Article

Feature Extraction and Evaluation for Driver Drowsiness Detection Based on Thermoregulation

Jasper Gielen and Jean-Marie Aerts *

Department of Biosystems, Division Animal and Human Health Engineering, M3-BIORES: Measure, Model & Manage of Bioresponses Laboratory, KU Leuven, Kasteelpark Arenberg 30, 3001 Heverlee, Belgium

* Correspondence: jean-marie.aerts@kuleuven.be

Received: 19 July 2019; Accepted: 28 August 2019; Published: 30 August 2019

Abstract: Numerous reports state that drowsiness is one of the major factors affecting driving performance and resulting in traffic accidents. In the past, methods to detect driver drowsiness have been developed based on physiological, behavioral, and vehicular features. In this pilot study, we test the use of a new set of features for detecting driver drowsiness based on physiological changes related to thermoregulation. Nineteen participants successfully performed a driving simulation, while the temperature of the nose (T_{nose}) and wrist (T_{wrist}) as well as the heart rate (HR) were monitored. On average, an initial increase in temperature followed by a gradual decrease was observed in drivers who experienced drowsiness. For non-drowsy drivers, no such trends were observed. In addition, HR decreased on average in both groups, yet the decrease in the drowsy group was more distinct. Next, a classification based on each of these variables resulted in an accuracy of 68.4%, 88.9%, and 70.6% for T_{nose} , T_{wrist} , and HR, respectively. Combining the information of all variables resulted in an accuracy of 89.5%, meaning that ultimately the state of 17 out of 19 drivers was detected correctly. Hence, we conclude that the use of physiological features related to thermoregulation shows potential for future research in this field.

Keywords: driver drowsiness; thermoregulation; distal skin temperature; decision tree

1. Introduction

Driver drowsiness is one of the major factors causing traffic accidents. Dependent on the source, between 10%–20% of all accidents are claimed to be caused by driver drowsiness [1,2]. Besides, the majority of this kind of accidents occur during the night or right after noon [3]. This is not surprising since our physiological clock is programmed to facilitate sleep onset at these moments [4]. Furthermore, drowsiness and sleep are by definition very closely related. For instance, Johns defines drowsiness as “the intermediate state between wakefulness and being asleep” [5]. More so, experiencing drowsiness is used as a synonym for fighting against falling asleep.

In the literature, different methods for monitoring the driver’s state have been presented and are based on features extracted from physiological, behavioral, or vehicular signals [6,7]. Physiological signals are measured directly or indirectly on the driver self. The relation between signals such as an electroencephalogram (EEG), electrocardiogram (ECG), electrooculogram (EoG), electromyogram (EMG), and sleep have been described in many studies [8]. Moreover, in polysomnography (PSG), these electrophysiological signals are interpreted according to the guidelines of the American Academy of Sleep Medicine (AASM) [9] for a gold standard reference measure of sleep and sleep quality [8]. Especially brain activity (EEG) shows the most distinct changes according to the state of the driver. Although, EEG technology has significantly evolved over the past years, setting up the devices still requires a lot of know-how and caution to obtain accurate measurements [10]. In addition, artifacts due to movement and other disturbing signals completely deteriorate the quality of the signal in less

controlled setups. Alternatively, the driver's behavior is also monitored for drowsiness detection. Features such as eye closure, blinking frequency, yawning, and head-nodding can be extracted by applying camera vision and face recognition techniques (e.g., [11,12]). The advantage of using cameras is that the measurements are completely unobtrusive. When the camera is placed near the sun visor or the rear mirror, the driver's field of view is not obstructed at all. However, issues with lighting conditions have been associated with these techniques [7]. Thirdly, features extracted from the vehicle itself are also widely used to monitor driver drowsiness. The angle of the steering wheel, the pressure applied to the acceleration and brake pedal, distance to the vehicle in front and a lot of other vehicular features have been used in recent studies [13–15]. Currently, commercial applications built into cars, busses, and trucks are already available on the market. These so-called advanced driver assistance systems (ADAS) are designed to automate, adapt, and enhance vehicles for safer and better driving.

Along with these ADAS, there is an ongoing evolution towards autonomous driving. Apart from the technical aspects of self-driving vehicles, the component of human attention and control should not be neglected. For instance, commercial and transport airplanes fly autonomously for the majority of the time, yet the pilots still have to monitor the plane's status continuously and must be ready for intervention to assure a smooth and safe flight. Similarly, low levels of autonomous driving also requires an attentive driver [16]. Moreover, legal implications related to fully autonomous driving are expected to influence integration on the market more than the technical challenges [17].

In this paper, we focus on a new set of features to monitor driver drowsiness based on physiological changes related to thermoregulation. The functional link between temperature regulation and sleep onset has been studied for a long time. In regulating temperature, the body is typically considered in terms of a core and an outer shell [18]. The core body temperature (T_C) is maintained within a specific narrow range (around 37 °C), whereas the outer shell can vary over multiple degrees. To illustrate, in thermo-neutral conditions, distal skin temperature is reported to be 7–8 °C below the core [19]. However, at a room temperature of 35 °C, the hands and feet are only 3–4 °C cooler. Furthermore, it is a well-known phenomenon that T_C varies periodically in a circadian way. Throughout the day, the temperature peaks in the evening around 21:00, followed by a gradual decrease overnight with a minimum around 05:00 [20]. To facilitate these temperature changes, heat production and dissipation are closely regulated. In thermo-neutral, sedentary conditions, this heat balance is mainly regulated by autonomous control of vasodilation and vasoconstriction of the arterioles in the skin. Accordingly, controlling blood flow to the skin is crucial for managing the human heat balance. Moreover, Rowel [21] states that skin blood flow takes up between 5% and 60% of our cardiac output depending on the environmental conditions. With respect to sleep and sleep onset, Gilbert et al. [19] presented the hypothesis that heat loss via the extremities is linked in feedback with activation of the sleep-promoting areas in the brain. More specifically, efferent warm sensitive neurons (WSN) in the preoptic anterior hypothalamus (PoAH) innervate and stimulate somnogenic brain structures, while other thermo-sensitive neurons innervating wake-promoting brain areas are inhibited. Quanten et al. [22] experimentally tested this hypothesis and showed that in conditions of unwanted sleepiness in active subjects there is a negative feedback connection between the distal-to-proximal temperature gradient and sleep onset.

In this work, we aim to investigate the use of features extracted from peripheral skin temperatures of the nose and wrist (heat dissipation) as well as the heart rate (heat production) to distinguish between subjects experiencing drowsiness while driving and subjects that remain alert.

2. Materials and Methods

2.1. Study Design and Participants

The data of 19 healthy subjects between 20 years old and 27 years old who volunteered to perform a driving test were used in this study. Each participant was asked to drive in a driving simulator while physiological variables related to thermoregulation were monitored. To simulate a monotonous driving

experience, a PlayStation 3 and racing game (Gran Turismo 5) were used together with a compatible steering wheel as well as acceleration and brake pedals (Logitech Driving Force EX). The race track named “Special Stage 7” was selected because it represents an empty highway at night. Additionally, a speed limit of 80 km/h was imposed. This setup was installed in a climate-controlled room of 2.3 by 3.6 m, where the temperature was held constant at 23 °C and the relative humidity kept at 50%. The constant thermal environment was deemed realistic since similar conditions have been reported in experimental measurements performed on the road [23].

Every participant arrived at the test facility at around 21:00. After an introduction to the setup and the purpose of this study, the volunteer was equipped with the different sensors (see Section 2.2. Data Collection). Subsequently, the driving simulation started between 21:30 and 22:15. The subjects performed the driving test in either one of the following conditions. In a first series of experiments, the lights inside the climate-controlled room were dimmed and there was no communication between the subject and the researcher present to follow up the measurements and the driver’s state. The driving simulation was stopped after a maximum duration of 2.5 h or until the driver was too sleepy to continue driving in a normal way. In a second set of experiments, the lights inside the room were left on and communication between the researcher and the participant was allowed. It has been shown that exposure to light elicits acute physiological effects in humans such as an increase in alertness, suppression of melatonin, and an increase in T_C [24]. In line with the hypothesis by Gilbert et al. [19], the thermoregulatory changes are thought to be related to the alerting effects instead of a direct relation between light exposure and thermoregulation. Evidently, the purpose of this second protocol was to increase the chances of staying alert during the simulation. However, the participants were not forced to stay alert. For these measurements, a maximum duration of 1.5 h was set in advance. The maximum duration was set to 1.5 h because this was the average in the first set of experiments. In other words, the first group of participants became on average too sleepy to drive after 1.5 h under the first conditions. Figure 1 visualizes the timing of the experiments and shows the driving simulator that was set up in the climate-controlled room.

Figure 1. Timing of the driving experiment (a) and a photo of the driving simulator installed in the climate-controlled room (b).

In total, 26 people completed the driving simulation. However, the data of 7 participants were excluded due to missing sensor data (i.e., in 3 simulation data collection were not stored correctly) and/or limited duration of the driving simulation (i.e., 5 participants were falling asleep within 40 min to 60 min after the start). In case a participant became too drowsy to drive within 1 hour, it was assumed that the process of sleep onset started before entering the simulator. As a result, the analysis was applied to a dataset of 10 participants who performed the first protocol and 9 subjects who performed the second version. Afterwards, we checked whether this first group actually experienced drowsiness during the test and the second group did not (see Section 2.2. Data Collection).

The study was approved by the Social and Societal Ethics Committee (SMEC) of the KU Leuven (case number G-2019 04 1632) and all participants provided written informed consent before starting the driving test.

2.2. Data Collection

During the driving simulation, a number of different physiological variables were monitored. First, information about the state of the driver was obtained by means of a self-assessment of the mental state. The Stanford Sleepiness Scale (Stanford Sleepiness Scale: <https://web.stanford.edu/~dement/sss.html>) (SSS) was used for this purpose [25]. The SSS is a validated method for determining sleepiness and has been used in other studies on driver drowsiness [26,27]. The scale consists of 7 scores/levels to indicate the degree of sleepiness with '1' being the most alert and '7' being almost asleep. The driver was asked to score his/her mental state every 5 min based on this scale. The timing of these moments was indicated by the controlling researcher present in the room by a small hand gesture (i.e., slowly raising their right hand). This way, distraction from the driving task was minimal and the driver did not experience any abrupt changes in his/her mental state. When the highest score of 7 was given 2 times in a row, the experiment was terminated prematurely because we assumed the driver to be too drowsy to continue driving in a normal way. In the SSS, a score of 7 corresponds to the following state: "No longer fighting sleep, sleep onset soon; having dream-like thoughts". Secondly, the skin temperature of the tip of the nose (T_{nose}) was measured using a VarioCAM infrared thermal camera with a resolution of 640 by 480 pixels (InfraTec GmbH, Dresden, Germany). The camera was pointed towards the face and from the thermal images, the position of the nose was manually determined. The frame rate of the camera was set to 1 frame per minute. Thirdly, the skin temperature at the wrist (T_{wrist}) was measured continuously with an Empatica E4 wristband at a sampling rate of 4 Hz (Empatica Inc., Cambridge, MA, USA). Fourth, the heart rate (HR) was monitored using a Zephyr HxM chest strap (MedTronic, Annapolis, MD, USA). The HR, expressed in beats per minute (bpm), was sampled every second with a smartphone and app for logging the data (Samsung Galaxy J1). In this research, we did not consider raw ECG measurements because it was decided beforehand that only the HR would be used in the analyses (see Section 2.4. Data Analysis). Lastly, the ratings of the SSS were used to check whether participants who performed the first protocol actually became drowsy during the driving test and the second group of subjects remained alert. This decision was based on the sleepiness scores. If a score of 5 was given at any point during the test, the participant was labeled as "drowsy". If this was not the case, the subject was labeled as "non-drowsy".

During one simulation, the data of T_{wrist} were not stored correctly. Hence, only 18 measurements are available. In addition, the data of the heart rate were missing for 2 simulations.

2.3. Data Pre-Processing

Since there was a difference in the length of the driving simulations and sampling frequency of the different variables, the first step in pre-processing was resampling the data to a fixed number of observations for each participant. To do so, all measurements (T_{nose} , T_{wrist} , HR, and SSS) were subdivided in a fixed number of 9 windows. This specific number was chosen based on the average duration of the driving test (90 min) and the average number of SSS ratings (18 per simulation). Below, Figure 2 visualizes this first pre-processing step schematically.

Afterwards, the mean value for the data inside each window was calculated. Subsequently, the data for each person were scaled by subtracting the average temperature or heart rate. This step was performed to account for inter-individual variability in the baseline of the variables. For instance, the heart rate of one participant fluctuated around 50 bpm, whereas this was 70 bpm for someone else. Furthermore, the slope of the data inside each window was also determined by means of least squares linear regression. These slopes will be used as the main features to distinguish between drowsy and non-drowsy drivers. A positive value corresponds to an increase in the variable in a certain window. Conversely, a negative value indicates a decrease. All operations described above were executed using MATLAB® software (The Math works Inc., Natick, MA, USA).

Figure 2. Visualization of pre-processing the data by subdividing each measurement into 9 windows. The black and grey dotted lines represent two fictional measurements for a certain variable x . Although the sampling frequency is the same for both measurements, the number of data points (i.e., length of the measurement) is not fixed. Accordingly, the number of data points inside the windows varies as well between the measurements.

2.4. Data Analysis and Classification

The first step in analyzing the data was looking for significant changes of T_{nose} , T_{wrist} , and HR for drowsy and non-drowsy subjects. The average trend over time was calculated for each variable as well as the distribution for each window. In addition, a Welch's ANOVA test was performed to check whether the data of the 9 windows were significantly different in their mean. The significance level (α) was set to 5%, meaning that a p-value smaller than 0.05 indicated a significant difference. The Welch's test was chosen because it does not assume an equal variance in both distributions.

Secondly, in our attempt to distinguish between drowsy and non-drowsy subjects, we applied a decision tree approach based on changes in skin temperature and heart rate. This methodology aims to start from biological knowledge for building our classifier, which is in contrast to traditional data-based approaches. The latter type of methods focused mainly on accuracy or performance but did not lead to physiological insights per se. In essence, this approach was inspired by the methodology described in patent EP 2 842 490 B1 published in 2016 [28] as well as the observed trends in the data. More specifically for the skin temperatures, we expected an initial increase to facilitate heat loss for lowering the core body temperature (T_C). Afterwards, a gradual decrease in skin temperature was anticipated. On the other hand, a decrease in the heart rate was expected because energy and heat production in the body was supposed to go down. Based on this background, 3 univariate classifiers were built, which made their decisions based on the increases and decreases in the data. Hence, the classifications were performed based on T_{nose} , T_{wrist} , and HR. For the classification based on one of the temperature variables, a decision tree with 2 nodes was designed. The first nodes check whether an initial increase in the temperature occurred by evaluating the slope in the selected initial window (number 1, 2 or 3). Subsequently, the second node evaluated whether there was a decrease in temperature by looking at the slope in windows 4 to 9. Accordingly, the thresholds for both decision nodes were set to zero. Positive values for the slope corresponded to an increase in temperature and negative values to a decrease. On the other hand, for classification based on the heart rate, a decision tree with only one node assessed if HR was significantly decreasing over time. Figure 3 shows a schematic overview of both structures for the decision tree. For the decision tree based on the temperature variables, every combination of each 2 windows was tested. For the decision tree based on HR, all windows were tested separately. Information extracted from specific windows was used for classification to indicate the position of the most valuable information present in the data. The choice for a (combination of) window(s) was based on the performance of the corresponding decision tree. For this purpose, the sensitivity, specificity, and accuracy for each combination were determined. Here a "positive" test referred to a participant being labeled as "drowsy" by the classification algorithm. The windows that were used in the decision tree with the highest accuracy were selected in the end. The receiver operating characteristic (ROC) curves and corresponding area under the curve (AUC) values were calculated as well by changing the threshold from zero. The advantage of this method is that the results allow for a meaningful interpretation since the approach is based on physiological knowledge.

Figure 3. Schematic overview of the structure for the decision tree classification. (a) $dT_{win I}$ refers to the slope of the temperature in one of the initial windows of the measurement (number 1, 2, or 3). Accordingly, $dT_{win II}$ refers to the slope in either one of windows 4 to 9. As a result, the driver is labeled as “drowsy” when an increase in temperature is followed by a decrease. (b) $dHR_{win I}$ indicates the slope of the heart rate inside one of the nine windows.

In the final part of the analysis, classification was performed based on all 3 variables. For this, no predetermined structure was imposed. The only constraints were that the features (the sign of the slope inside a certain window) from the univariate classifications were used. Similarly, the sensitivity, specificity, and accuracy were used to evaluate the performance. However, it was not possible to determine ROC curves for this multivariate approach.

3. Results

3.1. Data Trends

As described above, the average trend in the distal skin temperatures, heart rate, and sleepiness score was calculated for both the group of drowsy and non-drowsy subjects. These trends and the corresponding 95% confidence interval are visualized in Figure 4. We note that on average T_{nose} and T_{wrist} increased over the first three windows (+0.55 and +0.91 °C, respectively), followed by a decrease (−1.05 °C and −2.01 °C, respectively) in the drowsy subjects. The skin temperatures of the non-drowsy subjects showed no such pattern. When looking at the heart rate, we observed that there was a decreasing trend for both the drowsy and the non-drowsy subjects, yet the trend appeared to be more distinct for drowsy drivers with an average decrease of 10 bpm compared to 3 bpm for the group of non-drowsy drivers. The sleepiness scores showed a linear increase from on average 2.5 to 6.7 and 2.3 to 3.1 for the drowsy and non-drowsy group, respectively. Furthermore, the results of the Welch’s ANOVA are presented in Table 1. From these tests, we note that for drowsy subjects, the means of the windows differed significantly from each other in each of the four variables ($p < 0.001$). For the group of non-drowsy subjects, there was a significant difference in the mean of the heart rate between all nine windows, which indicated that the observed decreasing trend was significant ($p = 0.008$). For the other three variables (T_{nose} , T_{wrist} , and SSS), the hypothesis that the means of all windows were equal was accepted.

Figure 4. Average trend of T_{nose} , T_{wrist} , heart rate (HR), and Stanford sleepiness scale (SSS) in drowsy and non-drowsy subjects. (a) T_{nose} in drowsy subjects, (b) T_{nose} in non-drowsy subjects, (c) T_{wrist} in drowsy subjects, (d) T_{wrist} in non-drowsy subjects, (e) HR in drowsy subjects, (f) HR in non-drowsy subjects, (g) SSS in drowsy subjects, (h) SSS in non-drowsy subjects.

Table 1. Statistics of the Welch’s ANOVA tests. The “degrees of freedom” for each test are indicated by dF. F indicates the corresponding F-statistic. “D” refers to the group of drowsy drivers and “ND” to the non-drowsy drivers.

Welch’s ANOVA	T_{nose}		T_{wrist}		HR		SSS	
	D	ND	D	ND	D	ND	D	ND
dF	9	8	8	8	8	7	9	8
F	11.8	0.5	7.9	0.9	15.7	3.5	54.2	1.9
p-value	<0.001	0.820	<0.001	0.493	<0.001	0.008	<0.001	0.097

3.2. Decision Tree Classification

Based on these observed trends, a decision tree classification was tested to distinguish between drowsy and non-drowsy participants. Afterwards, information about both the distal temperatures and the heart rate was used in a multivariate classifier.

3.2.1. Univariate Classification

The performance measures, as well as the selected window numbers for the three univariate decision trees, are presented in Table 2, and the corresponding ROC curves and AUC-values are shown in Figure 5. Firstly, a combination of the slope inside window 1 and 6 performs the best for the decision tree based on T_{nose} . In total, the condition of 13 out of 19 drivers was classified correctly.

Additionally, in case the classification was solely based on the information from window 6, this resulted in a sensitivity, specificity, and accuracy of 90.0%, 66.7%, and 79.0%, respectively. Secondly, when analyzing the temperature measured at the wrist, the combination of windows 2 and 4 came out on top. As mentioned, the data of 18 drivers were available. All participants that did not experience drowsiness were classified correctly, as well as 7 out of 9 “drowsy” drivers. For T_{wrist} we did not observe an increase in the performance when the data of only one window were used. Third, the decision tree based on the information of the heart rate inside window 6 resulted in an overall accuracy of 70.6%. Here, these results are based on the data of 17 driving simulations.

Table 2. Performance of the different classifications based on information about T_{nose} , T_{wrist} , and HR. The “strict” multivariate classification refers to the fact that for all three variables, the imposed constraints must hold for a driver to be classified as drowsy. The “mild” classification represents the case when a driver is labeled as drowsy when only two out of three constraints are met. The sensitivity, specificity, and accuracy are expressed as a percentage. N_{drowsy} and $N_{non-drowsy}$ indicate the number of samples used in each case.

Classification	Windows	Sens. (%)	Spec. (%)	Acc. (%)	N_{drowsy}	$N_{non-drowsy}$
Univariate: T_{nose}	1 & 6	60.0	77.8	68.4	10	9
Univariate: T_{wrist}	2 & 4	77.8	100.0	88.9	9	9
Univariate: HR	6	88.9	50.0	70.6	9	8
Multivariate: “strict”	6, 4 & 6 *	50.0	100.0	75.0	8	8
Multivariate: “mild”	6, 4 & 6 *	100.0	77.8	89.5	10	9

* Window number for T_{nose} , T_{wrist} , HR respectively.

Figure 5. Receiver operating characteristic (ROC) curve and corresponding area under the curve (AUC) for the univariate classification based on T_{nose} , T_{wrist} , and HR.

3.2.2. Multivariate Classification

In the second classification approach, insights from the univariate classifications are combined to improve accuracy and robustness. Accordingly, the best performance was obtained by classifying the data based on the slope inside window 6 for T_{nose} , window 4 for T_{wrist} , and window 6 for HR. In a first, very strict classification approach, we imposed that for a certain simulation, the slope inside each of these windows had to be negative to classify the driver as drowsy. Accordingly, the order of the decisions was not of importance to this approach. In a second, milder classification approach, we assigned the label “drowsy” when at least two out of these three constraints were met. The performance of both classification types was also presented in Table 2 above. We note that the number of observations

was not the same for both analyses. As mentioned in Section 2.2., the data of three simulations were missing information about T_{wrist} or HR. Therefore, the first classification considered 16 out of 19 simulations. However, simulations with missing data can be considered in the second classification if the two available variables had the same sign of the slope inside the respective windows. Since this was the case for all three incomplete datasets, all 19 simulations were considered. Accordingly, we note that applying the strict constraints resulted in a specificity of 100.0% but also a relatively low sensitivity of 50.0%. Furthermore, an accuracy of 89.5% was observed under the mild constraints.

4. Discussion

4.1. Data Trends

From the first results presented in this study, it is clear that there are significant trends in the physiological variables of participants who became drowsy during the driving simulation. Moreover, these trends were not observed in participants that did not experience drowsiness. As expected, we observe an increase in both temperature variables in the first period of the driving simulation. This is in accordance with the fact that prior to sleep onset, the body dissipates heat via the extremities to lower the temperature of the core [29]. During this process, blood flow to the extremities is increased (vasodilation), which in turn increases the skin temperature. The result is an increased thermal gradient with the surrounding air to facilitate heat transfer to the environment. The subsequent decrease in skin temperature is also in line with the expectations. To avoid losing too much body heat, blood flow to the skin decreases again (vasoconstriction). Besides, the heart rate of drowsy drivers decreased on average more than that of non-drowsy drivers. We hypothesize that the decreasing trend rate can be interpreted as a secondary mechanism to control T_C , which is in accordance to the statements by Kräuchi and Wirz-Justice [20]. The human energy balance namely consists of energy producing and energy consuming (or reducing) components. Hence, the heart rate can be seen as a measure for providing energy to the body. Accordingly, lowering the heart rate means lowering the input to the energy balance. Nevertheless, we keep in mind that other physiological processes (e.g., stress) affect the heart rate and its dynamics as well. With respect to the SSS, we can infer that the test subjects evolved to a high sleepiness score in the first protocol, whereas this score did not increase significantly in the second protocol and fluctuated on average between two and three.

4.2. Decision Tree Classification

The first univariate classification based on information about the skin temperature of distal body parts (nose and wrist) shows promising results. The rather straightforward, predetermined structure of the decision trees, results in an accuracy between 68.4% and 88.9%. Moreover, AUC-values between 0.833 and 0.975 were observed. For both temperature variables, the specificity is higher than the sensitivity. This can be explained because two constraints have to be met for labeling the driver as “drowsy”, whereas he/she can be labelled as “non-drowsy” after only one decision in the decision tree (see Figure 3). Moreover, we note that the performance for the classification based on T_{wrist} are better than the performance of T_{nose} . When analyzing the raw data of both variables, we observed smaller changes in T_{wrist} from sample to sample. Obviously, this was related to the much higher sampling frequency of T_{wrist} (see Section 2.2. Data Collection). On the other hand, it has been shown that the temperature of the nose is also more variable. For example, it has been shown that the nose temperature changes periodically due to respiration [30]. Alternatively, an increased sensitivity and accuracy is observed when we only consider the information of window 6 for T_{nose} . From this, we conclude that the decrease in temperature is a more prominent feature for distinguishing between drowsy and non-drowsy drivers. Nevertheless, only considering this feature leads to more incorrect “drowsy” classifications, which is not surprising since the physiology behind temperature changes in relation to sleep onset suggests that both an increase and decrease occur. When considering the heart rate, a negative slope in window 6 is observed for 8 out of 9 drivers who reached the SSS threshold and were

labelled as “drowsy”. In addition, such a decrease was also observed in half of the “non-drowsy” participants. Given the simplicity of the decision criterion, an overall accuracy of 70.6% stresses the significance of the decreasing trend in HR in our test population.

In the last section of the results, two classifications based on T_{nose} , T_{wrist} and HR are presented. When the constraints for all three variables have to be met in order to label a driver as “drowsy”, only four out of eight drivers who actually experienced drowsiness were labeled correctly. In contrast, no drivers were falsely labelled as being drowsy. Accordingly, the use of such stringent constraints can be applied to situations where false positive classifications are unacceptable. In other words, this methodology can be used when it is important to be certain that a driver labelled as drowsy, is actually drowsy. Nevertheless, a sensitivity of only 50.0% is not sufficient to distinguish between drowsy and non-drowsy drivers consistently. Evidently, using less stringent constraints results in a larger number of “drowsy” labels. Hence, all 10 drowsy drivers were labeled correctly. While the specificity is now decreased, we note that still 7 out of 9 non-drowsy drivers are correctly identified. This last classification resulted in an accuracy of 89.5%.

When comparing the different classification approaches, we note that the univariate classification for T_{wrist} and the second multivariate classification resulted in similar accuracies (88.9% and 89.5%). Moreover, both results cannot be compared directly because the number of samples differs. The fact that a single temperature variable performs equally well as the multivariate classification indicates that including the heart rate for classifying the data does not add significant value to the current analysis.

4.3. Comparison to State-of-the-Art Methods for Drowsiness Detection

When comparing the performance of these classifications to related work on detecting driver drowsiness, we first note that, to our knowledge, no other studies make use of distal skin temperatures in their classification. Furthermore, only the patent by Berckmans et al. [28] presents a method based on similar signals, namely distal ear temperature and heart rate. Here, different conceptual approaches are described as well as the performance of one such method for a training and validation set; respective accuracies of 87.5% and 72.7% were obtained. On the other hand, multiple studies present methodologies to detect or monitor drowsiness based on information about the heart rate. More specifically, features related to heart rate variability (HRV) are most often used. For instance, Li and Chung [31] present a detection algorithm with an accuracy, sensitivity, and specificity of 95% each. In this study, two-minute-long drowsy and alert events were classified based on the ratio of low (0.04–0.15 Hz) to high frequency (0.15–0.40 Hz) variability in the heart rate obtained from photoplethysmography. In another study, ECG measurements were used to detect drowsiness based on HRV as well as respiratory frequency [32]. Here, a positive predictive value (PPV), sensitivity, and specificity of 96%, 59%, and 98%, respectively, was reported. In addition, an accuracy of around 78% corresponds to these results. In contrast to the two prior studies which used behavioral signs as a reference for the state of the driver, Fujiwara et al. validated their HRV-based drowsiness detection with EEG measurements [33]. Here, the algorithm was successful if drowsiness was indicated in a period of 15 min before sleep onset (NREM sleep stage 1). In 12 out of 13 subjects, sleep onset was detected in this timeframe. In comparison to these studies, our classification based on only the heart rate performs less well (an accuracy of 70.6%). However, HRV is interpreted as a measure for the dynamic control of the heart by different parts of the nervous system. In this study, the heart rate was interpreted as a part of the heat balance inside the human body. More specifically, HR represents the rate of aerobic heat production inside the body. Therefore, we also consider HR in combination with other components of the heat balance. Nonetheless, using ECG measurements in future research could make it possible to combine or compare these different approaches.

To put our own results further into perspective to the state-of-the-art, a couple of recent studies that used completely different variables and features are discussed hereafter. Firstly, Li et al. [34] present a smartwatch-based wearable EEG system for driver drowsiness detection. They designed a headband with three dry electrodes, instead of 21 electrodes used in the standard 10–20 system. In their analysis,

they distinguished between “alert” and “drowsy” epochs that were labelled based on the percentage of eye closure (PERCLOS) and the number of adjustments on the steering wheel. The authors reported an average accuracy of 91.25% and 91.92% for detecting the alert and drowsy epochs, respectively. Secondly, Mandal et al. [12] developed a method to determine PERCLOS from images of a dome camera in a bus. Their results showed a matching rate between their method and a ground truth measure for PERCLOS of 85.02% and 95.18% for normal and fatigued drivers, respectively. Since their method was based on commercial technology that is already implemented in busses, monitoring bus drivers can be done unobtrusively. Thirdly, the work by Li et al. [13] presents yet a completely different approach. The fatigue level, scored by experts, was linked with data about the steering wheel angles (SWA) recorded under real driving conditions. The system was able to run on-line and performed with an accuracy of 78.01% in distinguishing drowsy from non-drowsy participants. In this research, we obtained accuracies between 70% and 90%, which is below the performances of the related state-of-the-art. Nevertheless, the results show potential for future research.

Lastly, it should be noted that each study applies different protocols and builds their classifier on different events or epochs. This makes it even more difficult to compare all results directly.

4.4. Current Limitations and Future Perspectives

Given the preliminary nature of this pilot study, several limitations and suggestions for future work are listed in this section. Firstly, more advanced methodologies and equipment will improve the quality of the obtained measurements. For instance, now a simple driving simulator is used as a basis for the experiments. Nowadays, there are simulators available with a driving experience very close to driving a real car. Nevertheless, when working with a simulator, lacking the danger of causing an accident is thought to affect the driver in his behavior [35]. On the other hand, purposely sending out drowsy drivers on the road would be unethical and illegal. Using wearable technology for ambulatory monitoring of the driver in real-life conditions might be a good alternative for this issue. Furthermore, applying a crossover study design in which each participant performs the driving test in drowsy and non-drowsy conditions would allow for a pairwise comparison of the individual direction and magnitude of the changes in thermoregulation. This is not possible with the current data since each driver only performed the test one time. Another way to improve the experimental work is by working with more advanced measures for determining the state of the driver. Now, each driver evaluates his/her state of alertness based on the SSS. Presumably, individual differences in the interpretation of each level of the SSS causes variability in the data. The use of EEG measurements (brain activity) could provide a solution to this issue. As mentioned, EEG is used as a gold standard in sleep research. Alternatively, scoring the driver’s state based on an expert’s opinion is also done in the literature. An expert can be trained in scoring the driver based on behavioral expressions. Although such a score is also the result of human interpretation, the variability due to a difference in the scorer is eliminated. A final remark on the applied methodology is the timing of the simulations. Currently, one fixed moment throughout the day is selected to perform the driving simulation. By executing the measurements at different moments, a generalization of the results is possible. For instance, Fujiwara et al. [33] used data collected at 11:00 and after lunch. However, we have to keep in mind that the theory behind this method is related to actual sleep onset and does not relate directly to short drops in attention throughout the day. Accordingly, future research on the proposed methodology will focus on this specific type of event. Related to the timing of these events, we expect that the observed physiological changes do not only occur at night. For instance, night shift workers have an altered sleep-wake cycle. In general, diurnal changes in body temperature and heart rate follow the daily activity-rest cycle [36].

Secondly, we discuss a number of limitations related to data analysis. At the moment, we applied a very straightforward method to determine significant differences in drowsy and non-drowsy drivers. Subdividing the measurements in a fixed amount of time windows was an inevitable part of the analysis. As mentioned, this is due to a difference in the length of the experiments. By applying the windows to

the data and extracting the average value and the slope, a lot of detail, as well as information about the timing is lost. Nevertheless, this did not prevent us from identifying significant trends in the data and successfully distinguishing between drowsy and non-drowsy drivers. A follow-up study could either perform experiments of the same length or apply a time series approach to detect drowsiness over time and pinpoint key moments in the transition from an alert state to a drowsy one. Additionally, to allow for such a methodology to be applied in real-time and real-world situations, future research has to focus on using techniques such as recursive estimation and online time-series analysis to determine the increases and decreases in the variables in real-time. Overall, the current methodology for distinguishing between drowsy and non-drowsy drivers is elegant in its simplicity, allowing for a physiological interpretation of the results and indicates the potential of using thermoregulatory features for drowsiness detection. A final remark on the analysis is related to the lack of data for validation. Due to the limited number of measurements, no proper validation technique could be applied to the current dataset. Leave-one-out cross validation was briefly considered, however, this resulted in the same results for the training and test set. More specifically, only the selected window number could vary when using the decision tree classification. The selection for a (combination of) certain window(s), did not vary when only leaving out one observation at a time. Naturally, the lack of validation is identified as one of the most important limitations of the current communication.

5. Conclusions

In this work, the use of features extracted from distal skin temperatures and heart rate is tested for detecting drowsiness in driving simulations. In the first part of the analysis, we demonstrated that T_{nose} , T_{wrist} , and HR vary throughout the measurement in a specific pattern in participants who became drowsy. Initially, the temperature measured at the nose and the wrist increased to a maximal value. Subsequently, a gradual decrease in these temperature variables was observed. When studying the heart rate of the driver throughout each simulation, a significant decrease was observed in both drowsy and non-drowsy participants. However, the decreasing trend was more distinct in the group of drowsy drivers. Secondly, we showed that both populations of drivers (drowsy and non-drowsy) could be classified based on these observed trends in the data of T_{nose} , T_{wrist} , and HR. Despite the simplicity of these classifications, their performance indicates the potential for future research. The main advantage of the applied methodology is that it is based on knowledge about physiological processes related to sleep onset. Specifically, heat loss via the extremities and controlling of the heart rate regulates the decrease in core body temperature before and during sleep. A secondary classification approach was tested by using the information of all three different variables at the same time. From this analysis, we conclude that including the heart rate in our classification approach does not improve the performance significantly. Lastly, the current results were compared to the state-of-the-art in drowsiness monitoring and several limitations and suggestions for future research have been discussed.

Author Contributions: Conceptualization, J.G. and J.-M.A.; methodology, J.G.; formal analysis, J.G.; writing—original draft preparation, J.G. and J.-M.A.; writing—review and editing, J.G. and J.-M.A.; visualization, J.G.; supervision, J.-M.A.

Funding: This research received no external funding.

Acknowledgments: We would like to thank Steve De Raedt, Michiel De Vis, Arthur Devlieger, Robbert Jacobs, Jasmine Lalanto, Avitaj Mitra, Rahmath B. Paravakkal, Maarten Schoolmeesters, Laura Smets, Laura Verhelst, Wouter Verschuere, and Noor Vidts for their contribution to the experimental set-up and measurements, which was part of their students' project at the faculty of bio-science engineering at the KU Leuven.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tefft, B.C. *Prevalence of Motor Vehicle Crashes Involving Drowsy Drivers, United States, 2009–2013*; AAA Foundation for Traffic Safety: Washington, DC, USA, 2014.
2. Diependaele, K. *Sleepy at the Wheel. Analysis of the Extent and Characteristics of Sleepiness Among Belgian Car Drivers*; BRSI-KCRI: Brussels, Belgium, 2015.
3. Higgins, S.J.; Michael, J.; Austin, R.; Akerstedt, T.; Van Dongen, H.P.; Watson, N.; Czeisler, C.; Pack, A.I.; Rosekind, M.R. *Asleep at the Wheel: A National Compendium of Efforts to Eliminate Drowsy Driving*; NHTSA: Washington, DC, USA, 2017.
4. Waterhouse, J.; Fukuda, Y.; Morita, T. Daily rhythms of the sleep-wake cycle. *J. Physiol. Anthropol.* **2012**, *31*, 5. [[CrossRef](#)] [[PubMed](#)]
5. Johns, M.W. A sleep physiologist's view of the drowsy driver. *Transp. Res. Part F Traffic Psychol. Behav.* **2000**, *3*, 241–249. [[CrossRef](#)]
6. Sahayadhas, A.; Sundaraj, K.; Murugappan, M. Detecting driver drowsiness based on sensors: A review. *Sensors* **2012**, *12*, 16937–16953. [[CrossRef](#)] [[PubMed](#)]
7. Sikander, G.; Anwar, S. Driver Fatigue Detection Systems: A Review. *IEEE Trans. Intell. Transp. Syst.* **2018**, *20*, 2339–2352. [[CrossRef](#)]
8. Van De Water, A.T.M.; Holmes, A.; Hurley, D.A. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography—A systematic review. *J. Sleep Res.* **2011**, *20*, 183–200. [[CrossRef](#)] [[PubMed](#)]
9. Iber, C. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*; American Academy of Sleep Medicine: Darien, IL, USA, 2007.
10. Grummett, T.S.; Leibbrandt, R.E.; Lewis, T.W.; DeLosAngeles, D.; Powers, D.M.W.; Willoughby, J.O.; Pope, K.J.; Fitzgibbon, S.P. Measurement of neural signals from inexpensive, wireless and dry EEG systems. *Physiol. Meas.* **2015**, *36*, 1469–1484. [[CrossRef](#)] [[PubMed](#)]
11. Choi, I.H.; Jeong, C.H.; Kim, Y.G. Tracking a driver's face against extreme head poses and inference of drowsiness using a hidden Markov model. *Appl. Sci.* **2016**, *6*, 137. [[CrossRef](#)]
12. Mandal, B.; Li, L.; Wang, G.S.; Lin, J. Towards Detection of Bus Driver Fatigue Based on Robust Visual Analysis of Eye State. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 545–557. [[CrossRef](#)]
13. Li, Z.; Li, S.E.; Li, R.; Cheng, B.; Shi, J. Online detection of driver fatigue using steering wheel angles for real driving conditions. *Sensors* **2017**, *17*, 495. [[CrossRef](#)] [[PubMed](#)]
14. Wakita, T.; Ozawa, K.; Miyajima, C.; Igarashi, K.; Itou, K.; Takeda, K.; Itakura, F. Driver identification using driving behavior signals. In Proceedings of the 8th International IEEE Conference on Intelligent Transportation Systems, Vienna, Austria, 13–16 September 2005.
15. Dehzangi, O.; Masilamani, S. Unobtrusive Driver Drowsiness Prediction Using Driving Behavior from Vehicular Sensors. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018.
16. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. Available online: https://www.sae.org/standards/content/j3016_201806/ (accessed on 15 July 2019).
17. Joan, C.; Shaun, K. Autonomous vehicles: No driver ... no regulation? *Nature* **2018**, *361*, 2016–2018.
18. Aschoff, J.; Wever, R. Kern und Schale im Wärmehaushalt des Menschen. *Naturwissenschaften* **1958**, *45*, 477–485. [[CrossRef](#)]
19. Gilbert, S.S.; Van Den Heuvel, C.J.; Ferguson, S.A.; Dawson, D. Thermoregulation as a sleep signalling system. *Sleep Med. Rev.* **2004**, *8*, 81–93. [[CrossRef](#)]
20. Krauchi, K.; Wirz-Justice, A. Circadian rhythm of heat production, heart rate, and skin and core temperature under unmasking conditions in men. *Am. J. Physiol. Integr. Comp. Physiol.* **1994**, *38*, 819–829. [[CrossRef](#)] [[PubMed](#)]
21. Rowell, L.B. Human cardiovascular adjustments to exercise and thermal stress. *Physiol. Rev.* **1974**, *54*, 75–159. [[CrossRef](#)] [[PubMed](#)]
22. Quanten, S.; De Valck, E.; Cluydts, R.; Aerts, J.M.; Berckmans, D. Individualized and time-variant model for the functional link between thermoregulation and sleep onset. *J. Sleep Res.* **2006**, *15*, 183–198. [[CrossRef](#)] [[PubMed](#)]

23. Fiser, J.; Pokorný, J.; Podola, D.; Jicha, M. Experimental Investigation of Car Cabin Environment During Real Traffic Conditions. *Eng. Mech.* **2013**, *20*, 229–236.
24. Cajochen, C.; Münch, M.; Kobialka, S.; Kräuchi, K.; Steiner, R.; Oelhafen, P.; Orgül, S.; Wirz-Justice, A. High sensitivity of human melatonin, alertness, thermoregulation, and heart rate to short wavelength light. *J. Clin. Endocrinol. Metab.* **2005**, *90*, 1311–1316. [[CrossRef](#)]
25. Hoddes, E.; Zarcone, V.; Smythe, H.; Phillips, R.; Dement, W.C. Quantification of Sleepiness: A New Approach. *Psychophysiology* **1973**, *10*, 431–436. [[CrossRef](#)]
26. Ting, P.H.; Hwang, J.R.; Doong, J.L.; Jeng, M.C. Driver fatigue and highway driving: A simulator study. *Physiol. Behav.* **2008**, *94*, 448–453. [[CrossRef](#)]
27. Brown, T.; Johnson, R.; Milavetz, G. Identifying Periods of Drowsy Driving Using EEG. *Ann. Adv. Automot. Med.* **2013**, *57*, 99–108.
28. Berckmans, D.; Aerts, J.-M.; Exadaktylos, V.; Taelman, J. Method and Device for Accurate Real-Time Detection of Drowsiness in Operators Using Physiological Responses. European Patent 2,842,490, 1 June 2016.
29. Krauchi, K.; Deboer, T. The interrelationship between sleep regulation and thermoregulation. *Front. Biosci.* **2010**, *15*, 604–625. [[CrossRef](#)]
30. Boccanfuso, L.; O’Kane, J.M. Remote measurement of breathing rate in real time using a high precision, single-point infrared temperature sensor. In Proceedings of the 2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics, Rome, Italy, 24–27 June 2012; pp. 1704–1709.
31. Li, G.; Chung, W.Y. Detection of driver drowsiness using wavelet analysis of heart rate variability and a support vector machine classifier. *Sensors* **2013**, *13*, 16494–16511. [[CrossRef](#)] [[PubMed](#)]
32. Vicente, J.; Laguna, P.; Bartra, A.; Bailón, R. Drowsiness detection using heart rate variability. *Med. Biol. Eng. Comput.* **2016**, *54*, 927–937. [[CrossRef](#)] [[PubMed](#)]
33. Fujiwara, K.; Abe, E.; Kamata, K.; Nakayama, C.; Suzuki, Y.; Yamakawa, T.; Hiraoka, T.; Kano, M.; Sumi, Y.; Masuda, F.; et al. Heart Rate Variability-Based Driver Drowsiness Detection and Its Validation With EEG. *IEEE Trans. Biomed. Eng.* **2019**, *66*, 1769–1778. [[CrossRef](#)] [[PubMed](#)]
34. Li, G.; Lee, B.L.; Chung, W.Y. Smartwatch-Based Wearable EEG System for Driver Drowsiness Detection. *IEEE Sens. J.* **2015**, *15*, 7169–7180. [[CrossRef](#)]
35. Hallvig, D.; Anund, A.; Fors, C.; Kecklund, G.; Karlsson, J.G.; Wahde, M.; Åkerstedt, T. Sleepy driving on the real road and in the simulator—A comparison. *Accid. Anal. Prev.* **2013**, *50*, 44–50. [[CrossRef](#)] [[PubMed](#)]
36. Zulley, J.; Wever, R.; Aschoff, J. The dependence of onset and duration of sleep on the circadian rhythm of rectal temperature. *Pflügers Arch. Eur. J. Physiol.* **1981**, *391*, 314–318. [[CrossRef](#)] [[PubMed](#)]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).