Driver Drowsiness Classification Using Fuzzy Wavelet-Packet-Based Feature-Extraction Algorithm

Rami N. Khushaba*, Sarath Kodagoda, Sara Lal, and Gamini Dissanayake

***Abstract*—Driver drowsiness and loss of vigilance are a major cause of road accidents. Monitoring physiological signals while driving provides the possibility of detecting and warning of drowsiness and fatigue. The aim of this paper is to maximize the amount of drowsiness-related information extracted from a set of electroencephalogram (EEG), electrooculogram (EOG), and electrocardiogram (ECG) signals during a simulation driving test. Specifically, we develop an efficient fuzzy mutual-information (MI)- based wavelet packet transform (FMIWPT) feature-extraction method for classifying the driver drowsiness state into one of predefined drowsiness levels. The proposed method estimates the required MI using a novel approach based on fuzzy memberships providing an accurate-information content-estimation measure. The quality of the extracted features was assessed on datasets collected from 31 drivers on a simulation test. The experimental results proved the significance of FMIWPT in extracting features that highly correlate with the different drowsiness levels achieving a classification accuracy of 95%–97% on an average across all subjects.**

***Index Terms*—Biosignal processing, driver drowsiness, feature extraction.**

I. Introduction

DROWSINESS is an intermediate state between wakefulness and sleep that has been defined as a state of progressive impaired awareness associated with a desire or inclination to sleep [1]. In certain tasks, such as driving, drowsiness is considered as a significant risk factor that substantially contributes to the increasing number of motor vehicle accidents each year [2]. Critical aspects of driving impairments associated with drowsiness are slow reaction times, reduced vigilance, and deficits in information processing that all lead to an abnormal driving behavior [3], [4]. Driver drowsiness is usually used interchangeably with the term driver fatigue; however, each of these terms has its own meaning. Fatigue is considered as one of the factors that can lead to drowsiness and is a consequence of physical labor or a prolonged experience, and is defined as a disinclination to continue the task at hand [5]. Driver fatigue is believed to account for 35%–45% of all vehicle accidents [6]. Some authors distinguish fatigue from drowsiness as the former does not fluctuate rapidly, over periods of a few seconds, as drowsiness. Usually, rest and inactivity relieves fatigue, however, it makes drowsiness worse [7].

Different studies have been reported on driver drowsiness detection including methods identifying physiological associations between driver drowsiness/fatigue and the corresponding patterns of the electroencephalogram (EEG) (brain activity), electrooculogram (EOG) (eye movement), and electrocardiogram (ECG) (heart rate) signals [8]–[12]. Most of these studies reported that the physiological approach to drowsiness detection can provide very accurate results as strong correlation between these signals and the driver's cognitive state was found in many studies [9], [13]. Specifically, many of these studies suggested that the change in the cognitive state can be associated with significant changes in the EEG frequency bands, such as delta (δ: 0–4 Hz), theta (θ: 4–8 Hz), alpha (α: 8–13 Hz), and beta (β: 13–20 Hz) [14], [15], or their combinations [16], [17], and with changes in the eyelid parameters extracted from the EOG [18]. Additionally, heart rate variability was found to be applicable for the detection of drowsiness and fatigue using the ECG power spectrum [19].

The crucial step in most of the aforementioned reported studies includes the extraction of a set of features that correlate with drowsiness. As an example, Fu *et al.* [20] utilized probabilistic principal components analysis to extract features from 62 EEG channels to distinguish awake, drowsy, and sleep in a driving simulation study. A significant drop in classification accuracy from 97% to 89% was reported when classifying the extracted features into three classes rather than two binary classes (either awake and sleep) due to the fluctuations of the drowsy state. It was also reported that features extracted from high-frequency bands are more stable than features extracted from the low-frequency bands [20]. A smaller number of channels was reported in other studies, for example, Lin *et al.* [21] utilized 32 EEG/EOG channels and 2 ECG channels, and employed power spectrum and independent component analysis with fuzzy neural networks for alertness estimation. A maximum testing accuracy of 91.3% on an average across several subjects was reported. On the other hand, Yeo *et al.* [22] employed support vector machines for classifying different frequency-spectrum-based

Manuscript received June 1, 2010; revised August 6, 2010; accepted September 4, 2010. Date of publication September 20, 2010; date of current version December 17, 2010. This work is supported in part by the ARC Centre of Excellence programme, funded by the Australian Research Council and the New South Wales State Government, and in part by and the Centre for Intelligent and Mechatronics Systems, Sydney, Australia. *Asterisk indicates corresponding author.*

*R. N. Khushaba is with the ARC Centre of Excellence for Autonomous Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney, Broadway NSW 2007, Australia (e-mail: rkhushab@eng.uts.edu.au).

S. Kodagoda and G. Dissanayake are with the ARC Centre of Excellence for Autonomous Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney, Broadway NSW 2007, Australia (e-mail: sakoda@eng.uts.edu.au; gdissa@eng.uts.edu.au).

S. Lal is with the Department of Medical and Molecular Biosciences, Faculty of Science, University of Technology, Sydney, Broadway NSW 2007, Australia (e-mail: sara.lal@uts.edu.au).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2010.2077291

features reporting very high classification results of 99% accuracy with 19 EEG channels and an EOG recording. The authors also reported that the achieved classification accuracy would be compromised, if less EEG channels were used [22].

Despite the aforementioned results, there are a number of limitations associated with many of these previous studies. These include, most previous studies have attempted to use the same estimators for all subjects. However, the relatively large individual variability in EEG dynamics accompanying loss of alertness means that, for many drivers, group statistics cannot be used to accurately predict changes in alertness and performance as denoted by Jung *et al.* [23]. Another limitation in studies relating drowsiness/fatigue to EEG/EOG/ECG spectrum is the use of a small number of spectral bands defined *a priori*, rather than exploring the benefits of using the full spectrum [14]–[17]. It is also noticed that conventional techniques like the fast Fourier transform (FFT) were mostly employed for detecting changes in the spectral components of the corresponding signals [17], [23]. However, it is generally known that the FFT is not very suitable to extract features localized simultaneously in the time and frequency domains. The features extracted by FFT are of global nature either in time or frequency domains so that the interpretation of the results may not be straightforward. Additionally, the FFT is more suitable for analyzing stationary signals, while physiological signals like the EEG, tend to be nonstationary ones. This might in turn limit the accuracy of the drowsiness-detection system that is based on the aforementioned estimators and methods.

On the other hand, it is generally known that methods like the wavelet transform and the wavelet packet transform (WPT) are more suitable than FFT when dealing with nonstationary signals. Thus, an attempt for drowsiness detection with WPT-based feature-extraction method was utilized by Zhang *et al.* [24] on 13 EEG channels, 1 EOG, and 1 ECG channels. However, Shannon entropy was applied on the WPT to select the representative features, but it is generally known that Shannon entropy is not suitable for classification problems but for compression problems [25]. Thus, kernel-feature-projection techniques were necessary to provide an average accuracy of 91% as the original features were not powerful representatives of the underlying problem.

The aim of this paper is to develop an automated method of distinguishing different levels of drowsiness based on the EEG/EOG/ECG signals collected while performing simulated driving. Specifically, we propose to employ the WPT to construct features that highly correlate with alertness and the different levels of drowsiness. The WPT is chosen due to its ability to deal with stationary, nonstationary, or transitory characteristics of different signals including abrupt changes, spikes, drifts, and trends [26]. In order to automatically select the frequency components that are most suitable for constructing the drowsiness/alert features, a novel mutual information (MI) estimation measure is developed based on a generalization of the concept of fuzzy entropy. The justification for the need of such a measure is that it reduces the computational cost associated with estimating the MI (when comparing with other methods, such as Parzen density estimators and kernel density estimators for MI calculation) while providing an accurate estimation of the information contents of the different features. Thus, different features are extracted for each subject using the proposed method that automatically optimizes the full frequency spectrum of the aforementioned signals for the best representation for detecting drowsiness. Additionally, unlike most attempts in the literature necessitating a large number of channels, only 3 EEG, 1 EOG, and 1 ECG channels are utilized in this paper to provide a more practical approach.

The structure of the subsequent sections of the paper is as follows: Section II review the available WPT-based feature-extraction methods and justifies the need for the new method. Section III presents the proposed fuzzy-entropy-based MI measure. The data-collection procedure is described in Section IV. Section V describes first the data-collection procedure and then presents the experimental results. Finally, a conclusion is provided in Section VI.

II. Background

A. Wavelet-Packet Transform

Biomedical signals usually consist of brief high-frequency components closely spaced in time, accompanied by long lasting, low-frequency components closely spaced in frequency. Wavelets are considered appropriate for analyzing such signals as they exhibit good frequency resolution along with finite-time resolution; the first to localize the low-frequency components and the second to resolve the high-frequency components [27].The wavelet-packets transform, referred to as WPT subsequently, was introduced by Coifman *et al.* [28] by generalizing the link between multiresolution approximations and wavelets. The WPT may be thought of as a tree of subspaces, with $\Omega_{0,0}$ representing the original signal space, i.e., the root node of the tree. In a general notation, the node $\Omega_{j,k}$, with j denoting the scale and k denoting the subband index within the scale, is decomposed into two orthogonal subspaces: an approximation space $\Omega_{j,k} \rightarrow \Omega_{j+1,2k}$ plus a detail space $\Omega_{j,k} \rightarrow \Omega_{j+1,2k+1}$ [29]. This is done by dividing the orthogonal basis $\left\{\phi_j(t-2^j k)\right\}_{k\in Z}$ of $\Omega_{j,k}$ into two new orthogonal bases $\left\{\phi_{j+1}(t-2^{j+1}k)\right\}_{k\in Z}$ of $\Omega_{j+1,2k}$ and $\left\{\psi_{j+1}(t-2^{j+1}k)\right\}_{k\in Z}$ of $\Omega_{j+1,2k+1}$ [30], where $\phi_{j,k}(t)$ and $\psi_{j,k}(t)$ are the scaling and wavelet functions, respectively, that are given in [30] as

$$\phi_{j,k}(t) = \frac{1}{\sqrt{|2^j|}}\phi\left(\frac{t-2^j k}{2^j}\right) \tag{1}$$

$$\psi_{j,k}(t) = \frac{1}{\sqrt{|2^j|}}\psi\left(\frac{t-2^j k}{2^j}\right) \tag{2}$$

where the dilation factor 2^j, also known as the scaling parameter, measures the degree of compression or scaling. On the other hand, the location parameter $2^j k$, also known as translation parameter, determines the time location of the wavelet.

The difference to the wavelet transform is that, for the subsequent decomposition levels, the WPT not only decomposes the approximation coefficients, but also the detail coefficients.

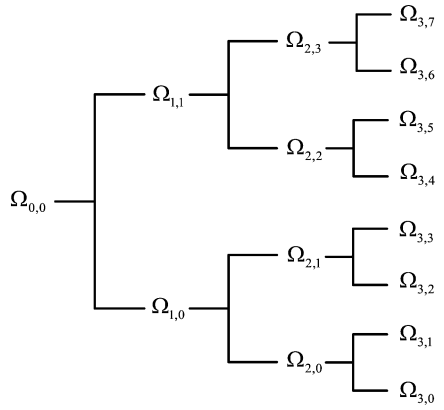

Fig. 1. Example of the wavelet-packet decompositions of $\Omega_{0,0}$ into tree-structured subspaces.

This process is repeated J times, where $J \leq \log_2 N$, with N being the number of samples in the original signal. This in turn results in $J \times N$ coefficients. Thus, at resolution level j, where $j = 1, 2, \ldots, J$, the tree has N coefficients divided into 2^j coefficient blocks or crystals. This iterative process generates a binary wavelet-packet tree-like structure, where the nodes of the tree represent subspaces with different frequency localization characteristics. This is shown schematically in Fig. 1 with three levels of decomposition [29].

B. Wavelet-Packet-Based Feature-Extraction Methods

The problem of WPT-based feature extraction can be decomposed into two main tasks: feature construction and bases selection. In the feature-construction step, the goal is to utilize the WPT coefficients generated at each of the WPT tree subspaces (shown in Fig. 1) to construct variables or properties that can best represent the classes of the signals at hand. On the other hand, the bases-selection problem is related to the identification of the best bases (given an ensemble of bases) from which the constructed features can highly discriminate between the signals belonging to different problem classes. Early contributions to this field included the work presented by Coifman and Wickerhauser [31] in which they proposed the joint best basis (JBB) method utilizing Shannon entropy as a cost function for bases selection. A limitation of JBB method is that it is most suitable for compression tasks rather than classification tasks. The justification is that classification problems need a measure to evaluate the power of discrimination of each subspace in the tree-structured subspaces rather than the efficiency in representation. To overcome such a limitation, the local discriminant basis (LDB) algorithm was proposed by Saito [25] employing the symmetric relative entropy. Both JBB and LDB are concerned with the energy levels of signals and were reported by Deqiang *et al.* [26] to exhibit some drawbacks when it comes to accentuating the discriminatory power essential in classification tasks. A fuzzy-set-based criterion was proposed by Deqiang *et al.* [26] in their fuzzy wavelet-packet method (or simply the FWP) to aid in the selection of the best basis showing better performance than JBB and LDB when classifying biomedical signals. On the other hand, the optimal wavelet packet (OWP) method proposed by Wang *et al.* [32] using Davies–Bouldin criterion managed to outperform the FWP when classifying biomedical signals. The justification for such a performance is that the feature representation in OWP attempts to overcome the lack of the shift-invariance property observed in the feature representation of FWP, JBB, and LDB as stated by Wang *et al.* [32]. However, the distances between the classes or clusters' centers across each feature (or dimension) are used to judge on the features suitability for classification in both OWP and FWP (in the preprocessing step). It is generally known that considering distance measures alone may not be very useful to judge the corresponding features ability in separating different classes [33]. A justification is that classes' means alone cannot be considered to estimate the classification error, since the error also depends on the overlap between the class likelihoods. Thus, a more powerful measure is required to identify the information content of the different features.

C. Mutual Information-Based Information Estimation

One of the most suitable approaches for estimating the arbitrary dependency between random variables is based on the concept of MI. The concept of MI is tightly interconnected with the concept of uncertainty, where the MI between two random variables is viewed as the capacity to reduce the uncertainty about those variables [37], [38]. The entropy, being a measure of uncertainty, is usually used to represent the MI between each feature or variable (denoted as f) and the class label (denoted as C) according to the following:

$$\begin{aligned} I(f;C) &= H(f) + H(C) - H(f,C) \\ &= H(f) - H(f|C) \end{aligned} \tag{3}$$

where $H(f)$ and $H(C)$ are the marginal entropies of f and C, respectively, and $H(f,C)$ and $H(f|C)$ are the joint and conditional entropies of f and C, respectively. According to Fano [34], maximizing the MI between different features and the desired target can achieve the lowest probability of error. However, in order to estimate the probability distribution function associated with a random variable, a certain density-estimation method needs to be employed for which many variants exist in the literature. Examples include the histograms, k-nearest neighbors (*k*NN), Parzen, and kernel density estimators [35]. The simplest and most widely used is the histogram approach in which one simply constructs the bins of the histogram by counting the number of occurrences of the values. Unfortunately, this method requires a huge number of samples to capture the density accurately. However, this will in turn lead to higher computational complexities. As an alternative to the aforementioned approaches in MI computation, we propose to employ fuzzy memberships for the calculation of MI, or the associated entropies required for MI computation (depending on the approach utilized). The justification behind the proposed approach is that fuzzy entropy and MI measures are able to reflect the actual distribution of the classification patterns and this will in turn reduce the generated decision regions for classification [36] and

will also provide more accurate estimation of the information content.

III. Mutual Information-Based Wavelet-Packet Feature Extraction

There are two classical theories of uncertainty within which we can define the entropy, the first and the most well known is based on the notion of *probability*, while the second is based on the notion of *possibility* [38]. In the probabilistic approach, *Shannon entropy* is a well-known measure of uncertainty and is extensively covered in the literature. An extension to Shannon entropy is the concept of *fuzzy entropy*, in which fuzzy sets are used to aid the estimation of the entropies. It should be highlighted that the measuring of the fuzzy entropy is quite different from the classical Shannon entropy since fuzzy entropy contains fuzziness uncertainties (possibilistic), while Shannon entropy contains randomness uncertainties (probabilistic). However, a well-defined fuzzy entropy must satisfy the four Luca–Termini axioms [39] given as follows . The first task was to estimate the required memberships of the samples along each dimension (or feature) in all of the problem classes. Several methods were reported in the literature for estimating the membership functions including the kNN approach [40], the well-known fuzzy c-means method (employed within the FWP) [41], and many other variants [33], [42]. Given the high computation cost associated with the kNN approach and the singularity problem of the fuzzy c-means, we propose to use the following approach for estimating the membership values.

Given a universal set with elements $\mathbf{x}_k$ distributed in a pattern space as $\mathrm{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_l\}$, where $k = 1, 2, \ldots, l$ with l being the total number of patterns. For simplicity, It will be useful to describe the membership value that the kth vector has in the ith class with the following notation:

$$\mu_{ik} = \mu_i(\mathbf{x}_k) \in [0, 1]. \tag{4}$$

Denote the mean of the data sample that belong to class i as $\overline{\mathbf{x}_i}$ and the radius of the data as r

$$r = \max \left\| \overline{\mathbf{x}}_i - \mathbf{x}_k \right\|_\sigma . \tag{5}$$

Then the fuzzy membership μ_{ik} can be calculated as follows

$$\mu_{ik} = \left(\frac{\left\| \overline{\mathbf{x}}_i - \mathbf{x}_k \right\|_\sigma}{r + \epsilon} \right)^{\frac{-2}{m-1}} \tag{6}$$

where m is the fuzzification parameter, and $\epsilon > 0$ is a small value to avoid singularity, and σ is the standard deviation involved in the distance computation. Finally, the membership of each of the samples in all of the problem classes is normalized according to $\sum_{i=1}^{c} \mu_{ik} = 1$.

A. Fuzzy Entropy and Mutual Information

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a discrete random variable with a finite alphabet set containing n symbols, and let $\mu_A(x_i)$ be the membership degree of the element x_i to fuzzy set A, and F be a set-to-point mapping $F : G(2^X) \rightarrow$ [0, 1]. Hence, F is a fuzzy set defined on fuzzy sets. F is an entropy measure if it satisfies the following Luca–Termini axioms [36], [39], [43]:

1) $F(A) = 0$ iff $A \in 2^X$, where A is a nonfuzzy set and 2^X indicates the power set of set A.
2) $F(A) = 1$ iff $\mu_A(x_i) = 0.5$ for all i;
3) $F(A) \leq F(B)$ if A is less fuzzy than B, i.e., if $\mu_A(x_i) \leq \mu_B(x_i)$ when $\mu_B(x_i) \leq 0.5$ and $\mu_A(x_i) \geq \mu_B(x_i)$ when $\mu_B(x_i) \geq 0.5$;
4) $F(A) = F(A^c)$;

where $A^c = (1 - \mu_A(x_1), \ldots, 1 - \mu_A(x_n))$. Shannon entropy satisfies the aforementioned four De Luca–Termini axioms, where for a discrete random variable X with a probability mass function $p(x_i)$, Shannon entropy is defined by

$$H(X) = -\sum_i p(x_i) \log_2 p(x_i). \tag{7}$$

Using the proposed membership function in (6), we construct c-fuzzy sets along each specific feature f, each of these will in turn reflect the membership degrees of the samples in each of the c problem classes. The fuzzy equivalent to the joint probability of the training patterns that belong to class i is given here as

$$P(f, c_i) = \frac{\sum_{k \in A_i} \mu_{ik}}{\mathrm{NP}} \tag{8}$$

where $P(f, c_i)$ can be interpreted as the degree by which the samples that are predefined to belong to class i does actually contribute to that specific class. A_i is the set of indices of the training patterns belonging to class i, and NP is the total number of patterns. The joint fuzzy entropy of the elements of class i, denoted as $H(f, c_i)$, is then equal to

$$H(f, c_i) = -P_{f,c_i} \log P_{f,c_i} . \tag{9}$$

In order to account for the entropy along all c-classes, the aforementioned entropy has to be summed along the universal set to generate the complete fuzzy entropy $H(f, C)$

$$H(f, C) = \sum_{i=1}^{c} H(f, c_i). \tag{10}$$

The aforementioned entropy satisfies the four De Luca–Termini axioms and is termed as the *joint fuzzy entropy*. The aforementioned equations can be applied on the samples along each feature, thus computing the entropies associated with each feature.

In order to find the marginal entropy $H(f)$ of each feature, we add the estimated membership values of the samples along each of the c-fuzzy sets S_i as follows:

$$P(f_{S_i}) = \frac{\sum_k \mu_{ik}}{\mathrm{NP}}. \tag{11}$$

Then the marginal entropy is found by

$$H(f) = -P_{f_{S_i}} \log P_{f_{S_i}} . \tag{12}$$

Similarly, in order to construct the class marginal entropy H_C, we first find the fuzzy equivalent to the class probability $P(c_i)$. This is found by adding the estimated membership values along all of the generated fuzzy sets shown as follows:

$$P(c_i) = \frac{\sum_{k \in A_i, \forall S} \mu_{ik}}{\mathrm{NP}}. \tag{13}$$

Then the marginal class entropy is found by

$$H(C) = -P_{c_i} \log P_{c_i}. \tag{14}$$

The required MI between each feature and the class label is then measured using (3) for each of the features. Finally, to compute the MI between each two variables (to measure the dependency between the two variables), one can multiply the corresponding membership vales from each feature along each fuzzy set and then employ (11) and (12) again to find $I(f_1; f_2)$ by using (3).

B. Proposed Feature-Extraction Method

Given a dataset of collected signal records, the WPT transform is applied on each record of the data acquired from each channel to generate a complete tree up to a level J decomposition. Considering each of the subspaces $\Omega_{j,k}$ of the WPT decomposition tree as a feature space, then features were simply constructed by using the normalized filter bank energy. This is simply formed for each subspace by accumulating the squares of the transform coefficients for that subspace divided by the number of coefficients in that subspace. Additionally, the logarithmic operator was applied to normalize the distribution of the generated features as given below

$$E_{\Omega_{j,k}} = \log\left(\frac{\sum_n (\mathbf{w}_{j,k,n}^T \mathbf{x})^2}{N/2^j}\right) \tag{15}$$

where $E_{\Omega_{j,k}}$ is the normalized logarithmic energy of the wavelet-packet coefficients extracted from the subspace $\Omega_{j,k}$. $\mathbf{w}_{j,k}\mathbf{x}$ is the wavelet-packet transformed signal (or simply the coefficients) evaluated at subspace $\Omega_{j,k}$, and $N/2^j$ is the number of the coefficients in that specific subspace.

For a dataset of n features f_i, where $i = 1, 2, \ldots, n$, one can either utilize the estimated fuzzy MI measure or its normalized variant as

$$F_i = \frac{I(C; f_i)}{H(f_i)} \quad i = 1, 2, \ldots, n. \tag{16}$$

Using (16), calculate the fuzzy-set-based criterion on all the features to evaluate their classification ability. Finally, rank the features according to the aforementioned measure, then remove their ascending and descending nodes or features, and choose the remaining features for classification. The algorithm is then summarized, in a similar manner to that of the OWP and FWP algorithms but with different feature representation and information measure than those utilized by these algorithms, as follows:

Algorithm: The Fuzzy MI-based Wavelet-Packet Algorithm - FMIWPT)

Given a training dataset consisting of labeled original signals,

1) *Step 0:* For each labeled original signal, perform a full WPT decomposition to the maximum level J. For all $j = 0, 1, \ldots, J$ and $k = 0, 1, 2, \ldots, 2^j - 1$, construct features according to (15).
2) *Step 1:* Construct the associated fuzzy sets and compute the fuzzy entropies and MI. Then calculate $F(\Omega_{j,k})$ according to (16), where $\Omega_{j,k}$ is the subspace representing each of the features.
3) *Step 2:* Determine the optimal WPT decomposition $\mathbf{X}^*$, being the one that corresponds to the maximum value of F.
4) *Step 3:* In descending order, sort the subspaces by F, $\mathbf{\Omega} = \{\Omega(1), \Omega(2), \ldots, \Omega(l)\}$. Let $\mathbf{X}^* = \emptyset$.
5) *Step 4:* Move first element in $\mathbf{\Omega}$ to $\mathbf{X}^*$.
6) *Step 5:* $\forall \Omega(k) \in \mathbf{\Omega}$, if $\Omega(k)$ is an ascendant (father) or descendant (child) (direct or indirect) of $\Omega(j_1)$, remove $\Omega(k)$ from $\mathbf{\Omega}$.
7) *Step 6:* if $\mathbf{\Omega} = \emptyset$, stop. Otherwise go to (3) and continue.
8) *Step 7:* The set $\mathbf{X}$ is the final FMIWPT-based decomposition.

The aforementioned algorithm is applied to optimize the WPT tree of each data channel. That is, the aforementioned algorithm is utilized to extract features from each channel and then these features are all concatenated to form one large feature vector that will be used for classification.

IV. Data-Collection Procedure

Thirty-one subjects (volunteer drivers, all males) aged between 20–69 years were recruited to perform a driving simulation task. All participants provided informed consent prior to participating in the study. Lifestyle appraisal questionnaire was used as a selection criteria, which required participants to have no medical contraindications such as severe concomitant disease, alcoholism, drug abuse, and psychological or intellectual problems likely to limit compliance [44]. Most of the studies were conducted between 9:00 AM and 1:00 PM with the total study time involving 2–3 h per subject. This included completing questionnaires, physiological sensor attachment, and performing the driving simulator task. The required ethical approval was obtained from the University Human Ethics Committee. Participants were asked to refrain from consuming caffeine, tea, or food, as well as smoking approximately 4 h and alcohol 24 h before the study, and reported compliance with these instructions.

STISIM driver, from Systems Technology, Inc. (STI), was utilized as driving simulator. The video display showed other cars, the driving environment, the current speed, pedestrians, and other road stimuli. The driving simulator equipment consisted of a large display unit with in-built steering wheel, brakes, and accelerator, as shown in Fig. 2. All subjects were given instructions on the operation of the simulator prior to the study. Two driving sessions were completed by each of the drivers. The initial driving session was approximately 25 min of alert driving, with a track involving many cars and stimuli on the road to serve as the baseline measure. The alert driving session was followed by monotonous driving session, in which participants were required to drive continuously for approximately 1 h. This session involved the participants driving with very few road stimuli in a track resembling country-side driving.

Simultaneous physiological measurements were recorded during the driving sessions. FlexComp Infiniti encoder, from

Fig. 2. Example of the experimental set up.

Thought Technology Inc., was utilized as the physiological data-acquisition device (encoder channel bandwidth ranges from dc to 512 Hz with 2048 samples/sec). The 10–20 international standard of electrode placement was applied [45], with EEG electrodes placed at Fz (frontal), T8 (temporal), and Oz (occipital) sites. The Fz and Oz channels are commonly used in the drowsiness/fatigue studies, while the T8 channel is a well-established region for detecting changes in the drowsiness/fatigue state [17], [46]. All electrodes were referenced to linked earlobes. Vertical EOG was recorded from the left eye, and later used to identify drowsiness by observing the blink rate of each subject. A surface electrode measuring the ECG signal was utilized with the reference electrode placed on the shoulder and the active electrodes placed on the right and the left chest. Blood pressure and heart rate were collected before and after the driving task. The whole data were prefiltered to remove the effect of the line noise (50-Hz noise).

A camera, which was synchronized with the physiological data, was utilized to capture the video image of the vehicle operator's face from which the level of drowsiness was estimated based on the Wierewille and Ellsworth criteria [47]. Specifically, the extracted information from the face data included the facial tone, slow eyelid closure, and mannerisms (rubbing, yawning, nodding, etc.). A continuous drowsiness scale containing five descriptors was utilized. These descriptors are given as: alert or not drowsy (class 1); slightly drowsy (class 2); moderately drowsy (class 3); significantly drowsy (class 4); and extremely drowsy (class 5) [47]. Two final year engineering students volunteered to participate in scoring this study in addition to a postdoctoral fellow, all from the University of Technology, Sydney . The three observers were trained on the scoring criterion by a panel made of a neuroscientist, a senior mechatronics engineer, and a postdoctoral fellow who has got experience in human factors and engineering. The three observers rated the video segments of 1-min duration and assigned a corresponding drowsiness descriptor according to the aforementioned scale. In order to form the final class label for each subject's data segments, a majority voting process was utilized. In such a process, each segment is assigned the label that most of the observers agreed on. The final label of the scored segments was then utilized as the class label for the stages of drowsiness required to train any classification system.

V. Experiments and Results

One of the main advantages of utilizing the proposed FMIWPT method is that it can easily overcome the individual variability that usually affects the choice of the optimum power spectrum components required to identify drowsiness/fatigue. This is justified as the FMIWPT algorithm selects the required components for each subject by estimating the fuzzy MI between these components and the class label that is provided by observer rating of the video frames. In comparison to the FMIWPT algorithm, many methods were proposed in the literature, most of which utilized the well-known $\alpha, \beta, \theta, \delta$ bands or their combinations, e.g., $(\theta+\alpha)/\beta$ and β/α [48], and θ/β and $(\theta+\alpha)/(\alpha+\beta)$ [17], as measures for detecting the current state of the driver. However, the use of the proposed FMIWPT can provide more insight about the required frequency components to detect drowsiness, as it inspects the full spectrum to identify the best frequency components for each specific problem.

In the first phase of the experiments, the performance of the proposed FMIWPT method is tested against other WPT-based feature-extraction methods from the literature, which include the following.

1) The WPT-based feature-extraction method utilized by Zhang *et al.* [24] for drowsiness detection. This method utilizes the relative energies of the WPT subspaces with Shannon entropy as a measure for feature suitability for detecting drowsiness. We refer to this method as HWPT.
2) The optimal wavelet-packet feature-extraction method proposed by Wang *et al.* [32] for classifying biomedical signal classification, that we refer to as OWP. This method was already shown to highly outperform the FWP proposed by Deqiang *et al.* [26]. Thus, no comparison with the FWP was necessary.
3) In order to test the significance of the proposed fuzzy MI measure with respect to its statistical variant, we implement our proposed FMIWPT with MI estimated using the histogram approach, and refer to this method as MIWPT.

The physiological data collected from the five channels (including 3 EEG, 1 EOG, and 1 ECG) were all utilized in this phase of the experiment as the goal here is test the significance of the proposed method rather than the significance of data collected from each of the channels. A windowing approach was then utilized on each of the datasets in which a sliding window of 10 s length incremented each time by 2 s was employed. All of the aforementioned methods were applied on the windowed records of the data to extract the corresponding features. A Symmlet family of wavelets of order 5 was chosen. Given that the five channels of data were utilized in the experiments, there was a need for a dimensionality-reduction method to produce a small feature set to the classifier to reduce the computational cost. For this specific reason, the recently proposed spectral regression (SR)-based linear discriminant analysis (LDA) [49] and its kernel-based version (KSR) [50] were utilized in this

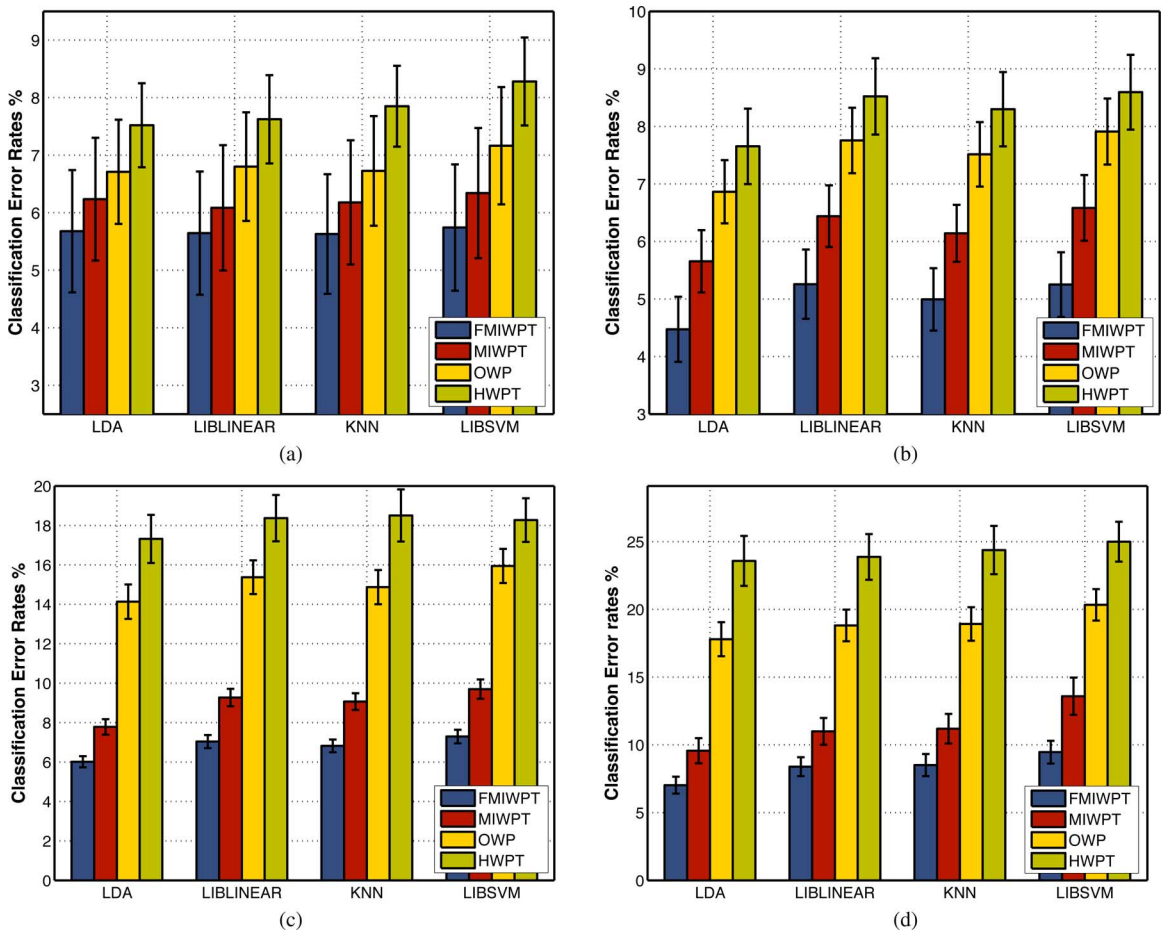

Fig. 3. Classification error rates across each group of subjects with different drowsiness levels using different variants of WPT-based feature-extraction methods with SR projection and different classifiers. (a) Average across three subjects exhibiting two drowsiness levels. (b) Average across 12 subjects exhibiting 3 drowsiness levels. (c) Average across ten subjects exhibiting four drowsiness levels. (d) Average across six subjects exhibiting five drowsiness levels.

paper as two possible variations of dimensionality reduction. These were mainly chosen due to their computational efficiency and powerful performance. However, given that the EEG, ECG, and EOG have very different signal characteristics, e.g., amplitudes and spectral ranges, then two normalization measures were utilized to correct the signal ranges. The first was the application of the logarithmic operator in (15), while the second was to map the resultant data range between a minimum of 0 and a maximum of 1 before attempting to use the dimensionality-reduction methods. In such a case, the features generated from one signal type will not dominate upon those generated from the other signals. Four different classifiers were then utilized to test the classification error rates, these included: the LDA classifier, linear support vector machine classifier (the LIBLINEAR implementation), *k*NN, and the kernel support vector machine classifier (the LIBSVM implementation). The testing scheme utilized in this paper was the two-way data split, i.e., the order of the samples in each data set was randomized and then half of the samples were utilized for training, and the other half, which is unseen during training, was used for testing. This process was repeated 50 times across each of the 31 data set.

Given that different drivers exhibited different levels of drowsiness, then the classification error rates were first given for each group of drivers sharing the same number of drowsiness levels, as shown in Fig. 3 with SR dimensionality reduction and bars denoting the standard errors. In this way, one can judge on the effectiveness of the proposed method when dealing with different number of drowsiness levels. The average classification error rates across 31 subjects using different WPT-based feature-extraction methods are also shown in Fig. 4 with both SR and KSR dimensionality reduction while using different classifiers. This in turn represents the average of the results given in Fig. 3.

These results clearly indicate that the proposed FMIWPT was able to outperform the rest of the methods on the same data sets with different classifiers. The results also indicate that when using KSR for dimensionality reduction, the performance of all of the methods was enhanced upon that offered by employing SR for dimensionality reduction. This is justified by the nonlinear mapping of the features within KSR using the kernel trick. It is also shown that the HWPT and OWP were not able to compete with both of the FMIWPT and its statistical version MIWPT when using either SR or KSR. On the other hand, the MIWPT was only capable of competing with the proposed FMIWPT when using kernel dimensionality-reduction technique. However, when using a linear dimensionality-reduction method, the MIWPT was clearly outperformed by FMIWPT.

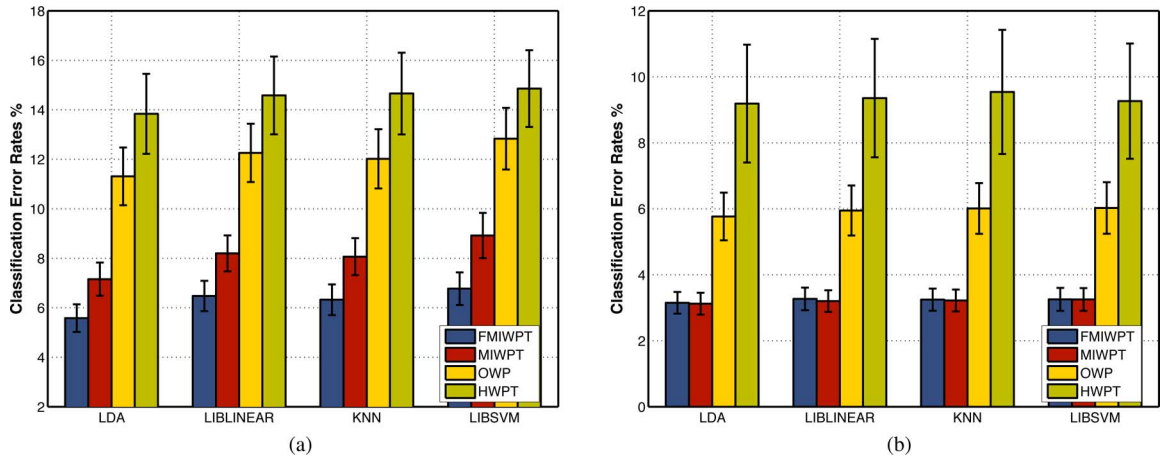

Fig. 4. Average classification error rates across 31 subjects using different variants of WPT-based feature-extraction methods and classifiers. (a) Using SR for dimensionality reduction. (b) Using Kernel-SR for dimensionality reduction.

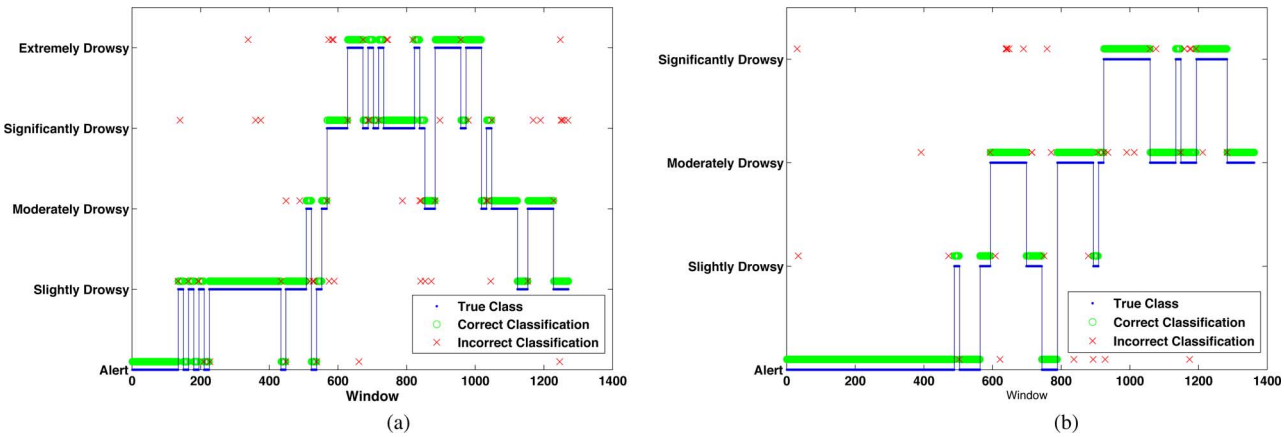

Fig. 5. Classification time plot showing the actual class and the estimated class label by the combination of FMIWPT with SR for dimensionality reduction and LDA classifier. (a) Subject-1 results. (b) Subject-2 results.

A plot of the classification sequence for two subjects is also shown in Fig. 5 using the proposed method with SR dimensionality reduction and LDA classifier. One can see that the proposed combination is quite successful at classifying the features extracted from the EEG + EOG + ECG data. Many of the errors that are present occur during the transitional periods from one drowsiness level to another, which are expected as the system is in an undetermined state between transitions. The rest of the errors occur due to the overlapping stages of drowsiness (mostly removed when using KSR).

In order to test the statistical significance of the achieved results by the proposed FMIWPT in Fig. 4, the following measures are utilized here.

1) *Geometric mean error ratio (GMER)*: For each two feature-extraction methods with associated classification errors $A_1, A_2, \ldots, A_n$ and $B_1, B_2, \ldots, B_n$, respectively (where n represents the number of observations), the geometric error ratio is

$$\exp \frac{\sum_{i=1}^{n} \log(A_i/B_i)}{n}. \tag{17}$$

This measure reflects the relative performance of one method with respect to another. If the outcome is less than 1, then it is an indication that the first method outperforms the second method in terms of error reduction.

2) *Win-tie-loss (WTL):*The measure presents three output values, these are the number of data sets (or subjects) on which the first feature-extraction method presented better, equal, or worse performance than the second feature-extraction method.
3) *Two-way analysis of variance test (ANOVA):*In order to further test the statistical significance, the two-way analysis of variance test (ANOVA with significance level set to 0.05) was again implemented between the testing results achieved by FMIWPT and each of the other four methods across four different classifiers.

Given the different feature-extraction methods, we apply the statistical significance tests on the results achieved by the features projected with SR rather than that with KSR. The justification for this is that all methods perform well with KSR. However, such performance is not related to how good the different features are, but its more related to the ability of KSR in

TABLE I
STATISTICAL SIGNIFICANCE TEST RESULTS OF FMIWPT1 AGAINST ALL OTHER METHODS USING DIFFERENT CLASSIFIERS

FMIWPT1 vs. →	OWP	MIWPT	HWPT
LDA classifier	GMER = 0.4956 WTL = 30-0-1 ANOVA($\rho \leq 0.001$)	GMER = 0.7711 WTL = 28-0-3 ANOVA($\rho \leq 0.001$)	GMER = 0.4199 WTL = 30-0-1 ANOVA($\rho \leq 0.001$)
LIBLINEAR	GMER= 0.5320 WTL = 30-0-1 ANOVA($\rho \leq 0.001$)	GMER = 0.7810 WTL = 27-0-4 ANOVA($\rho \leq 0.001$)	GMER = 0.4584 WTL = 30-0-1 ANOVA($\rho \leq 0.001$)
KNN	GMER= 0.5317 WTL = 30-0-1 ANOVA($\rho \leq 0.001$)	GMER = 0.7774 WTL = 27-0-4 ANOVA($\rho \leq 0.001$)	GMER = 0.4482 WTL = 31-0-0 ANOVA($\rho \leq 0.001$)
LIBSVM	GMER= 0.5283 WTL = 30-0-1 ANOVA($\rho \leq 0.001$)	GMER = 0.7696 WTL = 27-0-4 ANOVA($\rho \leq 0.001$)	GMER = 0.4614 WTL = 31-0-0 ANOVA($\rho \leq 0.001$)

nonlinearly separating the feature along different classes, which is out of context in this paper. Thus, using different feature-extraction methods with the same dimensionality reduction and classification methods, we test the significance of the achieved classification error rates. The results for running the geometric mean error ratio, WTL, and ANOVA tests are all reported in Table I.

Since the achieved geometric mean error rates were all reported to be less than 1 (and in many cases noticeably less than 1), it can be concluded that the proposed FMIWPT was able to outperform all other methods across different classifiers on which the aforementioned results were computed. These results also indicated that the performance of the FMIWPT was statistically more significant, by means of ANOVA, than all other methods while using the LDA, LIBLINEAR, *k*NN, and LIBSVM classifiers. Specifically, FMIWPT outperformed each of the MIWPT, OWP, and HWPT by 1.5737% ($\rho = 0.333 \times 10^{-6}$), 5.7289% ($\rho = 0$), and 8.2572% ($\rho = 0$), respectively, when using the LDA classifier, and by 1.7237% ($\rho = 0.211^{-5}$), 5.7839% ($\rho = 0.0002 \times 10^{-3}$), and 8.1048% ($\rho = 0$), respectively, when using the LIBLINEAR classifier and so on for the rest of the classifiers. This in turn clearly indicates the significance of the proposed method in comparison to other methods.

In the second phase of the experiments, we investigated the effect of employing different combinations of the collected EEG, EOG, and ECG signals for detecting drowsiness. In order to evaluate the significance of using different channel combinations, two sets of experiments were conducted. In the first, the FMIWPT was utilized with SR for dimensionality reduction and LDA classifier on the EEG channels only. Specifically, we evaluate the significance of using the Fz channel alone, Oz channel alone, T8 channel alone, then Fz plus Oz (denoted as Fz + Oz), Fz plus T8 (denoted as Fz + T8), Oz plus T8 (denoted as Oz + T8), and Fz plus Oz plus T8 (denoted as Fz + Oz + T8), as shown in Fig. 6(a) denoting average classification error rates across all subjects with bars denoting the standard errors. These results indicate two important points including firstly that each of the Fz, Oz, and T8 electrodes sites is significant to the drowsiness-detection problem. This in turn agrees upon with the results shown by Jap *et al.* [17] in which the authors proved that significant changes in the $\alpha, \beta, \theta, \delta$ activities existed in the frontal, temporal, and occipital regions. Secondly, despite the effectiveness of each of the utilized channel combination including Fz + Oz, Fz + T8, and Oz + T8, however, practical results indicated significant differences between the results achieved by using all of Fz + Oz + T8 together upon that of the rest of the combinations. Specifically, Fz + Oz + T8 was more significant than Fz + Oz by 2.2625% ($\rho \leq 1 \times 10^{-5}$), Fz + T8 by 2.8607% ($\rho \leq 1 \times 10^{-5}$), and Oz + T8 by 2.0819% ($\rho \leq 1 \times 10^{-5}$). This in turn justifies the need to use all of the three EEG channels together.

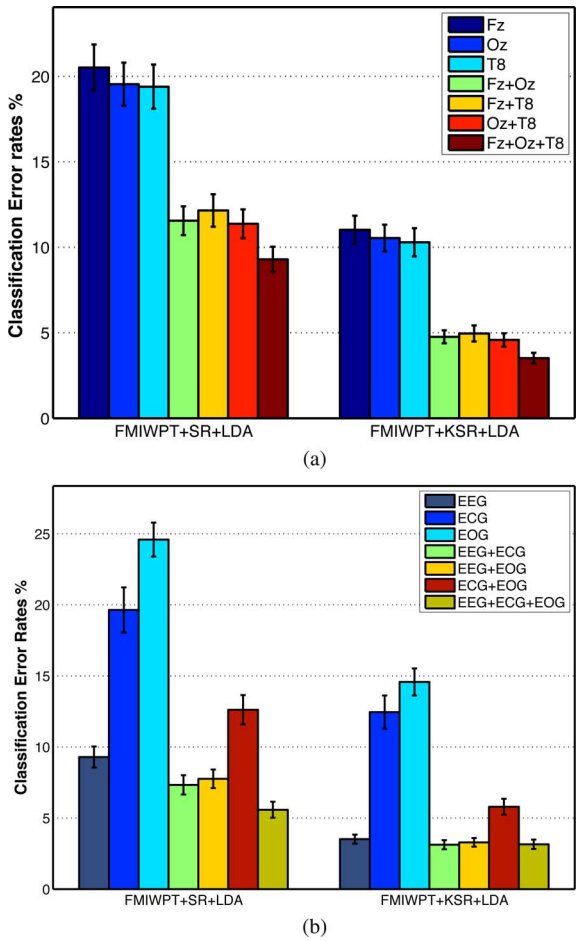

Fig. 6. Average classification error rates across 31 subjects using different channels combinations. (a) Using different EEG channels combinations. (b) Using different EEG, ECG, and EOG channels combinations.

The focus of the experiments was then shifted toward the use of FMIWPT to estimate the classification accuracies when utilizing each of the following channels: EEG (3 channels), ECG (1 channel), EOG (1 channel), EEG (3 channels) plus ECG (1 channel), and EEG (3 channels) plus EOG (1 channel), ECG (1 channel) plus EOG (1 channel) all in comparison to that when using all of the five channels together. In all cases, the classification error rates were calculated using the LDA classifier with SR and KSR as dimensionality-reduction techniques as using other classifiers provided almost the same results. Fig. 6(b) shows the average across all subjects with error bars denoting the standard errors. These results indicate many important points including: first, the EEG channels alone were capable of achieving very promising results as shown previously. These results were further enhanced by adding either EOG or ECG channels resulting in a decrease in the system's error rates, as these channels provided more insight about the actual physiological state of the driver. Second, it can be noticed that, by considering EEG with ECG signal only, the achieved error rates were slightly lower than that achieved by EEG with EOG only across all subjects when using different classifiers classifier with SR or KSR for dimensionality reduction. This in turn further suggests that the information extracted from the ECG signal are more representative than that provided by the EOG signal due to the role of the heart rate variability in detecting drowsiness. Additionally, one can notice that when using the ECG signal alone, the classification error rates , being 19.6377% with SR and 12.4505% with KSR, were actual lower than that achieved by the EOG signal only, being 24.5856% with SR and 14.5758% with KSR. This in turn further proves the effectiveness of the heart rate signal in detecting drowsiness. However, ECG or EOG alone cannot provide very powerful results as that provided by EEG + ECG or EEG + EOG, due to the role of EEG in detecting drowsiness. Finally, a combination of all channels together was shown vital to achieve the best classification error rates when using linear dimensionality-reduction methods achieving an average of 5.5812% with LDA classifier, 6.4770% with LIBLINEAR, 6.3246 % with *k*NN classifier, and 6.5727% with LIBSVM classifier. On the other hand, when using the KSR for dimensionality reduction, it was noticed that the EEG signals alone were enough to achieve accurate results. Adding EOG or ECG signals to EEG while considering KSR for dimensionality reduction, presented no greater statistical significance than using EEG alone. Thus, one can further reduce the number of input channels by considering nonlinear feature-projection techniques. However, using KSR is computationally more expensive than SR, due to the use of the kernel trick.

In the final part of the experiments, the time required by each method to select the best WPT bases for classification was calculated as the average time (per second) required across the different channels. Specifically, the required average best bases selection time for each method was found to be FMIWPT = 1.2273 s, MIWPT = 1.6643 s, OWP = 1.1882 s, and HWPT = 0.9950 s on a computer with i7 processor (1.6 GHz) and 4 GB of RAM. These results indicate that the HWPT required the minimum amount of time followed by OWP, FMIWPT, and MIWPT, respectively. However, given that the performance of the HWPT and OWP was not as accurate as that of the FMIWPT and MIWPT, then this in turn justifies the extra time required by both of these methods, with the proposed FMIWPT requiring less time than the MIWPT with the histogram-based approach. Thus, the computational cost of the proposed FMIWPT seems reasonable enough in comparison to that of the other available methods given the enhancements in the classification results achieved by the FMIWPT.

VI. Conclusion

In this paper, a new feature-extraction algorithm was developed to extract the most relevant features required to identify the driver drowsiness/fatigue states. This was achieved by analyzing the corresponding physiological signals from the brain, eye, and heart. Three observers rated the video segments and the final labeling for the different drowsiness levels was obtained by using majority voting on the observers' scores. A new fuzzy MI estimation scheme was developed and applied to identify the wavelet-packet best bases which were utilized to extract features. These features were then reduced in dimensionality using two methods, SR and KSR, and tested on four different classifiers. The achieved results outperformed the accuracy of state of the art WPT-based feature-extraction methods. The high significance of the statistical analysis supported the accuracy of the results in terms of the classification accuracy rates achieving nearly 95% with SR or nearly 97% with KSR across different classifiers. Finally, a combination of all channels was shown vital to achieve very high classification accuracies when using linear feature-projection methods. On the other hand, using EEG channels alone or a combination of EEG + ECG or EEG + EOG were shown to achieve highly accurate classification results when using nonlinear feature-projection methods like KSR.

Acknowledgment

The authors would like to thank S. Ridwan, L. de Leon, B. Jap (for data collection and assistance), M. D. Antonio, and W. Bastari (for video data analysis).

References

[1] J. D. Slater, "A definition of drowsiness: One purpose for sleep?" *Med. Hypotheses*, vol. 71, pp. 641–644, 2008.
[2] A. Campagne, T. Pebayle, and A. Muzet, "Correlation between driving errors and vigilance level: Influence of the driver's age," *Physiol. Behav.*, vol. 80, no. 4, pp. 515–524, 2004.
[3] D. F. Dinges and N. Kribbs, "Performing while sleepy: Effects of experimentally-induced sleepiness," in *Sleep, Sleepiness, and Performance*, T. Mank, Ed. New York: Wiley, 1991, pp. 98–128.
[4] D. F. Dinges, F. Pack, K. Williams, K. A. Gillen, J. W. Powell, and G. E. Ott, "Cumulative sleepiness, mood disturbance and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 h per night," *Sleep*, vol. 20, pp. 267–277, 1997.
[5] I. D. Brown, "Driving fatigue," *Endeavour*, vol. 6, no. 2, pp. 83–90, 1982.
[6] K. Idogawa, "On the brain wave activity of professional drivers during monotonous work," *Behaviormetrika*, vol. 30, pp. 23–34, 1991.
[7] M. W. Johns, R. Chapman, K. Crowley, and A. Tucker, "A new method for assessing the risks of drowsiness while driving," *Somnologie*, vol. 12, pp. 66–74, 2008.
[8] I. D. Brown, "Car driving and fatigue," *Triangle Sandoz J. Med. Serv.*, vol. 8, pp. 131–277, 1967.

[9] M. H. Chase, "Brain electrical activity and sensory processing during waking and sleep states," in *Principles and Practice of Sleep Medicine*, M. H. Kryger, T. H. Roth, and W. C. Dement, Eds. New York: Saunders, 2000, pp. 93–111.
[10] E. I. Dureman and C. Boden, "Fatigue in simulated car driving," *Ergonomics*, vol. 15, pp. 299–308, 1972.
[11] S. K. L. Lal and A. Craig, "Critical review of the psychophysiology of driver fatigue," *Biol. Psychol.*, vol. 55, pp. 173–194, 2001.
[12] E. Zilberg, Z. M. Xu, D. Burton, M. Karrar, and S. Lal, "Statistical validation of physiological indicators for noninvasive and hybrid drowsiness detection system," *African J. Inf. Commun. Technol.*, vol. 5, no. 2, pp. 75–83, 2009.
[13] Q. Ji, Z. Zhu, and P. Lan, "Real-time nonintrusive monitoring and prediction of driver fatigue," *IEEE Trans. Veh. Technol.*, vol. 53, no. 4, pp. 1052–1068, Jul. 2004.
[14] S. K. L. Lal and A. Craig, "Psychological effects associated with drowsiness: Driver fatigue and electroencephalography," *Int. J. Psychophysiol.*, vol. 55, pp. 183–189, 2001.
[15] T. Akerstedt, G. Kecklund, and A. Knutsson, "Manifest sleepiness and the spectral content of the EEG during shift work," *Sleep*, vol. 14, no. 3, pp. 221–225, 1991.
[16] K. A. Brookhuis and D. Waard, "The use of psychophysiology to assess driver status," *Ergonomics*, vol. 39, no. 9, pp. 1099–1110, 1993.
[17] B. T. Jap, S. Lal, P. Fischer, and E. Bekiaris, "Using EEG spectral components to assess algorithms for detecting fatigue," *Expert Syst. Appl.*, vol. 36, pp. 2352–2359, 2009.
[18] H. Shuyan and Z. Gangtie, "Driver drowsiness detection with eyelid related parameters by Support Vector Machine," *Expert Syst. Appl.*, vol. 36, pp. 7651–7658, 2009.
[19] A. Tsuchida, M. S. Bhuiyan, and K. Oguri, "Estimation of drowsiness level based on eyelid closure and heart rate variability," in *Proc. 31st Annu. Int. Conf. IEEE Eng. in Med. Biol. Soc. (EMBS)*, 2009, pp. 2543–2546.
[20] J. W. Fu, M. Li, and B. L. Lu, "Detecting drowsiness in driving simulation based on EEG," in *Proc. 8th Int. Workshop Auton. Syst.-Self-Org., Manage., Control*, Oct. 6–7, 2008, pp. 21–28.
[21] C. T. Lin, L. W. Ko, I. F. Chung, T. Y. Huang, Y. C. Chen, T. P. Jung, and S. F. Liang, "Adaptive EEG-based alertness estimation system by using ICA-based fuzzy neural networks," *IEEE Trans. Circuits Syst.—I*, vol. 53, no. 11, pp. 2469–2476, Nov. 2006.
[22] M. V. M. Yeo, X. Li, K. Shen, and E. P. V. Wilder-Smith, "Can SVM be used for automatic EEG detection of drowsiness during car driving ?," *Safety Sci.*, vol. 47, pp. 115–124, 2009.
[23] T. P. Jung, S. Makeig, M. Stensmo, and T. J. Sejnowski, "Estimating alertness from the EEG power spectrum," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 1, pp. 60–69, Jan. 1997.
[24] C. Zhang, C. X. Zheng, and X. L. Yu, "Automatic recognition of cognitive and fatigue from physiological indices by using wavelet packet transform and kernel learning methods," *Expert Syst. Appl.*, vol. 26, pp. 4664–4671, 2009.
[25] N. Saito, "Local feature extraction and its applications using a library of bases," Ph.D. Thesis, Department of Mathematics, Yale Univ., New Haven, CT, 1994.
[26] L. Deqiang, W. Pedrycz, and N. J. Pizzi, "Fuzzy wavelet packet based feature extraction method and its application to biomedical signal classification," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 6, pp. 1132–1139, Jun. 2005.
[27] M. Akay, "Wavelet applications in medicine," *IEEE Spectrum*, vol. 34, no. 5, pp. 50–56, May 1997.
[28] R. R. Coifman, Y. Meyer, S. Quake, and V. Wickerhauser, "Wavelet analysis and Signal processing," in *Wavelets and Their Applications*. Sudbury, MA: Jones and Barlett, 1992, pp. 153–178.
[29] K. Englehart, "Signal representation for classification of the transient myoelectric signal," Ph.D. Thesis, Department of Electrical and Computer Engineering, Univ. New Brunswick, Fredericton, NB, Canada, 1998.
[30] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd ed. New York: Academic, 2009.
[31] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 713–718, Mar. 1992.
[32] G. Wang, Z. Wang, W. Chen, and J. Zhuang, "Classification of surface EMG signals using optimal wavelet packet method based on Davies-Bouldin criterion," *Med. Biol. Eng. Comput.*, vol. 44, no. 10, pp. 1741–0444, 2006.
[33] R. Schalko, *Pattern Recognition: Statistical, Structural and Neural Approaches*. New York: Wiley, 1992.
[34] R. M. Fano, *Transmission of Information: A Statistical Theory of Communications*. Cambridge, MA: MIT Press, 1961.
[35] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Boca Raton, FL: CRC Press, 1998.
[36] H. M. Lee, C. M. Chen, J. M. Chen, and Y. L. Jou, "An efficient fuzzy classifier with feature selection based on fuzzy entropy," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 31, no. 3, pp. 426–432, Jun. 2001.
[37] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[38] G. J. Klir, *Uncertainty and Information: Foundations of Generalized Information Theory*. Hoboken, NJ: Wiley, 2006.
[39] A. De Luca and S. Termini, "A definition of a nonprobabilistic entropy in the setting of fuzzy set theory," *Inf. Control*, vol. 20, pp. 301–312, 1972.
[40] Z. P. Chen, J. H. Jiang, Y. Li, Y. Z. Liang, and R. Q. Yu, "Fuzzy linear discriminant analysis for chemical data sets," *Chemometrics Intell. Lab. Syst.*, vol. 45, no. 1–2, pp. 295–302, 1999.
[41] H. M. Lee, C. M. Chen, J. M. Chen, and Y. L. Jou, "Low-level segmentation of aerial images with fuzzy clustering," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-16, no. 4, pp. 589–598, Jul. 1986.
[42] W. Pedrycz, *Knowledge-Based Clustering: From Data to Information Granules*. Hoboken, NJ: Wiley, 2005.
[43] B. Kosko, "Fuzzy entropy and conditioning," *Inf. Sci.*, vol. 40, pp. 165–174, 1986.
[44] A. Craig, K. Hancock, and M. Craig, "The lifestyle appraisal questionnaire: A comprehensive assessment of health and stress," *Psychol. Health*, vol. 11, pp. 331–343, 1996.
[45] H. H. Jasper, "The ten-twenty electrode system of the International Federation," *Electroencephalogr. Clin. Neurophysiol.*, vol. 10, pp. 371–375, 1958.
[46] J. Santamaria and K. H. Chiappa, *The EEG of Drowsiness*. New York: Demos Publications, 1978.
[47] W. W. Wierwille and L. A. Ellsworth, "Evaluation of driver drowsiness by trained raters," *Accident Anal. Prevent.*, vol. 26, no. 5, pp. 571–581, 1994.
[48] H. J. Eoh, M. K. Chung, and S. H. Kim, "Electroencephalographic study of drowsiness in simulated driving with sleep deprivation," *Int. J. Ind. Ergon.*, vol. 34, no. 4, pp. 307–320, 2005.
[49] D. Cai, X. He, and J. Han, "SRDA: An efficient algorithm for large scale discriminant analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 1, pp. 1–12, Jan. 2008.
[50] D. Cai, X. He, and J. Han, "Efficient kernel discriminant analysis via spectral regression," presented at the Int. Conf. Data Mining (ICDM), Omaha, NE, Oct. 2007.

Authors' photographs and biographies not available at the time of publication.