IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Driver Drowsiness Detection Based on Respiratory Signal Analysis

FEDERICO GUEDE-FERNÁNDEZ[1], MIREYA FERNÁNDEZ-CHIMENO[1], (Member, IEEE), JUAN RAMOS-CASTRO[1], (Member, IEEE), AND MIGUEL A. GARCÍA-GONZÁLEZ[1]

[1]Department of Electronic Engineering, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain

Corresponding author: Federico Guede-Fernández (federico.guede@upc.edu)

This work has been partially financed by the Spanish Ministry of Economy, Industry and Competitiveness, DEP2015-68538-C2-2-R.

ABSTRACT Drowsy driving is a prevalent and serious public health issue that deserves attention. Recent studies estimate around 20% of car crashes have been caused by drowsy drivers. Nowadays, one of the main goals in the development of new advanced driver assistance systems is the trustworthy drowsiness detection. In this paper, a drowsiness detection method based on changes in the respiratory signal is proposed. The respiratory signal, which has been obtained using an inductive plethysmography belt, has been processed in real-time in order to classify the driver's state of alertness as drowsy or awake. The proposed algorithm is based on the analysis of the respiratory rate variability (RRV) in order to detect the fight against to fall asleep. Moreover, a method to provide a quality level of the respiratory signal is also proposed. Both methods have been combined to reduce false alarms due to changes of measured RRV associated not to drowsiness but body movements. A driving simulator cabin has been used to perform the validation tests and external observers have rated the drivers' state of alertness in order to evaluate the algorithm performance. It has been achieved a specificity of 96.6%, sensitivity of 90.3% and Cohen's Kappa agreement score of 0.75 on average across all subjects through a leave-one-subject-out cross-validation. A novel algorithm for driver's state of alertness monitoring through the identification of the fight against to fall asleep has been validated. The proposed algorithm may be a valuable vehicle safety system to alert drowsiness while driving.

INDEX TERMS advanced driver assistance systems, driver drowsiness, safety, respiratory signal.

I. INTRODUCTION

DROWSINESS is an intermediate state between wakefulness and sleep that may be defined as the progressive loss of cortical processing efficiency. It is also associated to a desire or inclination to sleep. [1]. Drowsy driving can be caused by a combination of sleep loss, driving when circadian rhythms are low (early morning hours or mid-afternoon) or for long periods of time. Drowsiness affects elements of human performance that are critical to safe driving such as: reaction time, alertness and information processing [2].

Drowsy driving is a prevalent and serious public health issue that deserves attention. The AAA Foundation for Traffic Safety in its 2015 Drowsy Driving Fact Sheet states that for the 2009-2015 period, the percentage of licensed drivers admitting drowsy driving (in the previous 30 days) has remained essentially constant, hovering around 30 percent. In the same report we can find that nearly all (97%) American drivers believe it is somewhat or completely unacceptable for somebody to drive when they are so sleepy that they have trouble keeping their eyes open. However, nearly a third (31.5%) of licensed drivers reported having driven when they were so tired that they had a hard time keeping their eyes open in the last 30 days. More than a fifth (22.3%) admitted to doing this more than once, and 3.5 percent reported having done this fairly often or regularly [3].

Road traffic injuries will rise to the worldwide seventh leading cause of death in 2030 according to the latest WHO's forecast [4]. Recent studies estimate around 20% of car accidents have been caused by drowsy drivers [5]. Moreover, recent research has examined the effects of sleep deprivation in crash rate. It was found that drivers who usually sleep for less than 5 hours daily, and driver who have slept for 1 or more hours less than their usual amount of sleep in the past 24 hours have significantly elevated crash rate. In addition, this estimated crash rate is similar to the risk associated with driving with a blood alcohol concentration equal to or slightly above the legal limit for alcohol [6].

Furthermore, feeling of being sleepy is a gradual process and drivers are not aware of their lack of attention during this time. Thus, a real-time drowsiness driver assessment system

to warn the driver when the first fatigue symptoms appear can avoid crashes by preventing and reducing sleep episodes, either for professional or for particular drivers.

Nowadays, one of the main goals in the development of new advanced driver assistance systems is the trustworthy drowsiness detection. The most widespread automatic drowsiness detection methods may be divided into three main categories based on: driving behaviour, visual and physiological features. Driving behaviour-based methods analyze information about the car position inside the lane, speed, usage of the steering wheel, brakes and gear changes [7]. The main weakness of this method is the variation in accuracy for particular characteristics of the vehicle and driving conditions.

On the one hand, other groups have focused on computer vision methods of driver's body analysis. These systems process eye-state information by calculating percentage of eye closure [8], eye closure duration and the frequency of eye closure [9], [10]. Other systems combine this information with head-movement, yawning and facial expression [11]. The main drawback of these systems is collecting eye-state information properly when the driver wears dark sun-glasses, which does not allow measuring these features properly. This scenario may be frequent in the car environment.

On the other hand, and focusing in sleep monitoring, polysomnography (PSG) is the gold-standard system for sleep disorders diagnosis. This method is based on the biomedical multisignal acquisition and processing, such as electroencephalogram (EEG), electrocardiogram (ECG), galvanic skin response and respiration. PSG is the most used method for clinical purpose because these biomedical signals provide useful information about the physiological response during sleep stages. Therefore, several of these biomedical signals have been proposed as measuring methods for drowsiness detection, the most used are EEG, ECG and respiration. The performance of EEG records based methods have been assessed by several authors [12]–[14], but this method suffers from some serious limitations related both to the EEG sensor ergonomics: it requires sensors and cables on the body that disturbs the driver and the information related to the open-eyes state of the driver disturbs the standard diagnosis methods of PSG. On the one hand, EEG is a low amplitude signal and car environment is electromagnetic noisy which make difficult the signal detection. On the other hand, the mechanical car vibrations may introduce artifacts in the skin-electrode interface, which also complicates the signal detection.

Heart rate variability (HRV) has been usefully employed as a fatigue indicator for athletes [15]. Moreover, in PSG, the extracted HRV from the ECG has been used for determining sleep stages [16]. However, one major drawback of ECG-based method is that it does not seem to be a feasible indicator of sleepiness [17]. A driver drowsiness detection algorithm has been proposed based on HRV analysis and the EEG was used to determinate the sleep onsets by a sleep specialist. The analysis is based on multivariate statistical process control to detect abnormalities in HRV. This work shows that FP rate is 1.7 times per hour on average, but it does not report the false negative rate [18]. Another algorithm for drowsiness detection, which is based on HRV analysis using linear discriminant analysis technique have been developed. It has achieved 98% of specificity, but 59% of sensitivity [19].

Furthermore, a multimodal drowsiness detection method based on the analysis of ECG and EEG has been proposed. These signals were processed using machine learning algorithms such as random forest, multilayer perceptron and support vector machines. A performance about 90% of accuracy, precision and recall has been achieved but it has been tested with only five volunteers [20]. Another, previous work which integrates features of ECG and EEG to detect driver drowsiness has been proposed. These features were combined using SVM and it was achieved a level of accuracy of 80% with 22 healthy subjects [21].

Finally, the analysis of respiratory rate variability (RRV) may be useful to collect accurate information of sleepiness cycles so, respiration-based methods can anticipate this risky situation while driving [22], [23]. Furthermore, there is research on non-contact respiratory signal acquisition systems through thoracic effort monitoring based on camera [24], [25]. This contactless solution will be more accepted by drivers.

The aim of this work is to propose a method for drowsiness detection based on changes in the respiratory signal. The respiratory signal has been obtained using an inductive plethysmography belt and it has been processed in real-time in order to classify the driver's state of alertness as drowsy or awake. This paper assesses the ability of the proposed algorithm to warn the driver when the early fatigue symptoms appear, thus it may be a valuable safety system in car environments to alert to these episodes. The paper is organized as follows: the data collection procedure (Section II), the description of the proposed algorithm (Section III) and the algorithm validation procedure (Section IV). The experimental results obtained and discussion are shown in Section V. Finally, the conclusion section remarks the findings of this paper.

II. METHODS

A. PARTICIPANTS

Twenty adult volunteers (ten females and ten males), aged from 20 to 60 years, participated in the study. 6 subjects were classified underweight (<18.5 kg/m^2), 7 normal weight (18.5 to 25 kg/m^2), and 7 overweight (25 to 30 kg/m^2). All participants were physically healthy and had no history of any sleep disorder, they had a valid driving license and they did not drink alcoholic or energy drink in the past 6 hours before the experiment. This study was performed in accordance with the principles of the Declaration of Helsinki [26] and all participants received detailed information about the study and gave informed consent. Participants should perform two driving tests on different days but four tests were not recorded properly so 36 tests were considered for the study.

(a) (b)

FIGURE 1. Measurement setup for driver drowsiness detection at IBV facilities. (a) Driving simulator room: front screen and car bodywork. (b) Participant wore three RIP bands around: abdomen, diaphragm and chest.

B. MEASUREMENT SETUP

In order to assess the accuracy of the proposed drowsiness detection algorithm, the volunteers of the study should drive while they are fighting against to fall asleep. Since these situations may be dangerous for volunteers, the test was carried on a driving simulator cabin under controlled conditions. Moreover, test conditions have been designed to observe the behaviour of drivers in their fight against to fall asleep while driving.

On the one hand, the experimental protocol is focused on drowsiness detection while driving in well-known hard conditions to keep alertness: night hours, boring and relatively less crowded roadways. This protocol has been setup with a front screen within a simulator with car bodywork in order to provide an immersive experience in the facilities of the Institute of Biomechanics (IBV) in Valencia, Spain as shown in Fig. 1a. The car was equipped with pedals, steering wheel and automatic transmission. A projector was used to display a virtual scenario on a screen in front of the car. The experiments were conducted with the room climate control to 24°C, low lighting and with highway sounds. The simulation scenario was a two-way highway with two lanes in each direction, low density of traffic, night environment, and path with no sharp curves. Moreover, a video camera, which has been focused on the subject's head, has recorded the experiment. The camera device used to record video was the Logitech webcam C120. Video recordings were used by several external observers in order to generate the ground truth driver drowsiness signal.

On the other hand, the breathing process involves several muscles which act on both inhalation and exhalation. Breathing is performed primarily by the diaphragm, a large muscle that separates the thoracic cavity from the abdominal cavity. In fact, the contraction and relaxation of diaphragm produce volume changes in the thoracic and abdominal cavities. Thus, the respiratory signal can be obtained from the tracking of the displacements of diaphragm, abdominal and rib cage. Respiratory inductive plethysmography (RIP) is the most widely accepted non-invasive method for quantitative and qualitative respiratory measurement [27]. Moreover, previous studies have demonstrated that RIP can be also used to quantify the breath-by-breath variability of ventilation [28]. As we can see in Fig. 1b, in our test set-up the subject has worn three RIP band sensors placed in thoracic, diaphragm and abdominal positions in order to guarantee the best signal quality regardless of driver anthropometric characteristics. The respiratory signals have been acquired by three synchronized channels of a Bitmed EximPro monitor (Bitmed®, Sibel S.A., Barcelona, Spain) and the sampling frequency was 40 Hz, which satisfy Nyquist's criteria for respiratory signals. The respiratory sensor was the inductive band SleepSense model. The inductive sensor is a commercial and validated device.

C. TEST PROTOCOL

Participants have performed the driving tests in two different days. In one of them, the subject had sleep deprivation for the last night so subjects have not slept in the 24 hours before the experiment. The other experiment took place with the normal sleep situation, so subjects must have slept at least 6 hours the night before the experiment. The two conditions were randomised and the tests were performed at the time of day at 9 a.m. Then, the difference between both experiments was the subjects' initial sleep condition. Subjects were asked to remain in the simulator car seat, wearing a seatbelt and keeping both hands on the steering wheel with both feet in the pedals. During the first 5 minutes of the experiment, subjects were asked to remain still and quiet. After that, they were asked to drive the simulator car for an hour and a half. Finally, they were asked to remain seated in the driving simulator with their eyes closed for 5 minutes.

D. EXTERNAL OBSERVERS

Trained external observers were asked to rate the subject's state of alertness as "drowsy" or "not drowsy" from face video recordings minute by minute in order to generate the ground truth signal. For this classification, it has been considered that drowsy state of alert is characterized by: loss of facial expressivity, yawns and slow blinks [29]. Otherwise the subject's state of alertness has been classified as "not drowsy". In addition, they were asked to classify minute by minute the quality of the breathing signals as "good" or "bad" from respiratory signal visual inspection.

According to minute by minute ratings of participant's state of alertness, we expect that a correct drowsiness detection places in the same minute that a drowsiness event in the gold standard. However, previous experiments have found that fight against to fall asleep ratings may have some limitations [29]. On the one hand, fight against to fall asleep is a phenomena that involves several events such as sighs and eye blinks that can be clearly identified from video face recordings by an external observer. Moreover, it is complicated for external observers to rate when the first signs of sleepiness appear, such as short periods of long blink duration, some yawns, and some changes in body position. Then, although several sleepiness states can be rated clearly, the starting point sometimes does not seem clear. Therefore, an advanced time (AT) region is defined to reduce the uncertainty in the observers' sleepiness ratings. The scored states which are

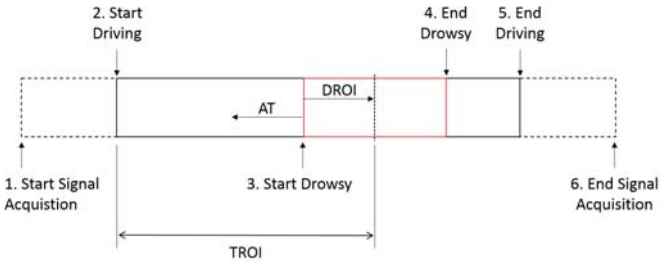

FIGURE 2. Timeline of tracked events over the measurement and defined time region of interest (TROI)

within the AT minutes before the first minute rated as drowsy by the observers, they have remained out of performance results of the drowsiness classification algorithm in order to reduce the uncertainty associated with the observers' ratings.

E. TIME REGION OF INTEREST (TROI)

In order to analyze the performance of the proposed algorithm, some events must be tracked during the measurement test. These events are defined as: event 1 is the beginning of signal acquisition, event 2 occurs when the subject starts driving, the event 3 happens when the subject starts fighting against to fall asleep, the event 4 takes place when the subject wakes up, the event 5 occurs when the subject stops driving and the event 6 is the end of signal acquisition. In addition, AT is the time region from AT minutes before event 3 happens. This time region is used to reduce the uncertainty as detailed above. The drowsiness region of interest (DROI) is the time interval within each subject test which has been identified for drowsiness detection. DROI has been defined from the first minute which the observers rate the subject as drowsy to 5 five minutes later.

The Fig. 2 shows the timeline of the events in the measurement test, it must be remarked that these events do not occur at the same time for all tests. So, the events have been marked for each subject by external observers from video recordings. Although, the time of event 2 and 5 are fixed by the protocol and it may vary from 1 to 2 minutes by subject due to protocol limitations, but 3 and 4 instants depend on how each subject feels during the test.

We have defined the start and endpoint as that which limit the time region of the respiratory signal which will be processed by the designed algorithm. We have established event 2 as a start point because the system is designed to be used under driving conditions. Then, the signal acquired before starting to drive has remained out of the analysis. On the other hand, the end point has been established at the end of DROI. We have considered that the subject's state of alertness is no longer apt to drive after five minutes have passed since the appearance of the first sleepiness symptoms. Then, DROI has been defined to analyze the signal until an affordable time after subjects start to be drowsy, so DROI has been set to 5 minutes for all subjects. The total time interval selected for each test has been called time region of interest (TROI).

TABLE 1. Amount of data available to be analyzed for different inclusion criteria

Inclusion criteria	# of selected tests[a]	Data from 2 to 5[b]			Data within TROI[b]		
		AW min	DW min	DW %	AW min	DW min	DW %
All available subjects	36	2246	1035	31.6	2015	103	4.7
Any drowsiness alert	21	872	1035	54.3	424	103	19.5
Drowsiness alert after 10 first minutes	15	768	592	43.5	404	73	15.3

[a] Number of tests for different constraints of presence of drowsiness.
[b] Amount of respiratory signal data for each alertness state rated by external observers for different time region of test. AW means awake and DW means drowsy.

The analysis of respiratory signal inside the TROI provides more reliable results than the analysis during the entire test because the main goal of the designed algorithm is the detection of the first time frame where signs of somnolence appear. It has been obtained different TROI for each subject based on external observers ratings. Thus, we have included in the analysis the measurements which have a presence of somnolence and the somnolence appears not before 10 minutes after start driving. 10 minutes has been chosen because the initialization of the algorithm takes 5 minutes and the duration of AT, which has been set to the same value as DROI, must be added to obtain reliable results. Table 1 shows the summary of the database: the number of participants, the amount of minutes which have been rated as drowsy and awake for different conditions and TROI. Finally, it must be remarked that the percentage of time that subjects were drowsy shows that more realistic results will be obtained from the analysis during the TROI than from event 2 to 5. Therefore, the proposed algorithm has been optimized and assessed from 15 different tests and 477 minutes of respiratory signal.

III. PROPOSED ALGORITHM

In this section, we will describe the proposed real time algorithm for drowsiness detection by Thoracic Effort Derived Drowsiness index (TEDD). It must be remarked that the proposed TEDD algorithm has been designed to analyze not only thoracic effort signals, but also diaphragmatic and abdominal effort signals. Mainly, the algorithm is based on the analysis of the variability of the respiratory rate along the time, as well as the presence of artifacts in the respiratory signal. Fig. 3 shows the main blocks of the proposed algorithm. The different steps of the drowsiness detection algorithm and respiratory signal quality assessment are discussed next.

The proposed algorithm for drowsiness detection while driving is based on the detection of the fight against to fall

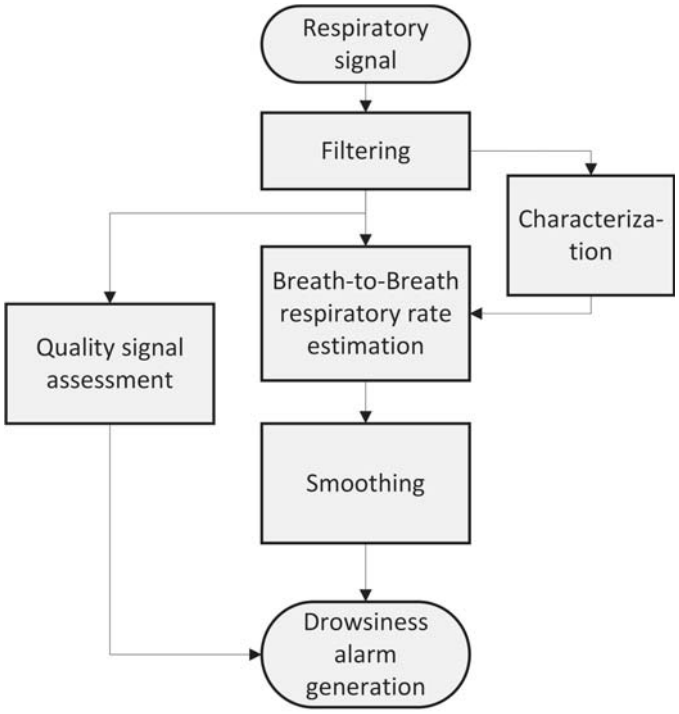

FIGURE 3. Main blocks of the proposed Thoracic Effort Drowsiness Detection (TEDD) algorithm.

asleep from the analysis of RRV. Because of the respiratory signal is acquired with an inductive band around the driver's chest while driving, artifacts can appear in the respiratory signal due to movements, coughing and speaking. Moreover, the respiratory signal amplitude may vary among subjects and along the time so the algorithm is designed based on frequency variations and not amplitude variations. Thus, the signal is filtered and compressed in order to mitigate these effects.

During the episodes of fight against to fall asleep, the RRV increases but it is not equal for everyone. Then, the algorithm searches for each subject the initial level RRV which is taken as reference. The algorithm searches a respiratory signal interval which is characterized by frequency and amplitude stability, mean respiratory rate within the normal respiration rate range and free of artifacts such as snores and movements. The RRV level of the selected signal interval, which is taken as reference, will be used by the following algorithm steps in order to assess the increase of RRV over time. This process has been called characterization, which is performed on-line. The reference signal is used to obtain a threshold to estimate the breathing rate.

Once the reference signal has been obtained, the respiratory rate (RR) is breath-to-breath estimated from the time elapsed between threshold crossings of respiratory signal. The RR has been averaged because it may be misled by several kinds of artifacts which are not related to respiration dynamics. This process was called breath-to-breath frequency estimation.

It has been observed that the RRV increases while the driver is fighting against to fall asleep and the driver's state of alertness decreases slowly. Hence, the RRV has been smoothed and the algorithm searches a sequence of peaks in RRV time series for a period of time in order to identify the fight episodes. The quasi-peak based method described below has been used for this purpose, the whole process was called RRV smoothing and maximum searching.

One of the major drawback of drowsiness detection based on RRV analysis is that high RRV are not only caused by fighting to fall asleep but also driver's movements, speaking and coughing. Therefore, the algorithm assess the quality of respiratory signal from the presence or absence of these events and the quality index has been combined with quasi-peak detection in order to avoid that events which are not related to drowsiness episodes may mislead. Speaking, coughing or other movements happened spontaneously during signal acquisition, without any action taken by the measuring device. Once the signal acquisition was finished, they were detected by visual inspection of the video recordings and the respiratory signal. In addition, the signal quality algorithm has been designed in order for the signal quality to decrease in presence of speaking, coughing and other movement events.

As was mentioned above, the state of alertness decreases slowly while driving and the aim of this algorithm is alert when the subject is not apt to drive, so the algorithm provides a minute-by-minute binary indicator. It has been found an optimized threshold in order to convert the level of one-minute averaged RRV into a binary indicator of drowsiness detection. The threshold is obtained by means of the maximization of the F1-scores of drowsiness detection algorithm. This optimization process is described in Section IV-B and the obtained results are shown in Section V-B1. The accuracy of the algorithm has been assessed from the comparison of this binary indicator with external observers' ratings of driver drowsiness. The mentioned different blocks of the proposed TEDD algorithm are particularly fully described below and illustrated in Fig. 4.

A. RESPIRATORY SIGNAL FILTERING

The normal breathing frequency varies with age and gender. The normal respiratory rate ranges from 14 to 20 breaths per minute in healthy adults [30]. The respiratory signal filtering block consists of a low-pass filter with a cut-off frequency of 0.5 Hz in order to reduce noise spikes, followed by a high-pass filter with a cut-off frequency of 0.05 Hz to remove the baseline of the signal. Both have been designed as a forth order Butterworth filter. The filter coefficients have been calculated as a function of the respiratory signal sampling rate according to digital filter theory. Then, the designed filtering block is suitable to reduce noise and eliminates the baseline in our measurement conditions. Finally, we have applied a non-linear enhancement function to reduce artifacts based on the arctangent that can be described as:

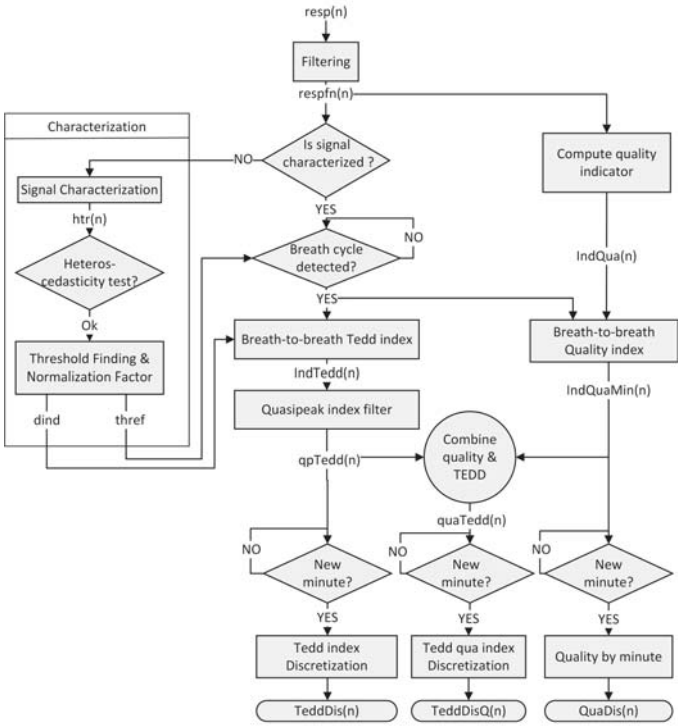

FIGURE 4. Detailed flowchart of TEDD algorithm

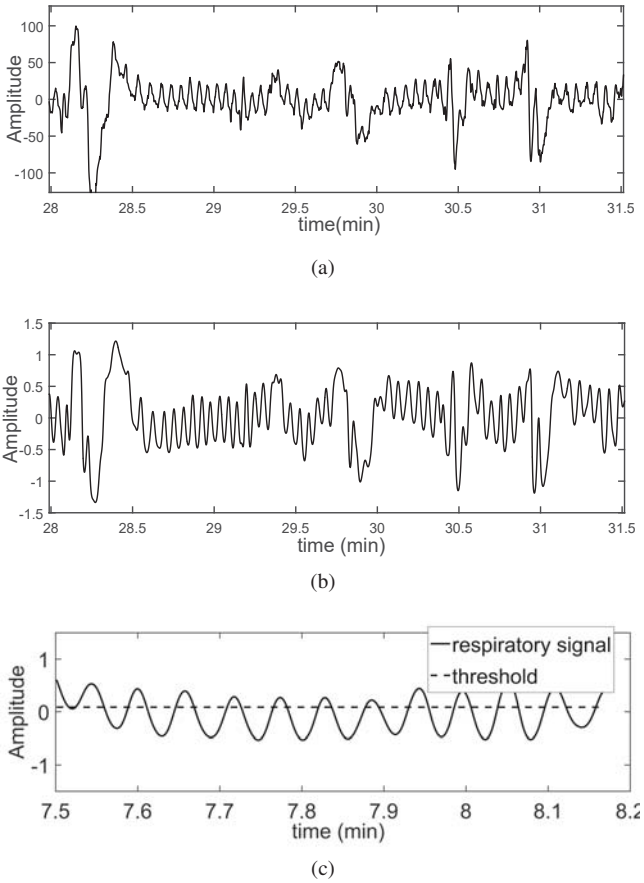

FIGURE 5. Example of one respiratory signal. (a) Raw respiratory signal. (b) Filtered respiratory signal. (c) The respiratory signal which has taken as reference and the obtained threshold

$$respfn(n) = \arctan \frac{respf(n)}{\frac{\sqrt{\sum_{k=1}^{300*f_s}(respf(k)-respfm)^2}*\sqrt{2}}{300*f_s}} \quad (1)$$

where f_s is the sampling frequency, respf is the previously filtered respiratory signal and respfm is the arithmetic mean of respf within the $300*f_s$ first samples. The nonlinear transformation normalizes the signal by the standard deviation and calculates the arctangent of the quotient [31]. In fact, it reduces the amplitude of large artifacts in the respiratory signal. Fig. 5 illustrates an example of raw respiratory signal and the same signal after filtering.

B. CHARACTERIZATION OF SUBJECT RESPIRATION RATE VARIABILITY

This block is in charge of finding the reference pattern of the initial subject's state. It enables us to deal with differences among subjects and different initial subjects' state of alertness. The constant algorithm parameters have been obtained in previous a experiment [32]. The algorithm searches a region of stable respiratory signal as a reference pattern for a maximum of five minutes of the signal from the start of the recording. The estimation of the duration of breath cycles is based on the elapsed time between consecutive threshold crossings of the respiratory signal, so the reference pattern will be also used to set this threshold. Finally, the reference RRV is also obtained from the reference pattern. The respiratory signal is analyzed in intervals of 40 seconds of signal with a sliding window and the pattern finding strategy is based on the analysis of signal stationarity from a heteroscedasticity test [33]. The most stationary signal frame will be selected as reference pattern. On the one hand, the stationarity indicator mHtr has been estimated as:

$$htr(n) = \frac{\sum_{j=0}^{n}(x(j)-\bar{x})^2}{\sum_{k=0}^{N}(x(k)-\bar{x})^2}$$
$$mHtr = \max_{k=1..N} htr(k) \quad (2)$$

where x is respfn, N is the number of samples ($N = 40 * f_s$) and $\bar{x}$ is the mean of respfn from 0 to N. On the other hand, two more indices have been used to find the reference pattern: the mean respiratory rate called mRR in Hz and the standard deviation of the respiratory rate in seconds called stdRR. In this block, the respiratory rate has been obtained from the zero-crossing rate of respfn. The selected reference signal will be the first 40 s interval that meets the next conditions: mHtr must be lower than 0.03, mRR values must be within the range 0.04 Hz to 0.5 Hz and stdRR must be lower than 0.7 s. If the mHtr index of all signal intervals are higher than 0.03, the respiratory interval with the lowest mHtr which satisfies the mRR and stdRR mentioned conditions is selected. In previous experiments,

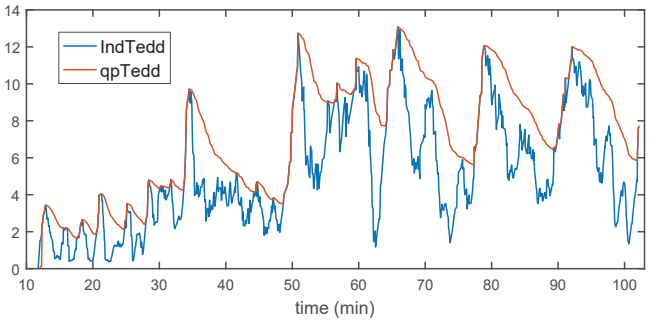

FIGURE 6. Comparison of *IndTedd* and *qpTedd* indices which have been obtained from one participant.

it has been tested that these three conditions are suitable in order to obtain a stable quasi-periodic reference pattern.

It has been used the moving average (MA) filter which is defined as follows:

$$MA_N(x(k)) = \frac{1}{N} * \sum_{j=k-N}^{k} x(j) \tag{3}$$

where x is the input data and N is the length of the moving average filter.

Once the reference signal has been obtained, it is used to obtain the threshold for breath cycle detection called thref. It is obtained from the value of the 60th percentile in the reference respiratory signal. An example of the respiratory signal which has taken as reference and the thref obtained is shown in Fig. 5. Moreover, a reference index of RRV has been computed by the following equation:

$$dind = \frac{1}{M} \sum_{k=5}^{M} |MA_4(T(k)) - MA_4(T(k-1)) \tag{4}$$

where dind is the RRV reference index, M is the number of breathing cycles detected within the reference pattern and T(k) is the breathing duration of the kth detected breath cycle. In addition, k starts at 5 in order to take care of initial transient of 4-point MA. Furthermore, dind has been lower bounded to a basal level, dind=7, in order to normalize stable and periodic respiratory patterns. The algorithm keeps the obtained dind in order to normalize the RRV index obtained along the recording.

C. BREATH-TO-BREATH RESPIRATORY RATE VARIABILITY ASSESSMENT AND DROWSINESS INDICATOR

The purpose of the algorithm is to obtain an indicator that responds to the changes in the respiratory rate associated with fatigue. This response must be cumulative, indicating the increase or decrease of fatigue level. The proposed indicator, which is based on breath-to-breath respiration rate variability (RRV), is described as:

$$\begin{aligned} T_{ma}(n) &= MA_{WLBC}(T(n)) \\ IndTedd(n) &= \frac{MA_{WLD}(|T_{ma}(n) - T_{ma}(n-1)|)}{dind} \end{aligned} \tag{5}$$

where IndTedd is the indicator, WLBC=4 is the window length of the MA of breathing cycles and WLD is the window length of the MA of the difference between successive respiration cycles. The value of WLD parameter will be found by the optimization method. T(n) has been obtained from the time interval between consecutive positive thref crossings and RRV has been computed from the first derivative of the estimated breathing rate. Then, the MA_{WLD} smooths the RRV changes in order to obtain the trend of RRV and it has been normalized by the initial reference state.

After that, a quasi-peak detector (QPD) has been applied to IndTedd in order to estimate RRV changes in a cumulative way. QPD is a peak detection method employed to measure conducted and radiated emissions of electronic equipment in electromagnetic compatibility testing. The QPD has the characteristics of fast-attack (charge time) and slow-decay response (discharge time) [34], which result from low-pass filter R-C components. The qpTedd has been obtained as follows:

$$\begin{aligned} qp(x_n) &= \begin{cases} \alpha * x_n + (1-\alpha) * qp_{n-1}, & \text{for } x_n \geq qp_{n-1} \\ \beta * x_n + (1-\beta) * qp_{n-1}, & \text{otherwise} \end{cases} \\ qpTedd_n &= qp(IndTedd_n) \end{aligned} \tag{6}$$

where the designed QPD has an increase rate faster than the decrease rate defined by α=1 and β=0.02 respectively. Then, qpTedd provides a cumulative indication of RRV changes along the time by obtaining a greater weighting to closer IndTedd peaks. Fig. 6 illustrates the result of applying the described QPD-based method over the IndTedd.

Finally, qpTedd was discretized by averaging over non-overlapped consecutive one-minute intervals. The discretized index has been called TeddDis, which not only takes into account the amplitude of RRV changes but also the time duration of RRV changes. It must be remarked that the events of fight against to fall asleep are rated by several external observers from video recordings of driver's face and these videos are usually rated minute-by-minute [35]. Thus, the TeddDis is an indicator of RRV changes for a specified time interval whose resolution is suitable to be compared with the external observers ratings. Therefore, TeddDis has been used to classify the subject's state of alertness based on decision threshold (ThTedd) which has been optimized by maximizing classification performance defined in Section IV-A.

D. QUALITY SIGNAL ASSESSMENT ALGORITHM AND COMBINATION WITH DROWSINESS DETECTION ALGORITHM

The proposed indicator for respiratory signal quality assessment is based on the analysis of the respiratory waveform changes. The quality indicator (qua) has been defined as follows:

$$qua(n) = \frac{MA_{WLR}(|x(n)|)}{\max\limits_{k=n-WLR\ldots n} x(k) - \min\limits_{k=n-WLR\ldots n} x(k)} \tag{7}$$

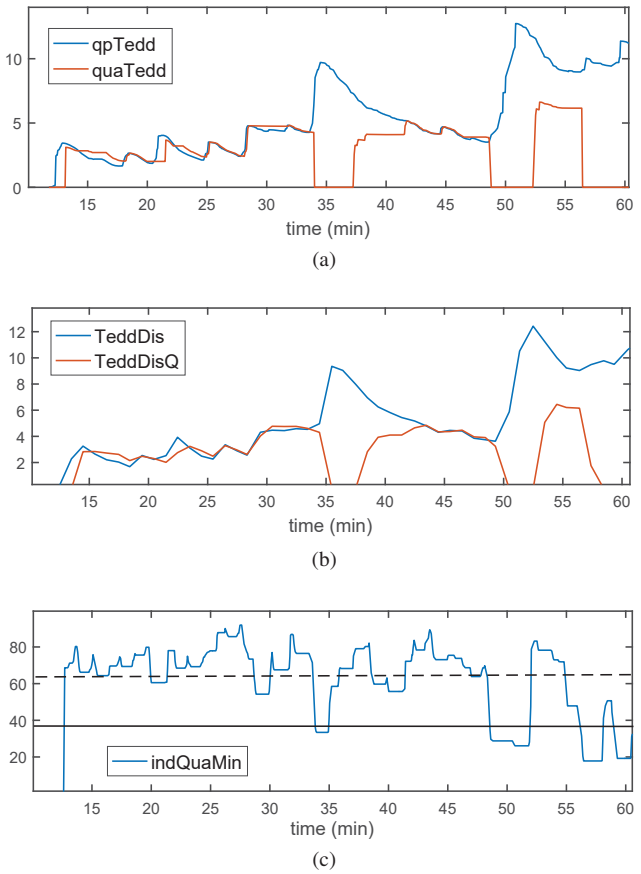

FIGURE 7. An example of results of TEDD algorithm in combination with quality assessment for one test. (a) Comparison of quasi-peak TEDD with and without quality: *quaTedd* and *qpTedd* respectively. (b) Comparison of discretized TEDD (*TeddDis*) and TEDD with quality (*TeddDisQ*). (c) *IndQuaMin* over the time, the dashed line is *QuaTh* and the solid line is *QuaTh*/*2*.

where x is respfn defined in (1), k is the kth respiratory sample and WLR is the window length of MA in samples. The best WLR has been found by optimizing from 20 to 60 seconds. A normalized index quan has been defined in order to analyze changes of quality over the time as:

$$quan(n) = \left| \frac{qua(n)}{quaRef} - 1 \right| \tag{8}$$

where quaRef is the arithmetic mean of qua over the first five minutes. Furthermore, the quan has been filtered in order to smooth sudden changes and it has been scaled. In fact, it can be expressed as:

$$IndQua(n) = 100 * \left(1 - \frac{MA_{WLQ}(quan(n))}{0.6}\right) \tag{9}$$

where WLQ is the window length of MA applied to quan. The best WLQ has been also searched from 20 to 60 seconds. Finally, the assessment of respiratory signal quality has been combined with the assessment of changes in the RRV. Thus, they are combined with a quasi-peak detector block in order to accumulate RRV changes in a suitable way through avoiding wrong estimations due to the low quality of respiratory

TABLE 2. Classifier performance metrics used for tuning parameters and model assessment

Performance Metric	Description
True positive (TP)	# correctly recognized drowsy states
True negative (TN)	# correctly recognized not drowsy states
False positive (FP)	# wrongly recognized not drowsy states
False negative (FN)	# wrongly recognized drowsy states
Sensitivity (SEN)	$\frac{TP}{TP+FN}$
Specificity (SPC)	$\frac{TN}{TN+FP}$
F1-score (F1)	$\frac{2*TP}{2TP+FN+FP}$
Gmean (G)	$\sqrt{SEN*SPC}$
Accuracy (ACC)	$\frac{TP+TN}{TP+TN+FN+FP}$
Cohen's kappa (K)	$pa = \frac{(TN+FP)*(TN+FP)}{TP+TN+FN+FP}$ $pb = \frac{(TP+FN)*(TP+FP)}{TP+TN+FN+FP}$ $p_e = \frac{pa+pb}{(TP+TN+FN+FP)}$ $K = \frac{ACC-p_e}{1-p_e}$

signal. $quaTedd_n$ denotes the combination of TEDD and quality indices, which has been defined as follows:

$$IndQuaMin_n = \min_{k=n_{NBC}\ldots n} IndQua(k)$$
$$quaTedd_n = \begin{cases} y_n, & \text{if } x_n \geq QuaTh \\ y_{n-1}, & \text{if } \frac{QuaTh}{2} \leq x_n < QuaTh \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

where x_n is $IndQuaMin_n$, y_n is $qpTedd_n$ and n_{NBC} is NBC breath cycles before n. IndQua has been minimized taking previous samples in order to obtain a conservative indicator because poor respiratory signal patterns may appear in consecutive breath cycles. As in qpTedd discretization, quaTedd has been discretized by averaging over non-overlapped consecutive one-minute intervals. The discretized index has been called TeddDisQ. Fig. 7 illustrates the effect of the signal quality assessment combination with qpTedd over time for one subject test.

IV. ALGORITHM PERFORMANCE EVALUATION

A. PERFORMANCE METRICS

The ratings generated by external observers were used as reference in order to assess the correctness of the proposed algorithm for drowsiness detection. In fact, it has been evaluated by computing the number of correctly recognized drowsy states (true positives), the number of correctly recognized not drowsy states (true negative) and states that either were incorrectly classified as drowsy (false positive) or that were not

recognized as drowsy (false negative). These counts can be represented as a matrix which is known as confusion matrix [36]. The metrics which have been used in order to evaluate the performance of TEDD algorithm classifier, are shown in Table 2. Sensitivity (SEN) and specificity (SPC) are well-known metrics most used joint in order to evaluate the performance of binary classification algorithms. On the one hand, SEN measures the proportion of positives, in our case drowsiness detection, which are correctly detected by the algorithm and SPC measures the proportion of negatives (not drowsy detection) which are properly identified. On the other hand, we have also used other performance metrics such as: F1-score (F1), G-mean (G) and Cohen's kappa coefficient (K). G is insensitive to changes in class distribution; on the contrary, F1 is well suited for tasks with lots of negative instances (not drowsy). K coefficient is a measure of agreement between classifiers and it has been proposed the following as standards for strength of agreement: $\leq$0=poor, .01–.20=slight, .21–.40=fair, .41–.60=moderate, .61–.80=substantial, and .81–1=almost perfect [37]. In addition, K has been used to measure the classifier performance, because it has measured the agreement between predicted and observed subjects' state of alertness, while correcting for an agreement that occurs by chance [38].

B. ALGORITHM OPTIMIZATION AND VALIDATION

The data have been partitioned into a training set and test set in order to obtain an unbiased estimation of accuracy of the optimized algorithm. Moreover, the training data must be partitioned into a training set and validation set again in order to get an unbiased estimation of accuracy during the optimizing process (e.g. to select the best size of WLD).

One of the main limitations of a single data partition is that we may not have enough data to make sufficiently large training, validation and test sets. Therefore, we have performed a cross-validation method to use the available data for validation and testing efficiently. A leave-one-subject-out cross-validation (LOSOCV) was used to select the best model parameters and to assess the performance of the selected drowsiness classifier. In LOSOCV technique, all but one tests' data were used to train the classifier and the remained test's data were used for testing. This process was repeated for every subject until each subject has been a test sample once [39]. Then, this cross-validation method is suitable for our scenario because any subject's data are not wasted.

The performance results of the optimized algorithm will be so optimistic because the cross-validation results were subject to maximization. Therefore, nested-loop cross-validation has been performed to combine model selection and performance evaluation. The method consists in two nested cross-validation loops: an inner one to select the algorithm parameters with the best cross-validation performance and an outer one to assess the tuned model on the remaining test set. Then, the results of the drowsiness classification algorithm with selected parameters have been obtained for the estimation of the performance metrics. The pseudo-code of the LOSOCV procedure has been described in Fig. 8.

1: Divide dataset in N folds
2: **for** $i = 1$ to N **do**
3: Use N-1 as validation set
4: Use the remaining as test set
5: Divide the validation set in N-1
6: **for** $j = 1$ to $N - 1$ **do**
7: Use N-2 to train the classifier
8: Use the remaining to evaluate the classifier
9: **end for**
10: Select the classifier with best AUC
11: Build the classifier selected with the validation set
12: Store the predictions in the test set
13: **end for**
14: Estimate the mean performance metrics in the test set

FIGURE 8. Nested LOSOCV algorithm for model selection and assessment

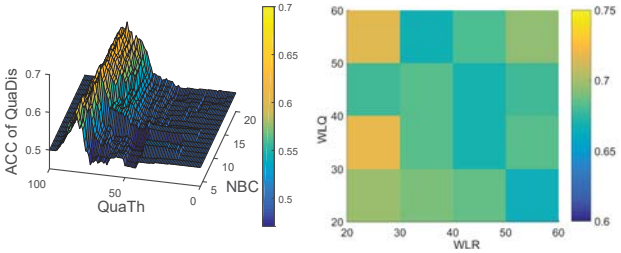

FIGURE 9. (a) An example of ACC of quality signal classification algorithm for different decision threshold (QuaTh) and NBC parameter of IndQuaMin. In this example, the other parameters that are involved in the optimization process have been fixed (WLR= 20 and WLQ= 20). (b) F1 classification results for different values of *WLR* and *WLQ* parameters. The *NBC* and *QuaTh* parameters have been optimized for each *WLR* and *WLQ*.

V. RESULTS AND DISCUSSION

A. QUALITY SIGNAL ASSESSMENT ALGORITHM: OPTIMIZATION AND PERFORMANCE EVALUATION

The algorithm performs a binary classification of respiratory signal quality as good or bad, which is based on QuaDis parameter, which has been defined in Section III-D. The classification performance results were used to optimize the algorithm parameters. These results have been obtained by comparing the classifications of the quality signal assessment algorithm with the quality signal rated by the external observers.

The optimized parameters are: WLR, WLQ, NBC and QuaTh. The range values of the parameters WLR and WLQ, have searched from 20 to 60 seconds in order to comprise several breath cycles without too much algorithm delays. Then, we have evaluated the classification results with different values of these parameters to obtain the best results. The algorithm has been optimized by following bottom-up steps. Firstly, the QuaTh which maximizes the ACC has been found in each one of the other parameters. An example is shown

TABLE 3. Performances metrics of quality signal classification algorithm

	Performance Metrics				Optimized parameters			
	ACC	SPC	SEN	F1	WLR	WLQ	NBC	QuaTh
MA	0.63	0.36	0.81	0.70	20	50	11	75
MI	0.63	0.35	0.83	0.72				

in Fig. 9a. It shows the ACC of classification for different QuaTh and NBC cycles when WLQ and WLR are fixed. Secondly, the obtained QuaTh has been used to find the best NBC that maximizes F1 for each combination of WLQ and WLR. In case of similar F1 results have been obtained for different NBC values, the lowest NBC has been selected in order to reduce the complexity of the algorithm. Thirdly, the WLR and WLQ which provides the best F1 results have been found. Fig. 9b shows the F1 values of these parameters with NBC previous optimized.

On the other hand, F1 is more sensitive to good quality classification and G provides balanced results for good and bad quality classification. Then, F1 is known to give a trade-off between high recall rates and accuracy for predictions, so the parameters have been optimized following the F1 maximization criterion. Table 3 summarizes the performance results of the signal quality classification algorithm. The table shows micro-average (MI) and macro-average (MA) performance results because while the MA-metrics are used when training independent classifiers per class, MI-metrics are used when all classes are trained jointly. Between MI and MA, MI have been used for tuning parameters because the same classifier was trained for all subjects and MA results have been used in order to assess the overall performance without bias the most populated ones. Table 3 shows that the obtained MI and MA results are quite similar. Furthermore, this optimized quality algorithm will have applied to TEDD algorithm in order to obtain more reliable drowsiness classification results.

B. TEDD ALGORITHM: DROWSINESS CLASSIFICATION RESULTS

1) TEDD algorithm parameters tuning

The proposed algorithm has been optimized by tuning WLD and ThTedd parameter, which has been defined in Section III-C, in order to obtain the algorithm with the best drowsiness classification results. The LOSOCV method was used to select the best WLD and ThTedd and to assess the performance of the selected drowsiness classifier. Therefore, each different WLD has been considered as a different model so the WLD with the best results has been selected in the inner loop whereas the generalized performance results have been obtained in the outer loop. The WLD ranges from 1 to 20, because higher to 20 may cause undesirable delays. The ThTedd was optimized within the range 0 to 15 and 0.025 step, the range is widely enough and the step has enough resolution for optimizing the algorithm. Fig. 10 shows the

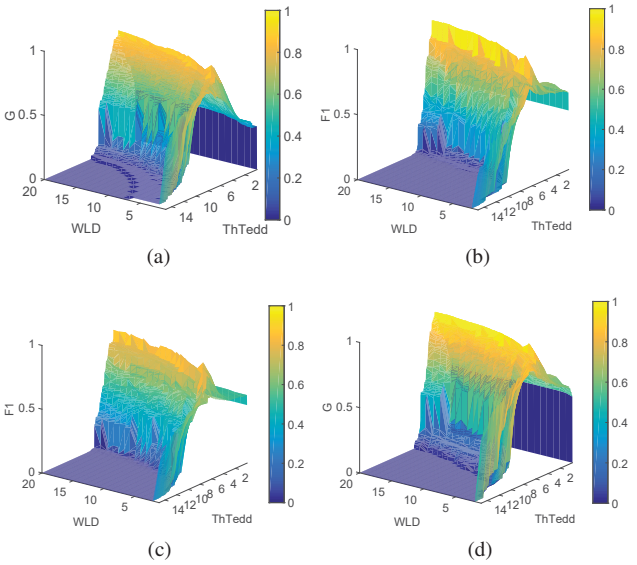

FIGURE 10. Drowsiness classification results of TEDD by varying *ThTedd* and *WLD* values for finding the best parameters. (a) and (c) G and F1 results of the TEDD without quality signal algorithm respectively. (b) and (d) G and F1 results of the TEDD with quality signal algorithm respectively

G and F1 of TEDD algorithm with and without quality signal assessment varying WLD and ThTedd values during tuning phase in the validation set. This figure also reveals that different optimal ThTedd have been found for each WLD parameter value. Whereas the best ThTedd has been searched by maximizing G or F1, the WLD optimization maximizes the area under the ROC curve (AUC). AUC is an overall indicator of classifier performance independently of the threshold decision chosen, in our case ThTedd. The AUC has been computed by the Wilcoxon–Mann–Whitney statistic, which is more efficient than integrating the ROC curve.

2) Evaluation of TEDD algorithm results

The results of drowsiness detection algorithm combined with the quality respiratory signal classification algorithm and without this combination have been obtained. These results have been obtained in order to assess the improvement of the contribution of the quality signal classification algorithm. Moreover, it have been assessed the effect of choosing G or F1 maximization criteria for optimizing ThTedd. The confidence intervals of each performance classification metrics have been obtained to compare and evaluate different algorithms. The uncertainty of the accuracy must be assessed taking into account inter-subject variability effects. The random effects analysis method has been used to estimate the average and confidence intervals of algorithm performance metrics taking into account this variability. Therefore, the average and 95% confidence interval of the cross-validation algorithm results over the folds have been obtained from pooling the across all subjects by using random effects analysis [40]. The values of these performance metrics for

TABLE 4. Drowsiness detection performance results for different advanced time (AT). TEDD algorithm (TeddDis) and TEDD combined with signal quality assessment (TeddDisQ)

TEDD method & max. criteria	AT=5					AT=4				
	SPC	SEN	F1	G	K	SPC	SEN	F1	G	K
TeddDis F1	80.2 ± 10.1	89.9 ± 14.9	77.2 ± 15.0	84.9 ± 8.8	0.59 ± 0.22	75.7 ± 10.5	89.9 ± 14.9	73.7 ± 17.6	82.5 ± 8.9	0.55 ± 0.23
TeddDis G	80.2 ± 10.1	89.9 ± 14.9	77.2 ± 15.0	84.9 ± 8.8	0.59 ± 0.22	75.7 ± 10.5	89.9 ± 14.9	73.7 ± 17.6	82.5 ± 8.9	0.55 ± 0.23
TeddDisQ F1	96.6 ± 3.6	90.3 ± 14.3	83.7 ± 12.7	93.4 ± 7.6	0.75 ± 0.19	93.1 ± 5.1	90.3 ± 14.3	80.4 ± 14.0	91.7 ± 7.7	0.69 ± 0.20
TeddDisQ G	96.6 ± 3.6	90.3 ± 14.3	83.7 ± 12.7	93.4 ± 7.6	0.75 ± 0.19	93.1 ± 5.1	90.3 ± 14.3	80.4 ± 14.0	91.7 ± 7.7	0.69 ± 0.20

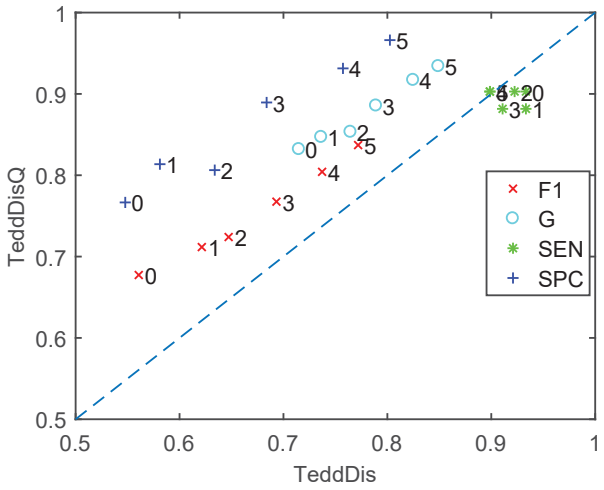

FIGURE 11. Performance metrics of TEDD drowsiness classification varying AT for *TeddDis* and *TeddDisQ*. The F1 maximization criteria has been used to find the most suitable *ThTedd*. The number over each symbol indicates the value of AT

different TEDD method (TeddDis and TeddDisQ), different maximization criteria (F1 and G) and for AT=5 and AT=4 are summarized in Table 4. On the one hand, this table shows that the same classification performance have been achieved using both maximization criteria.

On the other hand, Welch's t-test has been used to test the significance of differences between TeddDis and TeddDisQ methods. This test is commonly used to compare the means of two independent samples with unequal variances. It is an adaptation of Student's t-test and is more reliable when the variances of the two samples are not equal [41]. In fact, it has been found that SPC, F1, G, K obtained from TeddDisQ for both optimization criteria are significantly better than self TeddDis for both AT ($p<0.05$) while keeping the same SEN. Furthermore, this table shows that SPC of TeddDisQ has increased from 80.2% to 96.6% while keeping the same sensitivity around 90%. Finally, the obtained TeddDis K shows a moderate agreement while substantial agreement has been obtained with TeddDisQ.

Fig. 11 shows several performance metrics of classification results for TeddDis and TeddDisQ by varying AT. Both of them have been optimized with the same criterion (maximization F1) in order to compare only the effect of AT changing. The AT ranges from 5 minutes to 0 minutes and the number over the marker indicates the number of minutes of AT. This figure shows that SPC, F1 and G decrease when AT decreases, because of the FP increase near to drowsiness event keeping the same SEN for TeddDis and TeddDisQ. Moreover, the performance of TeddDisQ is better than TeddDis for all different AT values. Furthermore, it must be remarked that AT only affects to negative labels (TN increase or FP increase) so the increase of the number of FP close to drowsiness event can poor the SPC result. In fact, for AT=5 the obtained FPR (FPR=1-SPC) is 3% whereas the obtained FPR increases to 23% for AT=0.

The results obtained with the proposed method can only be compared with results obtained from other methods which do not use the respiratory signal, for example EEG, due to the novelty of the respiratory signal processing for driving drowsiness detection. The method [12] based on EEG records achieved an accuracy of 87.4% for drowsiness state and 83.6% for awake state. These results are worse than the ones obtained by the proposed method. The method [19] based on HRV variability obtained a high specificity (98%) but the sensitivity is poor (59 %). A method based on both EEG and ECG obtained a performance about 90% of accuracy, precision and recall but this performance has been evaluated with only five tests [20]. In our method the sensitivity is around 90% but our specificity is higher (96.6%) and the proposed method was tested with 15 tests.

C. STUDY LIMITATIONS

One limitation of this work is that no prior approaches of driver drowsiness detection based on respiratory signal analysis were found to compare the obtained results. Moreover, the results may be somewhat limited by the controlled experiment conditions because of the experiments were performed in a driving simulator. Then, a further study with more focus on real driving drowsiness detection is therefore suggested. Furthermore, this study was limited by the number of subjects that were 15 people and by the amount of data analyzed that was 474 minutes of respiratory signal.

VI. CONCLUSION

A novel algorithm (TEDD) for the driver's state of alertness monitoring through the identification of the fight against to fall asleep has been proposed based on the analysis of changes of RRV. Moreover, it has been also proposed a method to provide a confidence quality level of the respiratory signal. In addition, the obtained quality signal level has been combined with the drowsiness detection algorithm in order to improve the classification results by means of reducing the number of false positives due to changes of measured RRV associated not to drowsiness but body movements or talking.

Furthermore, the designed algorithm has been validated under a driving simulator and several sleepiness episodes have been identified for each one of 15 test sessions. The gold-standard of drowsiness events has been generated by external observers from video recordings of the subject while driving. The performance results have been obtained following LOSOCV procedure in order to achieve an unbiased estimate of the generalized algorithm performance. The improvement of drowsiness classification results due to the signal quality algorithm has been also assessed and the differences for each optimization criteria and AT have been discussed.

In summary, the obtained results are similar for different training criteria but significant results improvement have been found with respiratory signal quality combination. Therefore, the algorithm can be maximized with different criteria as long as the criteria were insensitive to class distribution. Finally, the optimal values for drowsiness detection TEDD algorithm parameters have been: WLD=17 and ThTedd=3.025.

Finally, the generalized performance of drowsiness detection has been assessed achieving a specificity of 96.6% $\pm$ 3.6%, sensitivity of 90.3% $\pm$ 14.3% and Cohen's Kappa agreement score of 0.75$\pm$ 0.19 on average across all subjects. Therefore, the proposed algorithm may be a valuable vehicle safety system to alert drowsiness while driving.

ACKNOWLEDGMENT

The authors would like to thank FICOSA International for helping to develop this work.

REFERENCES

[1] J. D. Slater, "A definition of drowsiness: One purpose for sleep?" Medical Hypotheses, vol. 71, no. 5, pp. 641–644, nov 2008. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0306987708002946

[2] N. H. T. S. Administration and Others, "Drowsy driving and automobile crashes," NCSDR/NHTSA Expert Panel on Driver Fatigue and Sleepiness. DOT Report HS, vol. 808, p. 707, 1998.

[3] L. S. Arnold and B. C. Tefft, "Prevalence of Self-Reported Drowsy Driving, United States: 2015," AAA Foundation for Traffic Safety, Tech. Rep., 2015. [Online]. Available: https://www.aaafoundation.org/sites/default/files/PrevalenceOfSelf-ReportedDrowsyDrivingReport.pdf

[4] World Health Organization. Violence and Injury Prevention and World Health Organization, Global status report on road safety 2015. World Health Organization, 2015.

[5] P. Jackson, C. Hilditch, A. Holmes, N. Reed, N. Merat, and L. Smith, "Fatigue and road safety: a critical analysis of recent evidence," London: Department for Transport, no. 21, 2011.

[6] B. C. Teff, "Acute Sleep Deprivation and Risk of Motor Vehicle Crash Involvement," AAA Foundation for Traffic Safety, Tech. Rep. December, 2016. [Online]. Available: https://trid.trb.org/view.aspx?id=1436965

[7] A. D. McDonald, J. D. Lee, C. Schwarz, and T. L. Brown, "Steering in a random forest: ensemble learning for detecting drowsiness-related lane departures." Human factors, vol. 56, no. 5, pp. 986–998, aug 2014.

[8] W. W. Wierwille, S. S. Wreggit, C. L. Kirn, L. A. Ellsworth, and R. J. Fairbanks, "Research on vehicle-based driver status/performance monitoring; development, validation, and refinement of algorithms for detection of driver drowsiness. final report," U.S. Department of Transportation, National Highway Traffic Safety Administration, Tech. Rep., 1994.

[9] T. D'Orazio, M. Leo, C. Guaragnella, and A. Distante, "A visual approach for driver inattention detection," Pattern Recognition, vol. 40, no. 8, pp. 2341–2355, aug 2007. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0031320307000544

[10] J. Jo, S. J. Lee, K. R. Park, I.-J. Kim, and J. Kim, "Detecting driver drowsiness using feature-level fusion and user-specific classification," Expert Systems with Applications, vol. 41, no. 4, pp. 1139–1152, 2014.

[11] S. J. Lee, J. Jo, H. G. Jung, K. R. Park, and J. Kim, "Real-Time Gaze Estimator Based on Driver's Head Orientation for Forward Collision Warning System," IEEE Transactions on Intelligent Transportation Systems, vol. 12, no. 1, pp. 254–267, mar 2011. [Online]. Available: http://ieeexplore.ieee.org/document/5688323/

[12] A. Garces Correa, L. Orosco, and E. Laciar, "Automatic detection of drowsiness in EEG records based on multimodal analysis." Medical engineering & physics, vol. 36, no. 2, pp. 244–249, feb 2014.

[13] J. W. Fu, M. Li, and B. L. Lu, "Detecting Drowsiness in Driving Simulation Based on EEG," in Autonomous Systems – Self-Organization, Management, and Control: Proceedings of the 8th International Workshop held at Shanghai Jiao Tong University, Shanghai, China, October 6–7, 2008. Dordrecht: Springer Netherlands, 2008, pp. 21–28. [Online]. Available: http://dx.doi.org/10.1007/978-1-4020-8889-6_3

[14] M. Hajinoroozi, Z. Mao, T. P. Jung, C. T. Lin, and Y. Huang, "EEG-based prediction of driver's cognitive performance by deep convolutional neural network," Signal Processing: Image Communication, vol. 47, pp. 549–555, sep 2016. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0923596516300832

[15] L. Schmitt, J. Regnard, M. Desmarets, F. Mauny, L. Mourot, J.-P. Fouillot, N. Coulmy, and G. Millet, "Fatigue shifts and scatters heart rate variability in elite endurance athletes." PloS one, vol. 8, no. 8, p. e71588, 2013.

[16] M. Ako, T. Kawara, S. Uchida, S. Miyazaki, K. Nishihara, J. Mukai, K. Hirao, J. Ako, and Y. Okubo, "Correlation between electroencephalography and heart rate variability during sleep." Psychiatry and clinical neurosciences, vol. 57, no. 1, pp. 59–65, feb 2003.

[17] J. van den Berg, G. Neely, U. Wiklund, and U. Landström, "Heart rate variability during sedentary work and sleep in normal and sleep-deprived states." Clinical physiology and functional imaging, vol. 25, no. 1, pp. 51–7, jan 2005. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15659081

[18] K. Fujiwara, E. Abe, K. Kamata, C. Nakayama, Y. Suzuki, T. Yamakawa, T. Hiraoka, M. Kano, Y. Sumi, F. Masuda, M. Matsuo, and H. Kadotani, "Heart Rate Variability-based Driver Drowsiness Detection and its Validation with EEG." IEEE transactions on bio-medical engineering, nov 2018.

[19] J. Vicente, P. Laguna, A. Bartra, and R. Bailon, "Drowsiness detection using heart rate variability." Medical & biological engineering & computing, vol. 54, no. 6, pp. 927–937, jun 2016.

[20] Y. Kim, M. Yeo, I. Sohn, and C. Park, "Multimodal Drowsiness Detection Methods using Machine Learning Algorithms," IEIE Transactions on Smart Processing & Computing, vol. 7, no. 5, pp. 361–365, oct 2018.

[21] M. Awais, N. Badruddin, and M. Drieberg, "A Hybrid Approach to Detect Driver Drowsiness Utilizing Physiological Signals to Improve System Performance and Wearability." Sensors (Basel, Switzerland), vol. 17, no. 9, aug 2017.

[22] Z. Mao, X.-p. Yan, and C.-z. Wu, "Driving Fatigue Identification Method Based on Physiological Signals," in Plan, Build, and Manage Transportation Infrastructure in China. ASCE, 2008, pp. 341–352.

[23] H. De Rosario, J. S. Solaz, N. Rodriguez, and L. M. Bergasa, "Controlled inducement and measurement of drowsiness in a driving simulator," IET intelligent transport systems, vol. 4, no. 4, pp. 280–288, 2010.

[24] M. Bartula, T. Tigges, and J. Muehlsteff, "Camera-based system for contactless monitoring of respiration," in 2013 35th Annual International

Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2013, pp. 2672–2675.
[25] J. Solaz, J. Laparra-Hernández, D. Bande, N. Rodríguez, S. Veleff, J. Gerpe, and E. Medina, "Drowsiness Detection Based on the Analysis of Breathing Rate Obtained from Real-time Image Recognition," Transportation Research Procedia, vol. 14, pp. 3867–3876, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2352146516304793
[26] World Medical Association, "World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects." Bulletin of the World Health Organization, vol. 79, no. 4, pp. 373–374, jul 2001. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2566407/
[27] P.-Y. Carry, P. Baconnier, A. Eberhard, P. Cotte, and G. Benchetrit, "Evaluation of Respiratory Inductive Plethysmography: Accuracy for Analysis of Respiratory Waveforms," Chest, vol. 111, no. 4, pp. 910–915, 1997.
[28] M. N. Fiamma, Z. Samara, P. Baconnier, T. Similowski, and C. Straus, "Respiratory inductive plethysmography to assess respiratory variability and complexity in humans," Respiratory Physiology and Neurobiology, vol. 156, no. 2, pp. 234–239, 2007.
[29] A. Anund, C. Fors, D. Hallvig, T. Åkerstedt, and G. Kecklund, "Observer Rated Sleepiness and Real Road Driving: An Explorative Study," PLoS ONE, vol. 8, no. 5, p. e64782, may 2013. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3665781/
[30] P. Hilton, Fundamental nursing skills. John Wiley & Sons, 2006.
[31] A. Martinez, R. Alcaraz, and J. J. Rieta, "Application of the phasor transform for automatic delineation of single-lead ECG fiducial points." Physiological measurement, vol. 31, no. 11, pp. 1467–1485, nov 2010.
[32] N. Rodriguez-Ibanez, M. A. Garcia-Gonzalez, M. Fernandez-Chimeno, and J. Ramos-Castro, "Drowsiness detection by thoracic effort signal analysis in real driving environments," in Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS. IEEE, aug 2011, pp. 6055–6058. [Online]. Available: http://ieeexplore.ieee.org/document/6091496/
[33] B.-J. Lee, "A Heteroskedasticity Test Robust to Conditional Mean Misspecification," Econometrica, vol. 60, no. 1, pp. 159–171, 1992. [Online]. Available: http://www.jstor.org/stable/2951681
[34] P. Russer, "EMC measurements in the time-domain," in General Assembly and Scientific Symposium, 2011 XXXth URSI. IEEE, 2011, pp. 1–35.
[35] C. Ahlstrom, C. Fors, A. Anund, and D. Hallvig, "Video-based observer rated sleepiness versus self-reported subjective sleepiness in real road driving," European Transport Research Review, vol. 7, no. 4, pp. 1–9, 2015. [Online]. Available: http://dx.doi.org/10.1007/s12544-015-0188-y
[36] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," Information Processing & Management, vol. 45, no. 4, pp. 427–437, 2009.
[37] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data." Biometrics, vol. 33, no. 1, pp. 159–174, mar 1977.
[38] I. H. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000.
[39] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-Validation," in Encyclopedia of Database Systems. Boston, MA: Springer US, 2009, pp. 532–538. [Online]. Available: https://doi.org/10.1007/978-0-387-39940-9_565
[40] M. Borenstein, L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein, "Fixed-Effect Versus Random-Effects Models," in Introduction to Meta-Analysis. John Wiley & Sons, Ltd, 2009, pp. 77–86. [Online]. Available: http://dx.doi.org/10.1002/9780470743386.ch13
[41] G. D. Ruxton, "The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test," Behavioral Ecology, vol. 17, no. 4, pp. 688–690, jul 2006. [Online]. Available: http://beheco.oxfordjournals.org/content/17/4/688.short

FEDERICO GUEDE-FERNÁNDEZ received his Telecommunication Engineering degree and Ph.D. from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, in 2012 and 2018. His research interests include driver monitoring, mobile health and machine learning.

MIREYA FERNÁNDEZ-CHIMENO (M'90) received the Ingeniero de Telecomunicación and Doctor Ingeniero de Telecomunicación degrees from the Universitat Politècnica de Catalunya, Barcelona, Spain, in 1990 and 1996, respectively. She has been a Vice-Dean of the Telecomunication Engineering School (ETSETB) from 1996 to 2000. She is currently an Associate Professor of Electronic Engineering at the Universitat Politècnica de Catalunya, Barcelona, Spain. She is also a Quality Manager of the Electromagnetic Compatibility Group (GCEM), Technical University of Catalonia. GCEM is one of the centers of the Technological Innovation Network of Generalitat de Catalunya (autonomical govern of Catalonia). She teaches courses of electronic instrumentation, acquisition systems, and electrical safety. She is the coauthor of Electronic Circuits and Devices (Edicions UPC, 1999), and Automatic Test Systems (Edicions UPC, 1999) both published in Spanish or Catalan. Her current research interests include biopotential measurements (high-Resolution ECG, beat-to-beat ECG monitoring, and heart rate variability, etc.) and electromagnetic compatibility, mainly oriented to medical devices and hospital environments.

JUAN RAMOS-CASTRO received his Telecommunication Engineering degree and Ph.D. from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, in 1992 and 1997. In 1992, he joined the Electronic Engineering Department as a Lecturer and, since 1997, he has been an Associate Professor, teaching courses in several areas of electronic instrumentation. He is a Member of the Biomedical Research Center at the UPC. His current research interests include biomedical and electronic instrumentation.

MIGUEL A. GARCÍA-GONZÁLEZ received the Ingeniero de Telecomunicación degree, in 1993, and the Doctor Ingeniero Electrónico degree, in 1998, both from the Universitat Politècnica de Catalunya, Barcelona, Spain. He is currently an Assistant Professor of Electronic Engineering, Universitat Politècnica de Catalunya. He teaches courses in several areas of medical and electronic instrumentation. He is engaged in research on instrumentation methods and ECG, arterial blood pressure, and EMG measurements. His current research interests include time series signal processing by time-domain, frequency-domain, time-frequency spectra, and nonlinear dynamic techniques, and noninvasive measurement of physiological signals.