic system that intelligently detects the driver's unfit status and makes a warning once necessary.

For developing an automatic system to measure drivers' workload and detect their internal status, it is crucial to select effective measurements. Usually, we can classify the measures into three categories: (1) vehicle behavior measures; (2) video-based measures; (3) physiological measures.

Vehicle behavior mainly includes: vehicle speed, acceleration, lane position deviation, steering, braking and gear changes (Horberry et al., 2006; Morris et al., 2015; Tamrin et al., 2014; Wang & Xu, 2015). These measures are easy to obtain however they are strongly dependent on the type of vehicle and the handling skill of drivers. Additionally, some vehicle features such as lane tracking technology are prone to data loss under the situation of lane markers missing, bad weather or darkness (Wang & Xu, 2015).

Video-based detectors that track head movements, monitor eye gaze status and interpret facial expression have been widely ex-

* Corresponding author.
E-mail addresses: llchen@ecust.edu.cn, chenlanlan104@gmail.com (L.-l. Chen).

http://dx.doi.org/10.1016/j.eswa.2017.01.040
0957-4174/

plored (Azim, Jaffar, & Mirza, 2014; Cyganek & Gruszczyński, 2014; Garcia, Bronte, Bergasa, Almazán, & Yebes, 2012; González-Ortega, Díaz-Pernas, Antón-Rodríguez, Martínez-Zarzuela, & Díez-Higuera, 2013; Jo et al., 2014; Kholerdi, TaheriNejad, Ghaderi, & Baleghi, 2016; Zhang, Cheng, & Lin, 2012). However, visual features cannot always return reliable results especially in the conditions of poor light, night driving and wearing glasses. Besides, inaccurate interpretation of the facial expression might be encountered when some introverted subjects intend to control their emotions during the process of driving test.

Recently, features extracted from physiological signals such as electrocardiogram (EKG), galvanic skin response (GSR), respiration, electromyogram (EMG) and electroencephalogram (EEG) show relatively high identification accuracy and get insight into driver's states directly (Correa, Orosco, & Laciar, 2014; Healey & Picard, 2005; Lee & Chung, 2012; Lee, Lee, & Chung, 2014; Li & Chung, 2013; Li, He, Fan, & Fei, 2012; Liu et al., 2015; Liu, Zhang, & Zheng, 2010; Singh, Conjeti, & Banerjee, 2013; Vicente, Laguna, Bartra, & Bailón, 2016; Wang, Zhang, Shi, Wang, & Ma, 2015; Zhao, Zhao, Liu, & Zheng, 2012; Zhao, Zheng, Zhao, Tu, & Liu, 2011). The disadvantage of these measurements is that they require sensors and cables attached on the body, which constrains the behavior of drivers in some extent. And the wireless and wearable acquisition solutions allow subjects participate the drive task more conveniently and comfortably (Fu & Wang, 2014; Lee et al., 2014). These physiological measures could be used automatically by in-vehicle intelligent systems in various approaches to help the drivers better manage their negative driving status.

It is a challenging and rewarding work for consecutive detecting of driver states in real driving environments. Computational techniques such as feature extraction, feature selection & reduction, and classification have the capacity to determine optimal sensors and automate driving status recognition.

(1) Many feature extraction methods have been developed for physiological signals such as waveform information (Correa et al., 2014; Healey & Picard, 2005), power spectral analysis (Vicente et al., 2016), wavelet coefficients (Li & Chung, 2013), and nonlinear analysis (Chen, Zhao, Zhang, & Zou, 2015).
(2) In order to reduce the dimensionality of feature sets or select the optimal features from the primary measures, Principal Component Analysis (PCA) (Wang, Lin, & Yang, 2013) and Linear Discriminant Analysis (LDA) (Correa et al., 2014) are the most commonly used. Besides, Mutual Information (MI) methods and Sparse Bayesian Learning (SBL) have been employed in the feature selection of Brain-Computer Interface (BCI) system and obtained outstanding performance (Atkinson & Campos, 2016; Hoffmann, Yazdani, Vesin, & Ebrahimi, 2008).
(3) Several modeling techniques have been developed to detect driving status such as support vector machines (Zhao et al., 2011; Zhao, Xu, & Tao, 2009), artificial neural networks (Correa et al., 2014; Singh et al., 2013), fuzzy logic inference (Chiang, 2015) and Bayesian networks (Yang et al., 2010).

These computational techniques will be discussed and compared in the discussion part.

Commonly, stress is viewed as a response to particular events. It is a normal reaction of human body preparing itself in the face of difficulties with focus, strength and improved alertness. Usually, stress can be both physical and mental. If subjects are involved in a long, monotonous driving or under sleep deprivation condition, physical stress might become the major consideration. In this research, three well-defined driving conditions are designed mainly to induce different levels of mental stress. Furthermore, there is a difference between eustress and distress, where eustress is a positive stress helping individuals to better cope with the challenging situations. However, in this research, we focus on distress only, which is usually caused by increased drivers' workload. A lot of literature has reported the relationship between driving distress and the diminished vehicle handling abilities, reduced alertness, and increased probability of traffic accidents (Bakker, Pechenizkiy, & Sidorova, 2011; Chiang, 2015; Okada et al., 2013; Sharma & Gedeon, 2012; Sun et al., 2010; Villarejo, Zapirain, & Zorrilla, 2012).

Stressful events could cause dynamic changes in autonomic nervous system (ANS), reflected as increased activity in sympathetic system and decreased activity in parasympathetic system. Physiological signals such as galvanic skin response (GSR), electrocardiogram (EKG), and electroencephalogram (EEG) are the main measures used for stress detection in literature. Heart rate variability (HRV) is a common non-invasive indicator to detect ANS activities and is used as a primary measure for stress detection in many monitoring systems (Chiang, 2015; Okada et al., 2013). HRV can dynamically reflect the accumulation of mental workload which makes it a good estimator for stress level. Short-term reduced HRV indicates acute stress, reflecting that stressful events negatively affect HRV. Skin conductance is another reliable measure for stress. Increased skin conductance reflects the individual under stress and reduced skin conductance indicates the individual under less-stressful situation (Bakker et al., 2011; Sharma et al., 2012; Sun et al., 2010; Villarejo et al., 2012). Other research demonstrates that relationships exist between brain activities and emotional stress. Stress assessment system has been developed using a new fusion link between EEG and peripheral signals such as blood volume pulse, skin conductance and respiration (Hosseini, Khalilzadeh, Naghibi-Sistani, & Homam, 2015).

Monitoring driving status has great potential in helping us decline the occurrence probability of traffic accidents. However, automatic detection is usually restricted to the stage of laboratory research mainly due to signal's features and noise, measurement constraints and subject-dependent issues. The aim of this paper is to develop an efficient system to determine driver's relative stress level during real-world driving tasks based on the analysis of physiological data. A novel system combining multimodal feature analysis and kernel-based classifiers is proposed. Considering the test differences between individual drivers and within individual drivers, two kinds of analysis are made: Analysis I used features from short intervals of well-defined data to discriminate three levels of diving stress at per-drive and cross-drive levels respectively. Analysis II made continuous stress evaluation throughout a complete driving test for each drive.

It is expected that: 1) different levels of driving stress could be characterized by specific set of physiological measures; 2) more features drawn from multimodal analysis are favorable for assessing the drive status more reliably and accurately; 3) the system combing efficient feature selection methods and kernel-based classifiers could enhance the capacity for identification of driver states.

In rest of this paper, Section 2 describes the materials including experiment design, participant information and data collection. Section 3 gives an introduction of our methodology including feature extraction, feature selection & reduction, and kernel-based classification technique. Section 4 presents our experimental results. Section 5 reviews the similar work and discusses the contributions and limitations of the present study. And the conclusion has been made in Section 6.

2. Materials

2.1. Experiment design

We use the database, contributed to PhysioNet by Healey & Picard, which collected a set of multiple physiological recordings

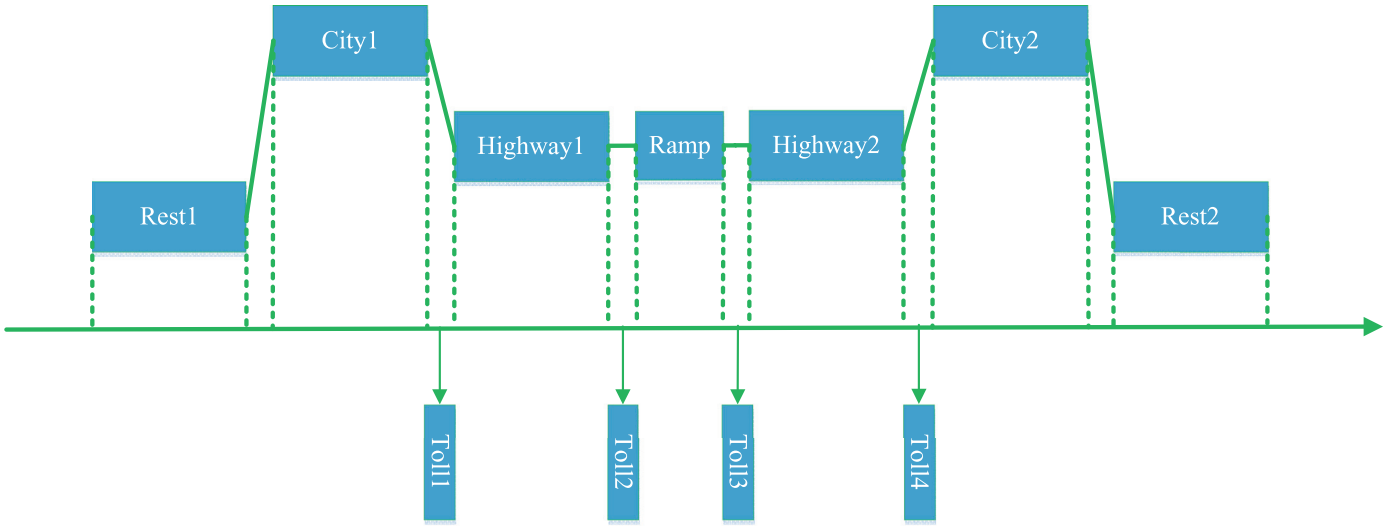

Fig. 1. Driving task.

from healthy subjects when they were driving along with a specified route in and around Boston.

The prescribed route was designed to make the drivers experience the situations which might produce different degrees of stress and also restore the stressful events encountered in daily commute. Fig. 1 shows the whole driving task which is mainly divided into six sections: Rest1, City1, Highway1, Highway2, City2, and Rest2. The total duration of the drive varied from about 50 min to 1.5 h, relying on the road conditions. The rest periods were designed as the baseline of the whole driving process. Actually, subjects cannot be completely free from stress at the very beginning and end of each test, so the rest period can be viewed as a relatively low-stress condition. In the city sections, the subjects drove at a busy main street in downtown area where drivers should frequently handle stop-and-start traffic conditions and cope with unpredicted emergencies probably created by jaywalkers and cyclists. This traffic condition easily led to a high-stress situation. All drives were implemented at midmorning or midafternoon to avoid congestion on the highway, thereby the subjects experienced a medium-stress condition during the uninterrupted highway driving. Besides, drivers need to be concentrated on some special events such as in-and-out toll, on-and-off ramp, which were also marked as high-stress situations, were not included in Analysis I but Analysis II.

2.2. Participant information

Nine subjects with valid driving licenses were recruited and were informed to having physiological signals and video recording during the experiments. Before each task, drivers were required to be familiar with the specific drive route and other instructions such as speed limits.

Totally 27 driving runs were accomplished. Six drives were executed by six different subjects and the left twenty-one drives were carried out by three subjects who followed the same route on different days. Of the initial 27 drives collected by Healey & Picard, 17 driving runs were available in public. One data record was incomplete due to the unstuck hand skin conductance sensor; two data records could not be analyzed because the markers for each driving condition were not provided. Finally, fourteen drives with complete data sets and clear markers were used in this study.

2.3. Data collection

Originally, the experiment designers, i.e., Healey and Picard (2005), collected contact (physiological) and contactless (video/behavioral) information to index stress during real world driving. Additionally, subjective questionnaires were used to assess the accuracy of the physiological measures. The data providers only supplied the physiological data in public. Although the raw data of video and subjective measures are not accessible, the results of both methods from the literature (Healey & Picard, 2005) could validate the assumptions that periods of rest, highway, and city driving produce low, medium, and high levels of stress.

In the following, we simply introduce the video and subjective measures, and only use them to validate the reasonability of the experiment design. The main focus of this research is to determine driver's relative stress level based on physiological data.

2.3.1. Subjective questionnaire

Drivers received subjective questionnaire immediately after each drive. Driving Questionnaires was designed by the data provider, i.e., Healey and Picard (2005), which included three parts, i.e., background Questions, today's driving experience and general questions. Two kinds of scoring were used: a free scale for stress feelings and a forced rating for stressful events. The results from both ratings show that subjects view the rest period as the least stressful condition, the highway task as more stressful situation, and the city drive as the most stressful task (Healey & Picard, 2005).

2.3.2. Videotape recording

Three digital cameras were used for videotape scoring: a small Elmo camera was placed on the steering wheel, a Sony digital video camera with a wide angle (0.42) lens was set up on the dashboard, and a third camera was used for monitoring event. Two video coders scored these video recordings based on the occurrence frequency of some typical behaviors in both vehicles and drivers. These observable stress indicators included vehicle turning, scram, collision and driver's head turning and gaze changes. Usually, eye gaze provides rich indices about human's attention focus and enables deducing mental status and intentions (Eckstein, Guerra-Carrillo, Singley, & Bunge, 2016). The spatial distribution and time/frequency characteristics of gaze are highly correlated with stress levels (Healey & Picard, 2005). In addition,

Table 1
The sampling frequency of the acquired physiological signals.

Signals	Abbreviation	Sampling frequency
Electrocardiogram	EKG	496 Hz
Hand Galvanic Skin Response	HGSR	31 Hz
Foot Galvanic Skin Response	FGSR	31 Hz
Respiration	RESP	31 Hz

stress models have been explored based on gesture and interaction features considering that real-driving environments are rich in stressors that induce drivers to react by gestures (Sharma & Gedeon, 2012). Gesture recognition has been applied to identify stress in car drivers (Liao, Zhang, Zhu, & Ji, 2005). Finally, the results from videotape scoring give a good support for the design assumptions that rest, highway and city conditions induce low, medium and high stress levels (Healey & Picard, 2005).

2.3.3. Physiological data collection

Four channels of physiological signals were analyzed: electrocardiogram (EKG); galvanic skin response (GSR, also known as skin conductivity); and respiration (through chest cavity expansion). Physiological sensors were connected to a FlexComp analog-to-digital converter, which was linked to an embedded computer in a modified Volvo S70 series car. The EKG electrodes were placed according to a modified lead II configuration which could effectively reduce the artifacts and enhance the R-waves. The skin conductance was recorded in two locations: on the palm of the left hand namely as HGSR and on the insole of the left foot namely as FGSR. Respiration was measured through chest cavity expansion with an elastic Hall Effect sensor strapped around the subject's diaphragm. The sampling frequency of each signal is listed in Table 1.

2.4. Data preparation

This experiment was designed to monitor driver's status in real driving environments under daily commuting conditions. The experiments performed in real traffic situations enable the results have more practical applicability. However, it may encounter many constraints such as data acquisition and subject-dependent issues. Considering these constraints, two types of analysis were implemented on the collected data.

Analysis I was designed to distinguish three levels of stress using typical segments drawn from rest, city, and highway periods. From each drive test, we selected a total of 30 min well-defined data, evenly distributed among low, medium and high stress driving periods. The low-stress segments were captured from the last 5 min of the rest conditions giving the subjects enough relaxation time. The high-stress segments were taken after the drivers enter into the main street and their bio-signals started to present obvious variation compared with the waveform in rest conditions. The medium-stress segments were selected from uninterrupted highway driving between two tolls. Each segment lasts for 100 s with an overlap of 10 s. Thus for each drive, 42 low stress samples, 42 medium stress samples and 42 high stress samples were obtained to construct a total sample size of 126 segments. For all 14 drives, the total sample set includes $126 * 14 = 1764$ segments.

Analysis II was designed to make a continuous assessment for drivers' internal states which could present how driving conditions affect drivers' stress level simultaneously. For a complete drive test, the continuous recordings were divided into 100 s length segments advanced by 10 s.

The segment length should be chosen with caution. It is true that a small segment length can reflect the subtle change of mental status however it is noticed that the shortest time for autonomic signals such as GSR and HRV to present meaningful variation is 100 s (Healey & Picard, 2005). Therefore, we set sample length to 100 s with an advance step of 10 s.

3. Methodology

The general block diagram of the proposed system is shown in Fig. 2. Physiological signals are acquired by multi-sensors. After the filtering and time segmentation, features extraction is executed including the wavelet decomposition, the time and the spectral analysis. The sparse Bayesian learning (SBL) is employed to select the optimal feature sets. To further reduce the feature dimensionality, principal component analysis (PCA) is utilized. Finally, two kernel-based classifiers, i.e., Support Vector Machine (SVM) and Extreme Learning Machine (ELM) with several different kernel functions are adopted to identify different drive stress levels.

3.1. Feature extraction

After initial noise filtering, each segment is analyzed in the time, frequency and wavelet domains.

3.1.1. Time analysis

Normalized mean N_{mean} and standard deviation σ are calculated for all four signals: the normalization process is implemented by subtracting the minimum of baseline and dividing by the range of baseline.

In the case of skin conductivity, more complex features, describing the orienting responses in a time segment, have been used.

Firstly, the correct peak position should be detected. We design an algorithm to detect the peak in GSR segments by first detecting slopes surpassing a specified threshold and then marking the local maximum. Compared with original algorithm provided by MATLAB, our algorithm makes more strict limitation on the slope and length of each rising edge. And our detection results are more close to the results of visual inspection as show in Fig. 3. Furthermore, four additional features, i.e., sum of the peak number, sum of the startle magnitude, sum of the rising duration, and sum of the rising area are calculated to describe the shape characteristics of GSR time series. The definition of these features are described in Table 2 and illustrated in Fig. 4.

3.1.2. Frequency analysis

In frequency domain, spectral features are computed for the respiration and electrocardiogram signals.

As for the respiration signal, four FFT-based spectral features are calculated to represent the energy in each band. Power spectrum is obtained using 1024-point Fourier transform and four spectral features are computed by summing the spectrum power in the bands 0–0.1, 0.1–0.2, 0.2–0.3, and 0.3–0.4 Hz, respectively.

Heart rate (HR) time series are obtained by extracting the R-R intervals from the electrocardiogram (EKG) signals. Then periodogram is applied to analyze the power spectrum of the heart rate time series. The power energy in the low frequency (LF) range (0–0.08 Hz) and in the high frequency (HF) range (0.15–0.5 Hz) are computed. Usually, low frequency components below 0.1 Hz can reflect how sympathetic system modulates heart rate and all frequency components ranging from 0 to 0.5 Hz (AF) can present how parasympathetic system affects heart rate. In this research, power sum in LF, power sum in HF, ratio of LF/HF, ratio of LF/AF, and ratio of HF/AF are estimated.

3.1.3. Wavelet decomposition

All four signals are decomposed into five scales using Daubechies 4 (db4) wavelet family because it provides better classification consequence than other common wavelets. The normalized mean value N_w and standard deviation σ_w for each sub-band signal $\{d_w$,

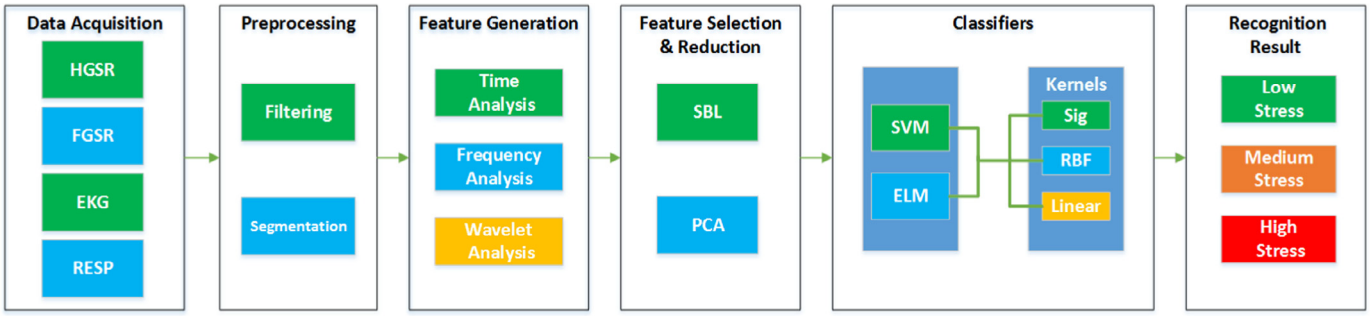

Fig. 2. General block diagram of the proposed system.

Table 2
Definition and descriptions for features.

Analysis	Features	Equation & description	Signals
Time	Normalize mean	Normalized *signals*:$S_N(i) = \frac{s(i)-\min(s_{base})}{\max(s_{base})-\min(s_{base})}$ Normalized mean :$N_{mean} = \frac{1}{N}\sum_{i=1}^{N} S_N(i)$	All
	Standard deviation	$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(S_N(i) - N_{mean})^2}$	All
	Peak number sum	Peak number = Number of peaks in a given segment	FGSR & HGSR
	Magnitude sum	Peak magnitude = GSR value at peak − GSR value at point of onset	FGSR & HGSR
	Duration sum	Duration = Time of occurrence of peak−Time of point of onset	FGSR & HGSR
	Area Sum	Area = Peak magnitude × Duration × 0.5	FGSR & HGSR
Spectral	Band power	$S = \sum P(f)$. $f1$: 0 − 0.1Hz, $f2$: 0.1 − 0.2Hz, $f3$: 0.2 − 0.3Hz, $f4$: 0.3 − 0.4Hz	RESP
	LF_power_sum	$S_{LF} = \sum P(f)$. f: 0 − 0.08Hz	HR
	HF_power_sum	$S_{HF} = \sum P(f)$. f: 0.15 − 0.5Hz	HR
	LF/HF	$Ratio_{LF/HF} = \frac{S_{LF}}{S_{HF}}$	HR
	LF/AF	$Ratio_{LF/AF} = \frac{S_{LF}}{S_{AF}}$	HR
	HF/AF	$Ratio_{HF/AF} = \frac{S_{HF}}{S_{AF}}$	HR
Wavelet	Wavelet mean Wavelet standard	$N_w = \frac{1}{N}\sum_{i=1}^{N} \lvert d_w(i)\rvert, w = 1, 2, ..., 5$	All
	Deviation	$\sigma_w = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \lvert d_w(i)\rvert - N_w)^2}$, $w = 1, 2, ..., 5$	All

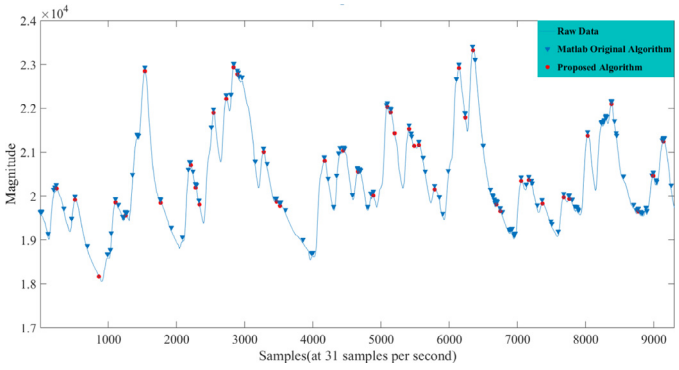

Fig. 3. A sample of peak detection in skin conductance signal.

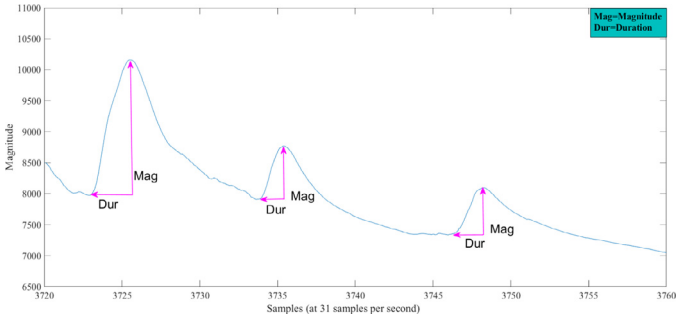

Fig. 4. A sample of shape features in skin conductance signal.

$w = 1, 2, ..., 5\}$ are computed. The wavelet coefficients are not used directly but the absolute value is used instead. The normalization process is similar to that in the time analysis.

Totally 73 features, which including 18 FGSR features, 18 HGSR features, 18 RESP features and 19 EKG features, are used to form a single vector denoting each segment for the identification analysis.

3.2. Feature selection and reduction

3.2.1. Sparse Bayesian learning (SBL)

In order to select the most optimal feature set for the subsequent classification, we apply sparse Bayesian learning (SBL) which shows its advance in estimating discriminant vectors with automatically determined sparsity. Recently, Bayesian-based modeling such as Bayesian LDA (BLDA) and sparse Bayesian learning (SBL) have been successfully applied as feature selection methods (Hoffmann et al., 2008; Tipping, 2001; Zhang et al., 2015). However, BLDA utilizes a standard Gaussian prior which does not result in a sparse discriminative vector. Therefore, the BLDA may encounter some difficulty in selecting the most significant features for dimensionality reduction. While the SBL algorithm learns a sparse discriminant vector through hierarchical Bayes modeling of a separate Gaussian prior. The sparsity degree is automatically and quickly determined based on training set under a Bayesian evidence framework. The general description for SBL algorithm is as follows:

Let $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N \in R^D$ be N samples with a feature dimensionality of D. Consider the following linear model:

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \varepsilon \quad (1)$$

where $\mathbf{y}$ is the target vector containing the class labels, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N]^T \in \mathrm{R}^{N\times D}$, and $\boldsymbol{\varepsilon}$ is the zero-mean Gaussian noise with variance σ^2. The likelihood function for weight $\mathbf{w}$ can be denoted as

$$p(\mathbf{y}|\mathbf{w},\sigma^2) = (2\pi\sigma^2)^{-N/2}\exp\{-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{X}\mathbf{w}\|_2^2\} \tag{2}$$

In order to automatically reduce the feature dimensionality and improve the classification accuracy, a sparse projection vector has been proposed. A zero-mean separate Gaussian prior in a probabilistic model is used to describe the penalty function

$$p(\mathbf{w}|\boldsymbol{\alpha}) = \prod_{i=1}^{D} N(w_i|0,\alpha_i^{-1}) = \prod_{i=1}^{D}(\frac{\alpha_i}{2\pi})^{\frac{1}{2}}\exp(-\frac{1}{2}\alpha_i w_i^2) \tag{3}$$

where $\alpha_i (i = 1, ..., D)$ separately controls the inverse variance of weight $w_i (i = 1, ..., D)$. According to the Bayesian rule with the likelihood function given in (2), the mean and covariance of the posterior $p(\mathbf{w}|\boldsymbol{\alpha}, \sigma^2, \mathbf{y})$ can be estimated by:

$$\boldsymbol{\mu} = \sigma^{-2}\boldsymbol{\Sigma}\mathbf{X}^T\mathbf{y} \tag{4}$$

$$\boldsymbol{\Sigma} = (\sigma^{-2}\mathbf{X}^T\mathbf{X} + \boldsymbol{\Lambda})^{-1} \tag{5}$$

where $\boldsymbol{\Lambda} = diag([\alpha_1, \ldots, \alpha_D])$.

To measure the hyper-parameters $\boldsymbol{\alpha}$ and σ^2, the marginal likelihood $p(\mathbf{y}|\boldsymbol{\alpha}, \sigma^2)$ is maximized to construct the following iterative regulations:

$$\alpha_i \leftarrow \frac{\gamma_i}{\mu_i^2} \tag{6}$$

$$\sigma^2 \leftarrow \frac{\|\mathbf{y}-\mathbf{X}\boldsymbol{\mu}\|_2^2}{N - \Sigma_{i=1}^{D}\gamma_i} \tag{7}$$

where μ_i is the ith posterior mean calculated by (5), and $\gamma_i = 1 - \alpha_i\Sigma_{ii}$, where Σ_{ii} is the ith diagonal entry of the posterior covariance obtained in (4) with the current $\boldsymbol{\alpha}$ and σ^2 values.

Then given a new test sample $\bar{\mathbf{x}}$, predictions can be made for the corresponding target $\bar{y}$, in terms of the predictive distribution:

$$p(\bar{y}|\alpha,\sigma^2,\bar{\mathbf{x}},\mathbf{y}) = \int p(\bar{y}|\mathbf{w},\sigma^2,\bar{\mathbf{x}})p(\mathbf{w}|\alpha,\sigma^2,\mathbf{y})\mathrm{d}\mathbf{w} \tag{8}$$

which is again Gaussian with mean and variance.

$$\bar{\mu} = \boldsymbol{\mu}^T\bar{\mathbf{x}} \tag{9}$$

$$\bar{\sigma}^2 = \sigma^2 + \bar{\mathbf{x}}^T\boldsymbol{\Sigma}\bar{\mathbf{x}} \tag{10}$$

3.2.2. Principal component analysis (PCA)

Principal Component Analysis (PCA) is used for reducing the dimensionality of feature sets or removing unwanted features from the set of primary measures. PCA has been applied to reduce features in ECG signals to identify stress level (Wang et al., 2013). In general, PCA is employed to search for a subspace where basis vectors correspond to the maximum-variance directions in the original space.

In this research, we first select the best feature sets based on SBL and further reduce the feature dimensionality using PCA.

3.3. Classification

In this work, we compare two kernel-based classifiers, i.e., Support Vector Machine (SVM) and Extreme Learning Machine (ELM). They are briefly described in this section. SVM is implemented using LIBSVM, which is an integrated software for support vector classification and is available at http://www.csie.ntu.edu.tw/~cjlin/libsvm/. Extreme Learning Machine (ELM) is introduced by Huang, Zhu, and Siew (2006) and its source code is accessible at http://www.ntu.edu.sg/home/egbhuang/.

3.3.1. Support vector machine (SVM)

SVM constructs a separating hyper-plane which maximizes the margin between the input data classes. For a 2-class problem, two parallel hyper-planes are established to calculate the margin from training samples, one on each side of the separating hyper-plane (Cortes & Vapnik, 1995). Support Vector Machine classifiers (SVM) have been applied into drive drowsiness and sleep-onset detection systems based on physiological measures such as eye movement (Jo, Lee, Park, Kim, & Kim, 2014) and combination of principal measures, for example, the combination of respiration and EEG (Lee et al., 2014). In this research, SVM with three kinds of kernel functions are used and compared, i.e., sigmoid, linear and radial basis function (RBF).

3.3.2. Extreme learning machine (ELM)

Extreme Learning Machine (ELM) is developed on the basis of the single-hidden layer feed-forward neural networks (SLFN) (Huang et al., 2006). Usually, neural networks cost large computational resource while ELM can accomplish learning extremely fast because ELM randomly selects the input weights and hidden layer biases only determining the output weights analytically. Huang (2014) extends ELM into kernel case which randomly chooses the centers and impact widths of kernels and simply assigns the output weights without complex iterative adjustment. ELM has been designed into automatic identification system of drowsy/alert status in daily mental working conditions based on EEG and EOG recordings (Chen et al., 2015). In this experiment, we test three kinds of ELM construction, i.e., ELM with sigmoid activation function, ELM with linear and RBF kernels.

The performance of different classifiers is evaluated by the estimators of precision (Prec.), sensitivity (Sens.), and specificity (Spec.). To compute these estimators, true positives (t_p), true negatives (t_n), false positives (f_p) and false negatives (f_n) should be measured first. For the multi-class classification, we divide the problem into three 2-class classifications. In each 2-class problem, we view one kind of road condition as positive case and the other two conditions as negative cases. The final estimators are calculated by taking the average among three classifications.

$$precision = \frac{t_p}{t_p + f_p} \tag{11}$$

$$sensitivity = \frac{t_p}{t_p + f_n} \tag{12}$$

$$specificity = \frac{t_n}{t_n + f_p} \tag{13}$$

The classification performance is evaluated at inter-drive and cross-drive two levels.

Inter-drive evaluation: As introduced earlier, each drive contributes to 126 segments, i.e., 42 rest segments, 42 city segments and 42 highway segments. We select two segments per road condition as test samples and the remaining 120 segments as training samples. This procedure is repeated 21 times to ensure that every segment is tested. Then the results are averaged as final performance.

Cross-drive evaluation: Note that totally 14 drives are analyzed. The data from 13 drives are used to train a classifier and then evaluate the classifier on the remaining one drive. This procedure is repeated for all 14 drives. The cross-drive validation allows us to evaluate the adaptability of previously trained classifiers on new participant. The generalized cross-drive evaluation procedure is illustrated in Fig. 5. The dataset is used to test the different classifiers.

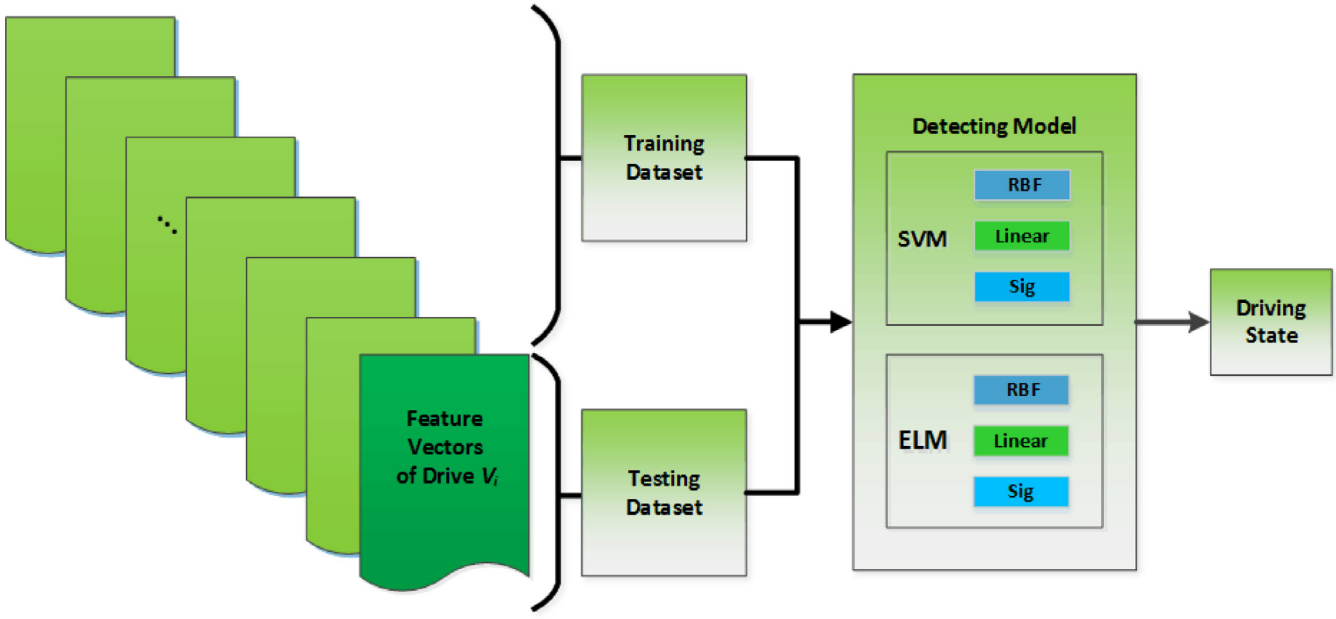

Fig. 5. Detecting algorithm for driving stress with leave-one-drive-out cross validation procedure, where V_i represents the feature vectors of the ith drive used as the testing set, whereas the withheld drives are used for the training set.

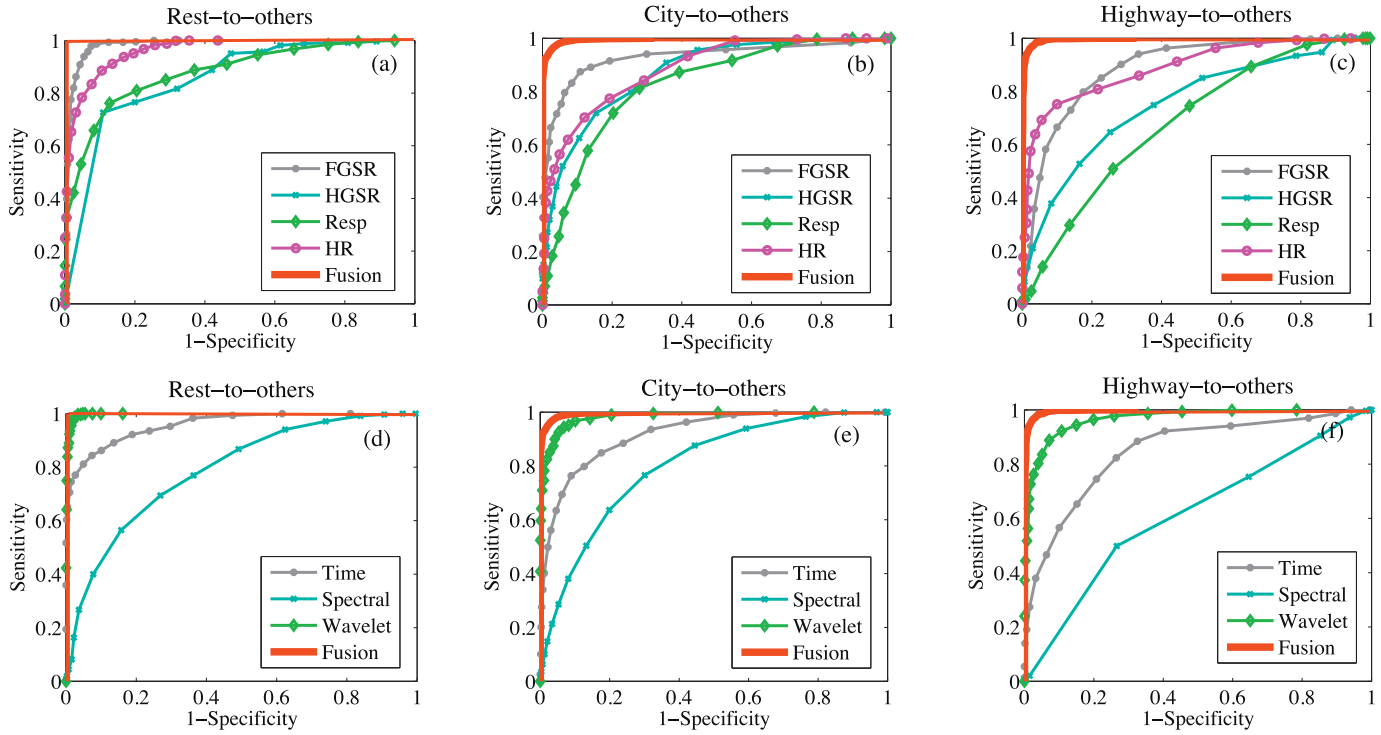

Fig. 6. ROC curves for each bio-signals and the fusion of all signals (a–c); ROC curves for time, spectral and wavelet features and the fusion of multi-domain features (d–f).

4. Results

4.1. Feature analysis with ROC

Receiver operating characteristic (ROC) curve is usually applied as a visual approach to demonstrate the relationship between true positive rate and false positive rate along with the change of a threshold parameter. Three driving conditions, i.e., rest, city and highway, were divided into three 2-class cases, i.e., rest to others, city to others, and highway to others. Fig. 6(a–c) are ROC curves for each bio-signals and the fusion of all signals; Fig. 6(d–f) are ROC curves for time, spectral and wavelet features and the fusion of multi-domain features.

Furthermore, a quantitative measure, i.e., the area under ROC curve (ROC_AUC) is computed to make comparisons among various signals and features (see Tables 3 and 4). For example, AUC_R was calculated when rest status was viewed as positive class while city and highway were negative. In the case of AUC_C and AUC_H, city and highway were regarded as positive classes, respectively.

Fig. 6(a–c) and Table 3 show that, considering the effect of single signals, AUC of foot skin conductivity is the largest, and the

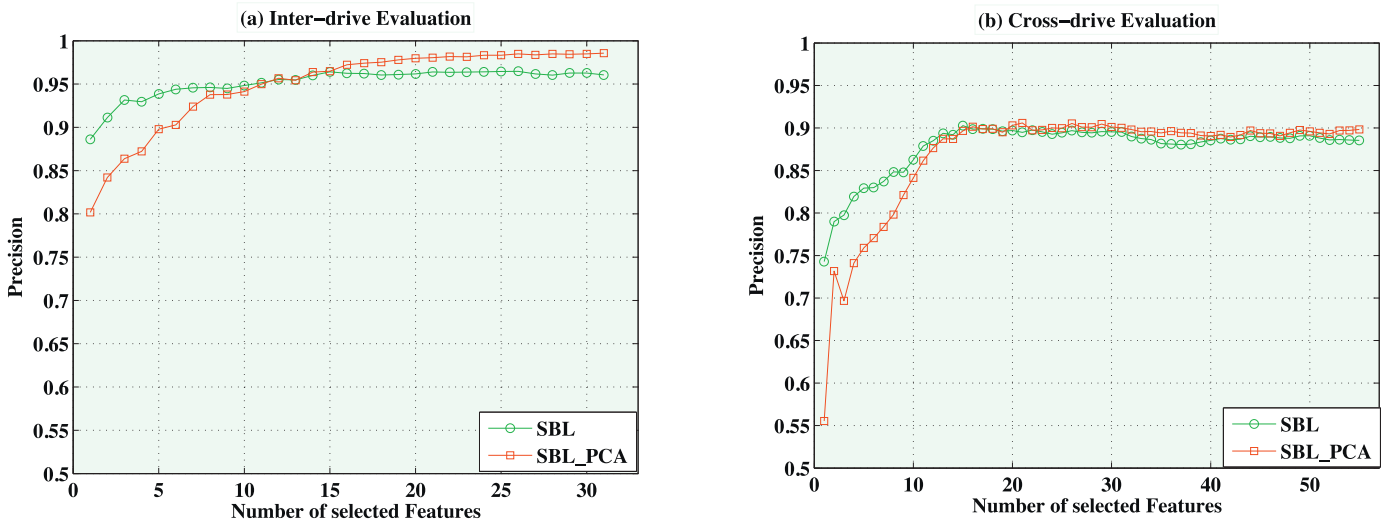

Fig. 7. Variation of precision on different number of selected features using SBL and SBL_PCA methods.

Table 3
ROC_AUCs (the area under ROC curve) for the different signals.

	FGSR	HGSR	RESP	HR	Fusion
AUC_R	0.984	0.903	0.881	0.963	1.000
AUC_C	0.931	0.876	0.829	0.890	0.997
AUC_H	0.890	0.759	0.687	0.834	0.997

Table 4
ROC_AUCs (the area under ROC curve) for the features in different domains.

	Time	Freq	Wavelet	Fusion
AUC_R	0.961	0.789	0.997	1.000
AUC_C	0.921	0.802	0.984	0.996
AUC_H	0.849	0.616	0.968	0.997

AUC of heart rate is the next. Besides, fusion of four signals could enhanced total classification capacity.

Fig. 6(d–f) and Table 4 show that the wavelet features generated larger AUC than that of the time and spectral features. Furthermore, fusion of time, frequency and wavelet features obtained stronger discriminative power than those using features of single domain.

4.2. Feature selection with SBL and PCA

An automatic feature selection from the full feature set was done by the sparse Bayesian learning (SBL). The mean value of optimal selected feature numbers for per-drive test and cross-drive test were 31.6 and 56.4 respectively (i.e., 68.4% and 43.6% features have been eliminated). After further transformation of PCA, the number of variables used was further reduced. The cumulative percents of the first 4 and 13 principal components have exceeded 90% at per-drive and cross-drive level respectively. Fig. 7 illustrates the variation of classification precision on the number of selected features using SBL and SBL_PCA two feature selection strategies. The mean precision of fourteen drives computed by SVM classifier with RBF kernel was used for illustration. When the selected feature number was small, SBL presented higher precision than SBL_PCA combined method. With the increase of selected features, SBL_PCA showed better identification performance especially at inter-drive evaluation.

4.3. Evaluation of different classifiers

To improve the classification performance as much as possible, two classifiers, i.e., SVM and ELM with three different kernel functions were utilized. The classifier performance tested on the original full feature set, the selected feature set with SBL, and the selected feature set with SBL_PCA respectively. The inter-drive and cross-drive results were listed in Tables 5 and 6 separately.

At inter-drive level: SVM with linear kernel achieved the best performance with a precision of 99.9% based on the full feature sets. SVM with RBF and sigmoid kernels obtained improved performance by using SBL_PCA combined selection methods.

At cross-drive level: SVM with RBF and sigmoid kernels got improved identification results by using SBL_PCA combined selection methods. The results obtained by ELM classifiers presented the same tendency. Finally, SVM with RBF kernel achieved the highest precision of 89.7%.

Obviously, the detection accuracy for per-drive level was higher than that of cross-drive level. Usually, cross-drive validation is important to ascertain whether the trained model could be applicable on unseen subjects or not. As seen in Table 6, the precision for cross-drive validation was 85.8%–89.7% with different classifiers. Note that the accuracy of our system was computed using training and test segments which were taken from different road conditions. Under this background, it is hard to ascertain that each segment drawn from the specific road condition corresponds to a unique stress degree. Levels of stress may change over time depending on the driving circumstances, individuals and testing dates. Moreover, our experiment was designed to provide relatively different stress levels created under different driving conditions. For instance, during rest conditions, the subjects were not completely free from stress. Initially, they might feel uncomfortable when wearing bio-signal sensors or feel uneasy under the monitoring of camera. Therefore, rest period was designed to create a relatively lower-stress situation just as city condition was designed to create a relatively higher-stress situation. Hence, the correct expectation for a good classifier is that a significant majority of segments in a specific road condition are correctly identified. Our cross-validation results could achieve this expectation.

4.4. Continuous analysis

Figs. 8 and 9 show the automatic detection of stress level during a continuous drive task accomplished by subject #6 and #11.

Table 5
Classifier performance comparison at inter-drive evaluation level.

		Full_featureset			SBL_featureset			SBL_PCA_featureset		
Feature number		73			31.6 ± 9.1			4.3 ± 1.3		
Classifiers	Kernels	Prec.	Sens.	Spec.	Prec.	Sens.	Spec.	Prec.	Sens.	Spec.
SVM	rbf	0.966	0.959	0.979	0.963	0.954	0.977	***0.985***	***0.984***	***0.992***
	lin	***0.999***	***0.999***	***0.999***	0.997	0.997	0.999	0.985	0.984	0.992
	sig	0.955	0.940	0.970	0.944	0.925	0.962	***0.977***	***0.974***	***0.987***
ELM	rbf	***0.979***	***0.976***	***0.988***	0.967	0.963	0.982	0.957	0.951	0.976
	lin	***0.998***	***0.998***	***0.999***	0.997	0.997	0.999	0.960	0.959	0.979
	sig	0.960	0.946	0.970	***0.965***	***0.964***	***0.982***	0.902	0.892	0.946

The best performance of each classifier with specific kernel function is in bold. The best performance across classifiers is underlined.

Table 6
Classifier performance comparison at cross-drive validation level.

		Full_featureset			SBL_featureset			SBL_PCA_featureset		
Feature number		73			56.4 ± 3.1			13.0 ± 0.4		
Classifiers	Kernels	Prec.	Sens.	Spec.	Prec.	Sens.	Spec.	Prec.	Sens.	Spec.
SVM	rbf	0.883	0.866	0.933	0.889	0.873	0.937	***0.897***	***0.885***	***0.942***
	lin	***0.885***	***0.873***	***0.937***	0.881	0.868	0.934	0.877	0.868	0.934
	sig	0.882	0.861	0.931	0.888	0.868	0.934	***0.889***	***0.878***	***0.939***
ELM	rbf	0.888	0.875	0.937	0.889	0.875	0.937	***0.892***	***0.882***	***0.941***
	lin	0.869	0.860	0.930	***0.873***	***0.867***	***0.933***	0.860	0.851	0.926
	sig	0.858	0.849	0.925	0.872	0.850	0.925	***0.884***	***0.864***	***0.932***

The best performance of each classifier with specific kernel function is in bold. The best performance across classifiers is underlined.

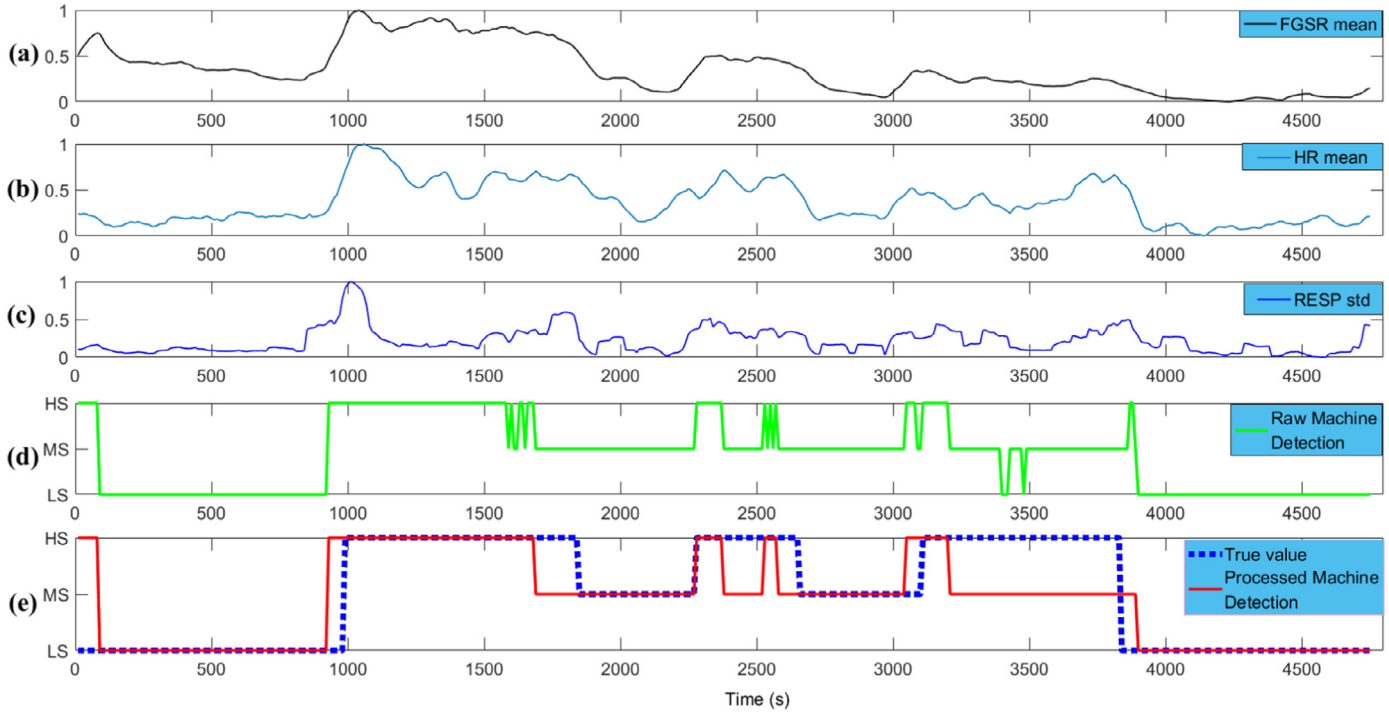

Fig. 8. Stress level detection during a continuous drive accomplished by subject #6. Fig. 8a, b and c shows the change of FGSR, HR and RESP over time. Fig. 8d gives the direct outputs of extremely learning machine with linear kernel. Fig. 8e illustrates the processed outputs of ELM compared with true value marked according to different traffic conditions by the task designers.

These two subjects implemented the same experiments on multiple days. Therefore, for each subject, some typical segments in the previous drives were selected as training data, the data on the last driving test was used for continuous detection. Fig. 8a, b and c shows the change of FGSR, HR and RESP statistical features over time. Fig. 8d gives the direct outputs of extremely learning machine with linear kernel. For a real driving stress warning system, too frequently alarms will disturb the drivers. Therefore, we merged the output pulse with the duration less than 100 s. Previous literature suggests that a driver state evaluation system with a 1–3 min time duration is fast enough to start customized changes to vehicle inside circumstances helping alleviate driver's negative moods (Healey & Picard, 2005). The final automatic detection results show high coincidence with the true road situation especially at the switching interval of traffic conditions. Our results matched the stress response in real driving environments.

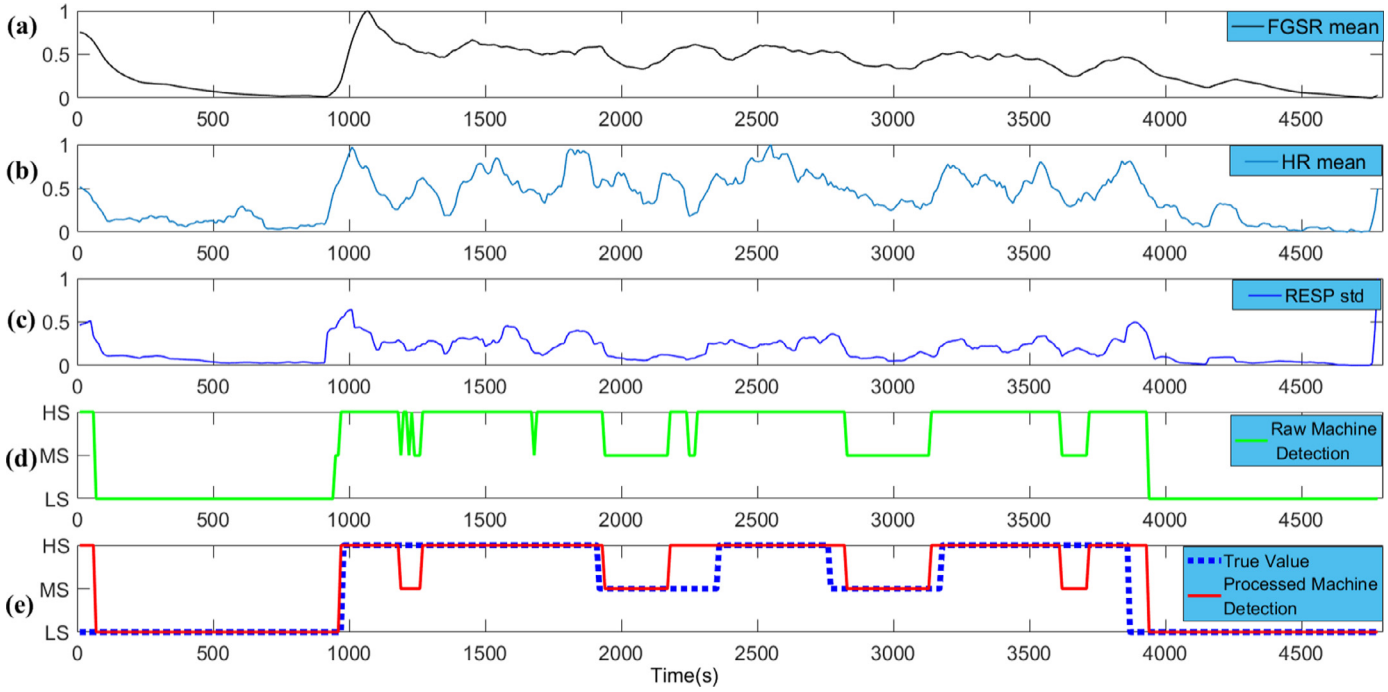

Fig. 9. Stress level detection during a continuous drive accomplished by subject #11. The description for each subgraph is similar to that of Fig. 8.

When a driver has just entered another new traffic condition, he/she is prone to produce relatively obvious physiological changes which can also be observed from the bio-signals directly. For instance, subject #6 finished the rest period and exited the garage for city driving around 1000 s at which FGSR_mean, HR_mean and RESP_std increased dramatically. And this driver accomplished the route in downtown area, started to drive away from the city and onto a highway around 1800 s at which FGSR_mean, HR_mean and RESP_std decreased obviously.

5. Discussion

5.1. Comparisons with similar work

This research, presented an analysis of multimodal features of physiological signals, to identify drive-related stress symptom. Then, it was proposed an automatic detection system to distinguish three different stress levels and provide detailed stress assessment during continuous drive tasks. Three important conclusions could be drawn from this study: 1) different levels of driving stress could be characterized by specific set of physiological measures. 2) multimodal features are favorable for assessing the drive status more reliably and accurately; 3) efficient feature selection methods and kernel-based classifiers shows promising applicability in the real-time monitoring and alerting systems in vehicles, which can help us to decrease the rate of accidents caused by drivers with negative status.

The literature reviews the similar work using ***contact measures***, which are extracted from ECG, GSR, RESP, EMG and EEG; and ***contactless features***, which are drawn from eye movements, facial expressions, and head movements. Table 7 lists some expert systems using these primary measures for mental states assessment during driving and daily working.

First, we compare the performance of several intelligent systems based on ***contact physiological sensors***. Heart rate variability (HRV) is a common non-invasive indicator to detect autonomic nervous system (ANS) activities and is applied as a primary measure for stress detection in many monitoring systems (Vicente et al., 2016; Chiang, 2015; Wang et al., 2013). Vicente et al. (2016) developed an online drowsiness detector based on HRV which can identify the instant status of the subject, ringing an alarm once drowsy or fatigue states were detected. Their system detected the driving drowsiness and sleep deprivation status with a predictive value of 96% and 80% respectively. Chiang (2015) created an ECG-based mental stress assessment system using a rule-based reasoning model by combining fuzzy logic and associative Petri net, which obtained a precision of 95.1%. Wang et al. (2013) presents a KNN classifiers with HRV-based features for driving stress detection. The common point of these studies is that their proposed systems only use ECG sensor to identify driver's states. Single sensor is convenient for wearing and application but sometimes may lead to inaccurate results because it only focuses on a specific aspect of human organism. The analysis from multi-sensors are favorable for assessing the drive status more reliably and accurately. For example, Healey and Picard (2005) developed a driving stress detection system by extracting features in EMG, EKG, GSR and respiration signals and classifying three general stress levels with linear discriminant analysis (LDA). Their system achieved over 97% identification rate and the physiological indicators revealed a high correlation with video-based stress indicators. Singh et al. (2013) proposed an algorithm based on artificial neural network (ANN) to learn driving overall stress level and correlated it with the changes of statistical, structural and time-frequency features in galvanic skin response (GSR) and photoplethysmography (PPG) signals. Their system achieved a precision of 89.23% based on the data collected from over 19 subjects. Lee et al. (2014) constructed a mobile healthcare system for automatic driving sleep-onset detection using support vector machine (SVM) by selecting the most descriptive features obtained from respiration and EEG. Their test results revealed that the integrated application of the respiration and EEG resulted in 98.6% identification accuracy.

Moreover, Table 7 introduces some video-based systems using ***contactless sensors*** (Kholerdi et al., 2016; Azim et al., 2014; Cyganek & Gruszczyński, 2014; Jo et al., 2014). Jo et al. (2014) developed a driver drowsiness detection system using eye movements and support vector machine (SVM) with feature-level and score-level fusion. Their system achieved robust classification results un-

Table 7
Comparison of the existing fatigue/drowsiness/stress detection system.

Feature Category	Year	Authors	Objective	Signals Used	Subjects	Scenario	Expert System Employed (Algorithm)	Performance (Accuracy)
	2016	Vicente et al.	Detecting driving drowsiness	ECG	30	Simulated driving & Real-time driving	Linear Discriminant Analysis (LDA)	96.0%
Contact measures	2015	Chiang et al.	Detecting driving stress	ECG	9	Real-time driving	Fuzzy theory	95.1%
	2013	Wang et al.	Detecting driving drowsiness	ECG	9	Real-time driving	KBCS+LDA +PCA+KNN	100.0%
	2012	Okada et al.	Detecting stress-level	ECG	2	Daily activities	N/A	98.3%
	2010	Sun et al.	Detecting stress-level	ECG,GSR	20	Sitting ,Standing, Walking	Decision Tree, Bayesian Network, Support Vector Machine	Per-subject:92.4%, Cross-Subject:80.9%
	2013	Singh et al.	Detecting driving stress	PPG, GSR	19	Real-time driving	Neural Network(NN)	89.2%
	2012	Villarejo et al.	Detecting stress state	GSR	16	Daily situations	BayesNet J48 SMO	76.6%
	2005	Healey & Picard	Detecting driving stress	ECG, EMG, GSR, RESP	9	Real-time driving	Linear Discriminant Analysis (LDA)	97.4%
	2015	Khan & Hong	Detecting driving drowsiness	EEG	13	Simulated driving	Linear Discriminant Analysis (LDA)	84.9%
	2015	Chen et al.	Detecting drowsiness	EEG and EOG	16	Mental work	Extreme Learning Machine (ELM)	97.3%
	2014	Lee et al.	Detecting driving sleep-onset	EEG, RESP	20	Real-time driving	Support Vector Machine (SVM)	98.6%
	2014	Correa et al.	Detecting drowsiness	EEG	16	Polysomnography	Artificial Neural Network (ANN)	87.4%
	2011	Zhao et al.	Detecting driving fatigue	EEG	10	Simulated driving	KPCA+SVM	81.6%
Contactless measures	2016	Kholerdi et al.	Detecting driving drowsiness	Image based eye, mouth and head state	3	Simulated driving	Multi-module Decision Making Algorithm	90.0%
	2014	Azim et al.	Detecting driving fatigue	Eye closure, Yawning	N/A	Simulated driving	Fuzzy Logic (FL)	100.0%
	2014	Cyganek & Gruszczyński	Monitoring driving fatigue, sleepiness and inattention	Eyemovement in the visible and NIR spectrum	4	Real-time driving	The Higher Order Singular Value Decomposition (HOSVD) Classifier	97.0%
	2014	Jo et al.	Detecting driving drowsiness	Eye state	12	Real-time driving	Support Vector Machine (SVM)	99.0%
Fusion method	2010	Yang et al.	Detecting driving fatigue	Contextual information and Eye movement, ECG, EEG	30	Simulated driving	Dynamic Bayesian Network (DBN)	N/A

der various conditions. Azim et al. (2014) applied image processing technique with fuzzy logic inference (FLI) to determine the level of driving fatigue and realized an accuracy of 100%. However, some defects of visual indicators for example limited illumination, wearing dark glasses and camera-induced distraction of the drivers should be taken into consideration. A hybrid system based on image processing technique for the detection of driving distraction, fatigue and sleepiness was proposed by Cyganek and Gruszczyński (2014). Two cameras in the visual and near infra-red (NIR) spectra make it possible to work both in day and night. Their system obtained a recognition accuracy of 97% except for the condition of wearing dark sunglass. However, long-time exposure to NIR may lead to extra visual fatigue of drivers.

Since the driver's status is not directly observable, but can only be deduced from the available messages, ***the integration of contact and contactless measures*** can be employed to make a more reliable inference (Yang et al., 2010; Chen et al., 2005). Yang et al. (2010) constructed a driving fatigue recognition model based on information fusion, which integrated multiple features extracted from contextual information, eye movements, ECG and EEG. Their proposed system considered the widest coverage of input information and obtained improved prediction compared with similar systems.

In addition to the focus on the input information, the computational techniques are very crucial for the construction of automatic systems in real driving environments. The systems described above applied different inference algorithms such as artificial neural networks, support vector machine, fuzzy logic inference and Bayesian networks. Usually, ANN and SVM spend long time on optimizing model structure and parameters and SVM is prone to suboptimality because of some constraints in the optimization process (Huang et al., 2006). Both Fuzzy logic inference and Bayesian networks need priori knowledge in some case it is difficult to obtain. In our previous research, we used kernel-based extreme learning machine (ELM) for the automatic identification of drowsy/alert status and got an accuracy of 97.3% (Chen et al., 2015). Considering the strong identification ability and very fast learning speed of ELM algorithm, it is also applied into the present study.

5.2. Merits of the proposed system

Assuming that different levels of driving stress could be characterized by specific set of physiological measures, our research purpose is to build an automatic detection system based on physiological data to indentify different drive stress levels and further make detailed inference of driver's stress in a dynamical and continuous task. Our experimental results demonstrate the feasibility and ef-

fectiveness of the proposed solution. The main contributions of the present study could be summarized as follows:

(1) Under real driving condition.
Unlike the drive executed in driving simulators where the stress statuses can be reproduced and controlled, real driving circumstances contain large uncontrollability and unpredictability. It is a challenging and rewarding task for identification of driver's status in real driving environments. This experiment is designed to monitor driver's emotion and mental status during real-world driving situations, which enables the results more practically applicable to use in daily commuting conditions.
However, the experiments performed in real traffic circumstances may encounter many constraints such as data acquisition and subject-dependent issues. Taking consideration of these constraints, we implemented two types of analysis not only enabling the system to correctly distinguish different stress levels from short-interval segments but also give detailed stress assessment during continuous drive tasks. Generally, our automatic detection results could match the stress response in real driving, which show high coincidence with the true road situation especially in the conversion interval of traffic conditions.
Moreover, Healey and Picard (2005) analyzed 5 min intervals of data, which is much longer than the duration used in our work (100 s length segments advanced by 10 s). It is critical to detect the driver states as soon as possible to initiate in-vehicle equipments to release warning signals for avoiding possible accidents.
(2) Widely extract physiological characteristics from multi-domains.
The current study provides drivers with a relative reliable assessment of their stress level by fusing physiological features from multiple aspects: waveform information from time series, components from power spectrum and statistics from wavelet coefficients. All the features we analyzed are easy to compute and can be further developed into the driver assistance system. Especially, wavelet decomposition provides more accurate information than that in original time series or full-spectrum analysis. In some sense wavelet transformation assists in revealing rewarding information about constituent electrophysiology activities embedded in the original signals (Chen et al., 2015; Li & Chung, 2013). Further, the integration of various input information is helpful to enhance the identification capacity.
(3) Achieve high identification performance with low computational cost.
We propose an efficient feature selection method by fusing sparse Bayesian learning (SBL) and Principal Component Analysis (PCA), which helps to find the optimal feature sets and reduce the computational cost of the subsequent detectors. Support vector machine (SVM) and extreme learning machine (ELM) were applied for developing a highly sensitive detector with fast learning speed. Among the proposed classifiers, ELM can accomplish learning steps very fast and achieve less-than-perfect but comparable accuracy, which makes it suitable in the development of intelligent monitoring system.

5.3. Limitations and future work

The limitations and future research directions could be summarized as follows:

First, current recording technique should be improved to a certain level which can be applied in daily activities to take full advantage of physiological measures. For the wired and stationary recording systems, the bigger issue may be the invasiveness of attaching and wearing these sensors. In future, to provide a more comfortable and convenient lifestyle solution for the users, it's better to transform the existing acquisition system into wearable and wireless mode with high signal quality and reliable data transmission.

Second, this research only analyzes the physiological measures for the identification of drive-related stress. As multiple features are favorable for detecting driver's status more reliably and robustly, a hybrid detection system could be developed considering vehicle behavior, physiological measures and image-based indicators. In addition, electroencephalogram (EEG) measures are not utilized in the present study. In fact, EEG is commonly regarded as the "golden standard" in the recognition of human emotion and vigilance level (Hosseini et al., 2015; Wang et al., 2015; Correa et al., 2014; Li et al., 2012; Liu et al., 2010), it can also be used as a promising approach in our future work.

Third, a problem associated with modeling is that the optimal feature sets are not constant for the different subjects. More generalized models have to be explored for the adaptability of the previously trained models on new participant. The practical applicability of drive status evaluation approaches could be enhanced based on the improvements of feature extraction, selection and inference algorithms.

Finally, viewed from a long-term aspect, more attention should be paid on the online warning and assistance countermeasures including vehicle inside and outside environmental aspects. Our design scheme for drive stress detection can facilitate the development of some in-vehicle intelligent systems using physiological measures, for example, drivers' emotion management, drivers' sleeping onset monitoring, and human-computer interfaces (HCI). HCI technique can interpret drivers' mental and emotion status intelligently and trigger the corresponding electronic terminals in various approaches to help the drivers better manage their negative driving status. For instance, during highly stressful road conditions, in-vehicle music playback system could automatically turn down the volume and switch to a more relaxing music chapter which is helpful to relieve tension. Other noncritical systems such as radios and voice navigation could also adapt working mode to create a better environment inside automobiles. Furthermore, based on the continuous evaluation of drivers' status during daily commuting route, city planners could optimize road layout management especially at highly stressful and fatigue spots.

6. Conclusions

Monitoring drivers' internal status has great potential in declining the occurrence probability of traffic accidents. The present research developed an automatic system for detecting the driving related stress level based on multichannel physiological records. A variety of features were extracted using wavelet decomposition, time and spectral analysis. Sparse Bayesian Learning (SBL) and Principal Component Analysis (PCA) were combined and adopted to search for the optimal and compact feature sets. Kernel-based classifiers were employed to improve the accuracy of stress detection task. Two kinds of analysis, i.e., classification of short-interval segment and continuous evaluation for complete drive were made. Our results reveal that physiological measures especially skin conductance and heart rate indicators can provide efficient assessment of driving stress. The proposed system combing efficient feature selection methods and kernel-based classifiers can be developed into a real-time monitoring and alerting systems in vehicles helping to decrease the rate of accidents caused by drivers under a variety of improper statuses.

Acknowledgement

This work is partly supported by National Natural Science Foundation of China (Nos. 61201124, 51407078) and Fundamental Research Funds for the Central Universities WH1414022.

References

Åkerstedt, T., Hallvig, D., Anund, A., Fors, C., Schwarz, J., & Kecklund, G. (2013). Having to stop driving at night because of dangerous sleepiness–awareness, physiology and behaviour. *Journal of Sleep Research, 22*(4), 380–388.

Atkinson, J., & Campos, D. (2016). Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. *Expert Systems with Applications, 47*, 35–41.

Azim, T., Jaffar, M. A., & Mirza, A. M. (2014). Fully automated real time fatigue detection of drivers through fuzzy expert systems. *Applied Soft Computing, 18*, 25–38.

Bakker, J., Pechenizkiy, M., & Sidorova, N. (2011). What's your current stress level? Detection of stress patterns from GSR sensor data. In *2011 IEEE 11th international conference on data mining workshops* (pp. 573–580). IEEE.

Chen, H., & Meer, P. (2005). Robust fusion of uncertain information. *Systems, Man, and Cybernetics, Part B, 35*(3), 578–586.

Chen, L. L., Zhao, Y., Zhang, J., & Zou, J. Z. (2015). Automatic detection of alertness/drowsiness from physiological signals using wavelet-based nonlinear features and machine learning. *Expert Systems with Applications, 42*(21), 7344–7355.

Chiang, H. S. (2015). Ecg-based mental stress assessment using fuzzy computing and associative petri net. *Journal of Medical and Biological Engineering, 35*(6), 833–844.

Correa, A. G., Orosco, L., & Laciar, E. (2014). Automatic detection of drowsiness in EEG records based on multimodal analysis. *Medical Engineering & Physics, 36*(2), 244–249.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

Craye, C., Rashwan, A., Kamel, M. S., & Karray, F. (2015). A multi-modal driver fatigue and distraction assessment system. *International Journal of Intelligent Transportation Systems Research*, 1–22.

Cyganek, B., & Gruszczyński, S. (2014). Hybrid computer vision system for drivers' eye recognition and fatigue monitoring. *Neurocomputing, 126*, 78–94.

Daanen, H. A., van de Vliert, E., & Huang, X. (2003). Driving performance in cold, warm, and thermoneutral environments. *Applied Ergonomics, 34*(6), 597–602.

Eckstein, M. K., Guerra-Carrillo, B., Singley, A. T. M., & Bunge, S. A. (2016). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience*..

Fu, R., & Wang, H. (2014). Detection of driving fatigue by using noncontact EMG and ECG signals measurement system. *International Journal of Neural Systems, 24*(03), 1450006.

Garcia, I., Bronte, S., Bergasa, L. M., Almazán, J., & Yebes, J. (2012). Vision-based drowsiness detector for real driving conditions. In *Intelligent vehicles symposium (IV), 2012 IEEE* (pp. 618–623). IEEE.

González-Ortega, D., Díaz-Pernas, F. J., Antón-Rodríguez, M., Martínez-Zarzuela, M., & Díez-Higuera, J. F. (2013). Real-time vision-based eye state detection for driver alertness monitoring. *Pattern Analysis and Applications, 16*(3), 285–306.

Hallvig, D., Anund, A., Fors, C., Kecklund, G., Karlsson, J. G., Wahde, M., et al. (2013). Sleepy driving on the real road and in the simulator-A comparison. *Accident Analysis & Prevention, 50*, 44–50.

Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *Intelligent Transportation Systems, IEEE Transactions on, 6*(2), 156–166.

Hoffmann, U., Yazdani, A., Vesin, J. M., & Ebrahimi, T. (2008). Bayesian feature selection applied in a P300 brain-computer interface. In *Signal processing conference, 2008 16th European* (pp. 1–5). IEEE.

Horberry, T., Anderson, J., Regan, M. A., Triggs, T. J., & Brown, J. (2006). Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident Analysis & Prevention, 38*(1), 185–191.

Hosseini, S. A., Khalilzadeh, M. A., Naghibi-Sistani, M. B., & Homam, S. M. (2015). Emotional stress recognition using a new fusion link between electroencephalogram and peripheral signals. *Iranian Journal of Neurology, 14*(3), 142.

Hostens, I., & Ramon, H. (2005). Assessment of muscle fatigue in low level monotonous task performance during car driving. *Journal of Electromyography and Kinesiology, 15*(3), 266–274.

Huang, G. B. (2014). An insight into extreme learning machines: Random neurons, random features and kernels. *Cognitive Computation, 6*, 376–390.

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing, 70*(1), 489–501.

Jo, J., Lee, S. J., Park, K. R., Kim, I. J., & Kim, J. (2014). Detecting driver drowsiness using feature-level fusion and user-specific classification. *Expert Systems with Applications, 41*(4), 1139–1152.

Kholerdi, H. A., TaheriNejad, N., Ghaderi, R., & Baleghi, Y. (2016). Driver's drowsiness detection using an enhanced image processing technique inspired by the human visual system. *Connection Science, 28*(1), 27–46.

Lee, B. G., Lee, B. L., & Chung, W. Y. (2014). Mobile healthcare for automatic driving sleep-onset detection using wavelet-based EEG and respiration signals. *Sensors, 14*(10), 17915–17936.

Lee, B. G., & Chung, W. Y. (2012). Driver alertness monitoring using fusion of facial features and bio-signals. *IEEE Sensors Journal, 12*(7), 2416–2422.

Li, G., & Chung, W. Y. (2013). Detection of driver drowsiness using wavelet analysis of heart rate variability and a support vector machine classifier. *Sensors, 13*(12), 16494–16511.

Li, W., He, Q. C., Fan, X. M., & Fei, Z. M. (2012). Evaluation of driver fatigue on two channels of EEG data. *Neuroscience Letters, 506*(2), 235–239.

Liao, W., Zhang, W., Zhu, Z., & Ji, Q. (2005). A real-time human stress monitoring system using dynamic Bayesian network. *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)-workshops*. IEEE 70-70.

Liu, J., Zhang, C., & Zheng, C. (2010). EEG-based estimation of mental fatigue by using KPCA–HMM and complexity parameters. *Biomedical Signal Processing and Control, 5*(2), 124–130.

Liu, Y. T., Lin, Y. Y., Wu, S. L., Chuang, C. H., & Lin, C. T. (2015). Brain dynamics in predicting driving fatigue using a recurrent self-evolving fuzzy neural network. *IEEE Transactions on Neural Networks and Learning Systems, 27*(2), 347–360.

Morris, D. M., Pilcher, J. J., & Switzer, F. S., III (2015). Lane heading difference: An innovative model for drowsy driving detection using retrospective analysis around curves. *Accident Analysis & Prevention, 80*, 117–124.

Okada, Y., Yoto, T. Y., Suzuki, T. A., Sakuragawa, S., Sakakibara, H., Shimoi, K., et al. (2013). Wearable ECG recorder with acceleration sensors for monitoring daily stress. *Journal of Medical and Biological Engineering, 33*(4), 420–426.

Pimenta, A. M., & Assunção, A. Á. (2015). Thermal discomfort and hypertension in bus drivers and chargers in the metropolitan region of Belo Horizonte, Brazil. *Applied Ergonomics, 47*, 236–241.

Sharma, N., & Gedeon, T. (2012). Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Computer Methods and Programs in Biomedicine, 108*(3), 1287–1301.

Singh, R. R., Conjeti, S., & Banerjee, R. (2013). A comparative evaluation of neural network classifiers for stress level analysis of automotive drivers using physiological signals. *Biomedical Signal Processing and Control, 8*(6), 740–754.

Smith, J., Mansfield, N., Gyi, D., Pagett, M., & Bateman, B. (2015). Driving performance and driver discomfort in an elevated and standard driving position during a driving simulation. *Applied Ergonomics, 49*, 25–33.

Sun, F. T., Kuo, C., Cheng, H. T., Buthpitiya, S., Collins, P., & Griss, M. (2010). Activity-aware mental stress detection using physiological sensors. In *International conference on mobile computing, applications, and services* (pp. 211–230). Berlin Heidelberg: Springer.

Tamrin, M., Bahri, S., Saliman, K., Ng, Y. G., Yusoff, M., & Syah, I. (2014). Stress; the vulnerability and association with driving performance. *American Journal of Applied Sciences, 11*(3), 448–454.

Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research, 1*(Jun), 211–244.

Vicente, J., Laguna, P., Bartra, A., & Bailón, R. (2016). Drowsiness detection using heart rate variability. *Medical & Biological Engineering & Computing*, 1–11.

Villarejo, M. V., Zapirain, B. G., & Zorrilla, A. M. (2012). A stress sensor based on galvanic skin response (GSR) controlled by ZigBee. *Sensors, 12*(5), 6075–6101.

Wang, H., Zhang, C., Shi, T., Wang, F., & Ma, S. (2015). Real-time EEG-based detection of fatigue driving danger for accident prediction. *International Journal of Neural Systems, 25*(02), 1550002.

Wang, J. S., Lin, C. W., & Yang, Y. T. C. (2013). A k-nearest-neighbor classifier with heart rate variability feature-based transformation algorithm for driving stress recognition. *Neurocomputing, 116*, 136–143.

Wang, X., & Xu, C. (2015). Driver drowsiness detection based on non-intrusive metrics considering individual specifics. *Accident analysis & prevention*. In Press Corrected Proof.

Yang, G., Lin, Y., & Bhattacharya, P. (2010). A driver fatigue recognition model based on information fusion and dynamic Bayesian network. *Information Sciences, 180*(10), 1942–1954.

Zhang, W., Cheng, B., & Lin, Y. (2012). Driver drowsiness recognition based on computer vision technology. *Tsinghua Science and Technology, 17*(3), 354–362.

Zhang, Y., Zhou, G., Jin, J., Zhao, Q., Wang, X., & Cichocki, A. (2015). Sparse Bayesian classification of EEG for brain-computer interface.

Zhao, C., Zhao, M., Liu, J., & Zheng, C. (2012). Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator. *Accident Analysis & Prevention, 45*, 83–90.

Zhao, C., Zheng, C., Zhao, M., Tu, Y., & Liu, J. (2011). Multivariate autoregressive models and kernel learning algorithms for classifying driving mental fatigue based on electroencephalographic. *Expert Systems with Applications, 38*(3), 1859–1865.

Zhao, S., Xu, G., & Tao, T. (2009). Detecting of driver's drowsiness using multiwavelet packet energy spectrum. In *Image and signal processing, 2009. CISP'09. 2nd international congress on* (pp. 1–5). IEEE.

Zheng, R., Yamabe, S., Nakano, K., & Suda, Y. (2015). Biosignal analysis to assess mental stress in automatic driving of trucks: Palmar perspiration and masseter electromyography. *Sensors, 15*(3), 5136–5150.