

Driver drowsiness detection: a comparison between intrusive and non-intrusive signal acquisition methods

Licínio Oliveira
University of Porto
Porto, Portugal
up201304838@fe.up.pt

Jaime S. Cardoso
INESC TEC and University of Porto
Porto, Portugal
jaime.cardoso@fe.up.pt

André Lourenço
CardioID Technologies and
Instituto Superior de Engenharia de Lisboa, Portugal
arl@cardio-id.com

Christer Ahlström
Swedish National Road and Transport Research Institute (VTI)
Linköping, Sweden
christer.ahlstrom@vti.se

Abstract—Driver drowsiness is a major cause of road accidents, many of which result in fatalities. A solution to this problem is the inclusion of a drowsiness detector in vehicles to alert the driver if sleepiness is detected. To detect drowsiness, physiologic, behavioral (visual) and vehicle-based methods can be used, however, only measures that can be acquired non-intrusively are viable in a real life application. This work uses data from a real-road experiment with sleep deprived drivers to compare the performance of driver drowsiness detection using intrusive acquisition methods, namely electrooculogram (EOG), with camera-based, non-intrusive, methods. A hybrid strategy, combining the described methods with electrocardiogram (ECG) measures, is also evaluated. Overall, the obtained results show that drowsiness detection performance is similar using non-intrusive camera-based measures or intrusive EOG measures. The detection performance increases when combining two methods (ECG + visual) or (ECG + EOG).

Index Terms—Driver drowsiness, Camera-based methods, ECG, EOG

INTRODUCTION

Driving a car is a complex, multifaceted and potentially risky activity requiring both physiological and cognitive resources to maintain performance over time. A reduction in any of these resources can have dramatic consequences, resulting in crashes. Most people are aware of the dangers of driving while intoxicated, but many do not know that drowsiness also impairs judgment, performance and reaction times, just like alcohol and drugs do. The International Transport Forum at the OECD estimates that 20-30% of fatalities are due to driver sleepiness and fatigue [1], and in Europe, a survey across nineteen countries showed that the prevalence of falling asleep while driving in the previous 2 years is 17%, and amongst those who fell asleep, the prevalence of sleep-related crashes was 7% [2].

In summary, the impact of drowsy driving in traffic crashes is of known severity and, as so, a problem that needs to be solved.

Still, with the current investment being made in autonomous cars by every major car manufacture, thus planning a future of self-driven cars, the study of this problem could be considered unnecessary. However, the following analysis to the car automation process needs to be considered [3].

First, essentially only vehicles with SAE level¹ ≤ 2 are available as of today, that is, cars that have advanced driver assistance systems (ADAS) with lane keeping and speed adjustment functionality. These vehicles are by no means are capable of driving by themselves.

Secondly, even though SAE level 3 and 4 vehicles, where automated driving is possible in certain situations, the driver still has to be ready to take over control, sometimes after a long period of non-driving. Indeed, driver handover strategies is currently an area of big interest, in which drowsiness plays a major role [4], only emphasizing the critical necessity to further explore drowsy driving detection.

Lastly, the level five of automation is only expected to be available, in the best scenario, at least after 2025 [3], meaning that the transition between manual and full autonomous driving will be slow and a substantial proportion of the population will drive vehicles that are not fully automated for many years to come.

DROWSINESS DETECTION METHODS

Drowsiness is related to a natural physiological need, as so, it simply cannot be eliminated [5]. For this reason, the current developed works approach the problem by monitoring and alerting drivers when they are drowsy.

The main approaches used for detecting driver drowsiness in real-time can be divided into three main categories: vehicle-based measures, behavioral measures (camera-based) and physiological measures [5]–[8].

¹SAE Levels were developed by the Society of Automotive Engineers (SAE) as a classification system for self-driving cars.

Vehicle-based measures estimates the drowsiness level based on for example steering wheel movements (SWM) or the standard deviation of lane position (SDLP) [6]. Car manufacturers such as Mercedes (MERCEDES Attention Assist™) [9] and Volvo (Driver Alert Control) [10] use this technique to evaluate driver drowsiness. Vehicle-based measures have the advantage that they can be easily acquired, however these measures are confounded by geometric properties of the road, by surrounding traffic and also by other driver states such as cognitive load and visual distraction [6], [11]. In addition, with the development of automated driving functions, as conditionally automated driving which gives drivers the freedom to let go of the steering wheel, steering behavior related measures will become less significant.

Behavioral measures consist in detecting driver visual features using a camera [12], [13]. Sources of visual information can include facial expressions (wrinkles, chin, nose, lip, nasolabial fold, lid tightener and yawn), eye movements (analyzing eyelid movement as the percentage of eyelid closure (PERCLOS) and the gaze) and head movements. Based on this technique, tier 1 suppliers like Bosch [14] and NVidia [15] are developing driver drowsiness detection systems. Similarly to vehicle-based measures, one of the major advantages of visual methods is that they can be acquired non-intrusively. However, light conditions and sunlight could complicate the task, both sensor and algorithm wise, but also anatomically since the eyes are not fully visible while squinting. Also, in partially automated driving, the driver may look away from traffic, requiring a larger head box or additional cameras to provide reliable imagery of the face.

Physiological methods are based on the fact that physiological signals start to change in earlier stages of drowsiness, which could allow for earlier driver drowsiness detection [5]. The electrocardiogram (ECG), the electroencephalogram (EEG) and the electrooculogram (EOG) are physiological signals from where we can extract measures to detect drowsiness [13]. The biggest limitation using physiological signals is that they are usually acquired using intrusive methods [12] and although some approaches have been made to acquire the signals using non-intrusive systems, loss of signal quality is still significant [13]. CardioID Technologies [16], Plessey [17] and StopSleep UK [18] are examples of companies that develop non-intrusive solutions to detect driver drowsiness based on physiologic signals. It is important to take into account, that, although, in comparison, physiologic signals are one of the most reliable ways to detect drowsiness, they still have some limitations [19]. For example, while EEG is excellent to distinguish between awake and sleep states, there might be some limitations when identifying the transition (drowsy) state. Heart rate measures also show differences between alert and drowsy states, nonetheless heart rate can be affected by multiple factors.

RESEARCH AIM

As previously introduced, there are already available multiple methods to detect drowsiness, some of which are used

in commercially available cars. While each method has its strengths and weaknesses, being important to assess the combination of different measures (hybrid method) to achieve the best possible performance, real world drowsiness detection solutions need to use non-intrusive acquisition methods.

Considering this, the main focus of this work is to compare drowsiness detection performance using a non-intrusive camera-based method with a physiological method based on EOG. In addition, the combination of different methods (ECG + Video) is also assessed in order to obtain the best possible performance, understanding how much performance can be improved by fusing multiple data sources.

METHODS

In order to accomplish the cited goals, we follow an overall approach similar to Silveira et al. [20], [21], which focused on driver drowsiness detection using physiological methods. Our approach corresponds to a standard supervised machine learning classification problem workflow.

On a practical level, the first step of the process is to perform feature extraction, that is, computing drowsiness related features from the available signals. Secondly, the data (the set of all observations, each characterized by the extracted features) is randomly split in training (70%) and test (30%) datasets. The training dataset is then used to train a machine learning classifier, with 10-fold cross validation determining the best hyperparameters. Lastly, the trained classifiers are evaluated on the test dataset using a performance measure, the accuracy. For a better generalization of the results, the entire process is repeated 10 times (10-fold validation).

Dataset

In this work, all the data used is part of the SleepEye project [22]. A total of 20 participants (10 women) were recruited from the Swedish National Register of vehicle owners. Inclusion criteria were: between 30 and 60 years old, good self-reported health, normal weight, and no shift workers or professional drivers. Approximately two weeks before the experiment the participants received detailed information about the upcoming experiment, together with sleep and wake diaries that were to be filled in the three nights and two days prior to the experimental day. They also received background questionnaires. In addition, the scale to use for reporting self-reported sleepiness (Karolinska Sleepiness Scale, KSS) was sent home together with instructions and a training program to learn how to rate KSS. Permission to conduct driving sessions with sleep deprived drivers on public roads was given by the Swedish government (N2007/5326/TR). The study was approved by the regional ethics committee in Linköping (dnr 2010/1533).

The participants prepared for the experiment by sleeping for at least seven hours the three nights prior to the test. On the experimental day they were instructed to get up no later than 7:00 a.m. The participants were also requested to avoid alcohol for 72 h and to abstain from nicotine and caffeine for 3h before the first driving session until the end of the

experimental day. Each driver performed one driving session during daytime (supposedly alert condition) and one session during night-time (sleep deprived condition). Two participants took part in the experiment on the experimental day. The car used in the experiment was a Volvo XC70 with an automatic gearbox. A test leader accompanied all participants and was seated in the front passenger seat. The car had dual command and the test leader was prepared to take control of the vehicle if the driver became too sleepy.

The participants were asked to evaluate their state using the KSS scale [23] (9 levels: 1 (extremely alert); 3 (alert); 5 (neither alert nor sleepy); 7 (Sleepy, but no effort to keep alert; 9 (Very sleepy, great effort to keep alert, fighting sleep)) at 5-minute intervals and physiological data, ECG and EOG signals, were recorded by a Vitaport 3 (TEMEC Instrument BV, Kerkrade, the Netherlands). Additionally, behavioral data, obtained by a SmartEye Pro 5.7 system [24], was recorded. More details can be found in [22].

Labels

In order to categorize sleepiness, the available KSS scale for each session was used. As described, the KSS is a 9-point Likert scale, a typical self-report rating format where subjects rank a quality from high to low or best to worst and one of the most common used drowsiness scales [25].

Specifically, a binary and a three level multiclass problem divisions were considered. In the binary approach, the driver is treated as either awake or drowsy, for the multiclass problem an intermediate medium state is also contemplated. The class labels are obtained from the KSS according to following scheme, Fig. 1, as in [20], [21].

2-class	awake						drowsy		
KSS rating	1	2	3	4	5	6	7	8	9
3-class	awake					medium		drowsy	

Fig. 1. KSS rating and their corresponding states for 2-class and 3-class problems.

Features

The features characterizing each observation were extracted from the ECG, EOG and video measures analyzing the signals in non-overlapping 2-minute windows. Regarding the physiological signals (ECG and EOG), the pre-processing and feature extraction steps are identical to the work by Silveira et al. [20], [21].

ECG: Before extraction of features from the ECG signal, a bandpass filter was applied to eliminate high and low frequency noise, especially related to power line interference (50 Hz) and baseline wander due to respiration or motions. After preprocessing, a modified version [26] of the Pan-Tompkins QRS detection algorithm [27] was used for R-peak detection. An example of the ECG signal and detected peaks is illustrated in Fig. 2.

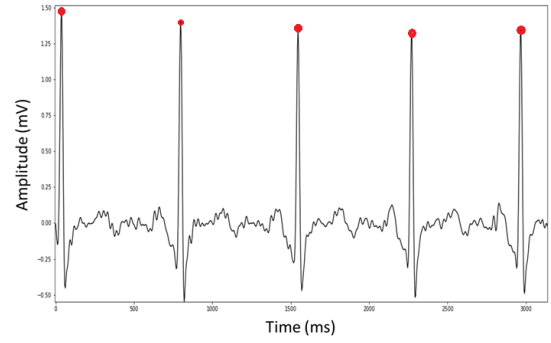


Fig. 2. Example of ECG signal with detected R peaks marked in red points.

After R-peak detection, 16 heart rate variability features from the time and frequency domain were extracted, 8 from time domain (e.g. heart rate) and 8 from frequency domain (mainly related with energy in different frequency bands).

EOG: The vertical and horizontal EOG channels were bandpass filtered similarly to the ECG signal. Eye blink information was extracted from the vertical EOG based on differentiation and thresholding. Extracted features include blink duration, blink frequency and blink velocity. An example showing the start and end points for each blink is provided in Fig. 3. The horizontal channel mostly contain information

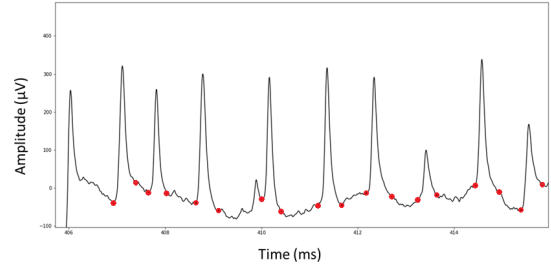


Fig. 3. Example of vertical EOG signal with detect blinks (start and end points) marked in red points.

about saccade eye movements. Using a similar technique to the one applied in the vertical channel, 8 features characterizing those movements were extracted. Accounting both channels, in total, 25 features are computed.

Video: The video measures obtained with the SmartEye system are the eyelid distance, gaze angles, head pose and pupil diameter. From the eyelid distance signal, 7 features similar to the vertical EOG features were extracted. This was done based on the work by Ji [28], using 4 points (t_1 , t_2 , t_3 , t_4) to characterize each blink. To compute these points, first, the signal valleys (closed eyes) were located based on the signal's first order difference and an adaptive threshold. Then, for each detected location, the signal was searched before and after the valley until it stops increasing within an established range, obtaining, respectively, the t_1 and t_4 points. Lastly, t_2 was defined by the position where the signal is approximately 60% of the difference between t_1 and the valley opening value. The analog process is performed for t_3 using t_4 .

An eye is denoted closed from point t2 to point t3 and the average eye opening/closing time is compute from point t1 to point t2/point t3 to point t4. An illustration of the blink detecting scheme is presented in Fig. 4.

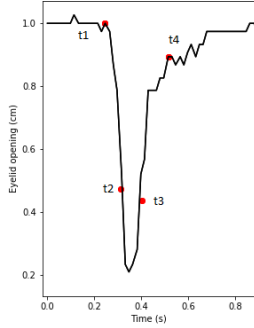


Fig. 4. Scheme of points (t1 to t4 points marked in red) characterizing an eye blink.

From the gaze direction measures, 3 features were extracted: the mean and standard deviation of the combined yaw and pitch gaze angles as well as the total fixation time. Several fixation segmentation algorithms are available [29]. In this work, a velocity based algorithm was implemented based on the two main velocity distributions that can be found on eye movements: low velocities for fixations (<100 deg/s) and high velocities for saccadic eye movements (>300 deg/s) [29].

Similar to the gaze direction measures, the mean and standard deviation of the combined yaw, pitch and roll head pose angles were calculated. In addition, head nods were also identified based on thresholding of the head pose pitch angle. The nodding frequency, average nodding duration and total nodding duration time were used to characterize the detected head nods. Regarding pupil diameter, average and standard deviation values were computed. A total of 17 features was extracted from the video data.

Models

As described, using the samples characterized by the extracted features and the KSS label, the driver state ground truth reference, a model was trained and evaluated. There are several machine learning models used typically in supervised classification tasks, however there is no uniform classification method for a certain application, each classifier has its strengths and weaknesses, but ultimately, the classifier performance depends on the data.

As such, in this work, five machine learning classifiers were used: Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Networks (ANN), Gradient Boosting Tree (GBT) and K-Nearest Neighbors (KNN).

Performance evaluation

In order to evaluate the models performance, accuracy was used. Accuracy expresses the percentage of samples correctly

classified, and for a two class problem it can be calculated as in (1).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (1)$$

Where TP are the true positives, expresses the number of drowsy states (given) that are correctly classified as drowsy, true negative (TN) the number of alert states (given) that are correctly classified as alert, false positive (FP) the number of alert states (given) that are incorrectly classified as drowsy and false negative (FN) the number of drowsy states (given) that are incorrectly classified as alert. Similarly, for a three class problem, accuracy is computed as in (2).

$$Accuracy = \frac{\sum_{i=1}^3 M_{i,i}}{\sum_{i=1}^3 \sum_{j=1}^3 M_{i,j}} \times 100\% \quad (2)$$

Where M represents the 3×3 confusion matrix.

To analyze the models performance per class, F1-score was used. F1-score is the weighted average of Precision and Recall (3).

$$F1\text{-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \quad (3)$$

Precision is the ratio of correctly predicted observations of a class i by the total predicted observations of that class. For example, the correctly predicted drowsy observations over the total number of samples classified as drowsy. Recall is the ratio of correctly predicted observations of class i by all the observations of the class. For example, the correctly predicted drowsy observations over all instances that have a label drowsy.

RESULTS

The obtained results using the previously described framework, for multiple features and classifiers is shown in Fig. 5. It is recognizable that the ECG alone shows poor performance

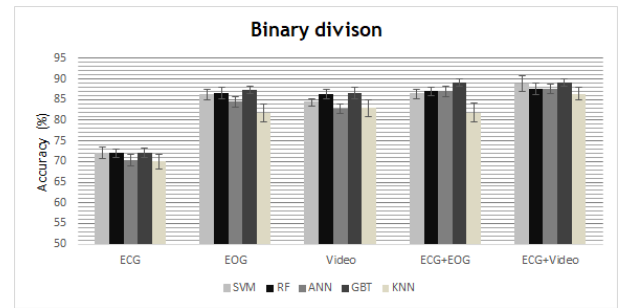


Fig. 5. Accuracy results (average and standard deviation) for all classifiers using different features and binary class division.

in comparison with EOG and video features. Also, the overall performance obtained using EOG features is similar to the performance using video features. The same observation can be made when comparing ECG+EOG and ECG+video features. There is a slight improvement when using a hybrid

approach including ECG features in addition to the EOG or video features. Regarding the different classifiers, there are some minor differences in their performance's; however, no classifier seems to be far superior to the others.

Although the overall accuracy results are satisfactory, particularly for the hybrid approaches, it is important to evaluate the individual classes performance. The obtained F1-score results for each class “alert” and “drowsy” are displayed in Fig. 6.

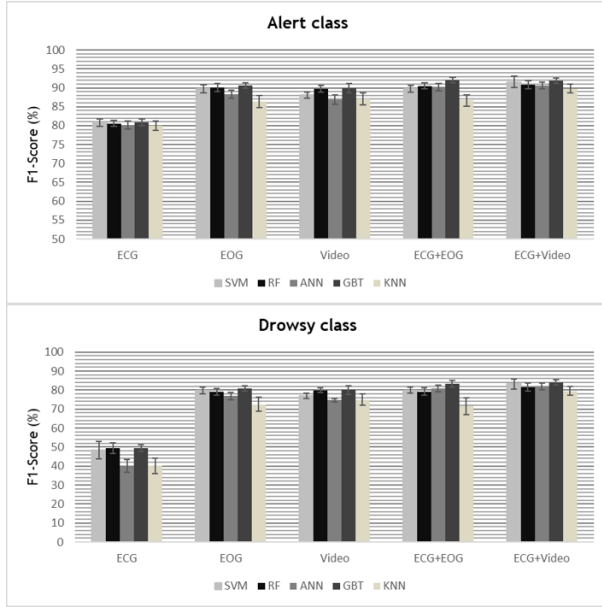


Fig. 6. Alert and drowsy class's F1-score (average and standard deviation) for all classifiers using different features and binary class division.

Comparing the F1-score for both classes, there is a clear difference between the performance of the alert and drowsy classes, with the alert class having for all cases a better performance. This difference is most noticeable for the approach using only ECG features, emphasizing that those features alone are not enough to accurately evaluate the driver drowsiness state.

The obtained accuracy results applying the same methods but considering a multiclass driver drowsiness categorization approach are displayed in Fig. 7.

For this approach, the observations previously made to the binary problem division can still be applied, however the overall accuracies are clearly worse, about 10% lower compared to the binary case.

Regarding the performance on individual classes (Fig. 8), it is clear that the alert class has the best performance also in multiclass classification. The performance on the intermediate class is approximately 10% lower than the alert class, and the drowsy class is 20% lower.

Imbalance between the classes might be one reason for these differences. However, there might also be individual subject differences in the KSS scale used as ground truth, as one subject's feeling of drowsiness can be considerable different from another subject, that can also decrease the model

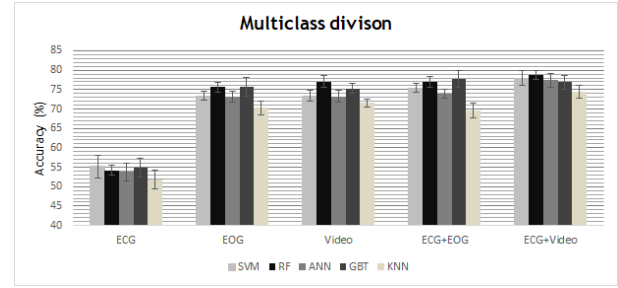


Fig. 7. Accuracy results (average and standard deviation) for all classifiers using different features and multiclass division.

performance to correctly determine what are the characteristics that define the drowsy state.

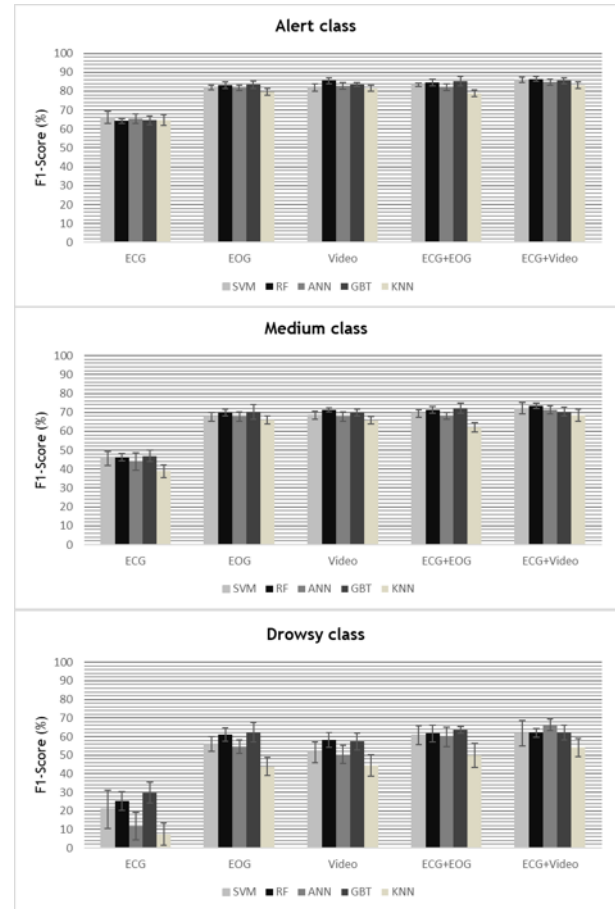


Fig. 8. Alert, medium and drowsy class's F1-score (average and standard deviation) for all classifiers using different features and multiclass class division.

CONCLUSIONS

In summary, the obtained results are satisfactory as the drowsiness detection performance was similar using non-intrusive video measures and intrusive EOG measures. However, the overall accuracy still needs to be improved, particularly for the multiclass approach where the drowsy class has

considerably lower performance. The hybrid approach using ECG, which has the potential to be acquired non-intrusively, and video measures can provide some extra robustness to the system, but still further development is needed.

ACKNOWLEDGEMENT

This work was partially financed by the ERDF European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme and by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project "POCI-01-0145-FEDER-030707", and in collaboration with the ADAS&Me project which is funded by the European Unions Horizon 2020 research and innovation program under grant agreement N^o 688900.

REFERENCES

- [1] I. T. Forum, *Road Safety Annual Report 2017*. OECD iLibrary, 2017.
- [2] M. Gonçalves, R. Amici, R. Lucas, T. Kerstedt, F. Cirignotta, J. Horne, D. Lger, W. T. McNicholas, M. Partinen, J. Tran-Santos, P. Peigneux, and L. a. Grote, "Sleepiness at the wheel across europe: a survey of 19 countries," *Journal of Sleep Research*, vol. 24, no. 3, pp. 242–253, 2015.
- [3] NHTSA, "Automated Vehicles for Safety." [Online]. Available: <https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety>
- [4] M. Walch, K. Mühl, J. Kraus, T. Stoll, M. Baumann, and M. Weber, *From Car-Driver-Handovers to Cooperative Interfaces: Visions for Driver-Vehicle Interaction in Automated Driving*. Springer International Publishing, 2017, pp. 273–294.
- [5] A. Colic, O. Marques, and B. Furht, "Driver drowsiness detection: Systems and solutions," 2018. [Online]. Available: <http://what-when-how.com/Tutorial/topic-2046rgjofl/Driver-Drowsiness-Detection-Systems-and-Solutions-bookmark.html>
- [6] A. Sahayadhas, K. Sundaraj, and M. Murugappan, "Detecting driver drowsiness based on sensors: A review," *Sensors (Switzerland)*, vol. 12, no. 12, pp. 16 937–16 953, 2012.
- [7] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 596–614, 2011.
- [8] M. Awais, N. Badruddin, and M. Driberg, "A hybrid approach to detect driver drowsiness utilizing physiological signals to improve system performance and Wearability," *Sensors (Switzerland)*, vol. 17, no. 9, pp. 1–16, 2017.
- [9] Mbusa, "Attention Assist - Vehicle Safety Video — Mercedes-Benz." [Online]. Available: <https://www.mbusa.com/mercedes/technology/videos/detail/title-safety/videoId-710835ab8d127410VgnVCM100000ccec1e35RCRD>
- [10] VolvoCars, "Volvo Cars Support." [Online]. Available: <https://support.volvocars.com/en-CA/cars/Pages/owners-manual.aspx?mc=v526t8hbat&my=2016&sw=15w46&article=2212bb11e5e1c274c0a8015154096fb6>
- [11] J. Schmidt, R. Laarousi, W. Stolzmann, and K. Karrer-Gauß, "Eye blink detection for different driver states in conditionally automated driving and manual driving using EOG and a driver camera," *Behavior Research Methods*, pp. 1–14, 2017.
- [12] B. G. Pratama, I. Ardiyanto, and T. B. Adj, "A review on driver drowsiness based on image, bio-signal, and driver behavior," *Proceeding - 2017 3rd International Conference on Science and Technology-Computer, ICST 2017*, pp. 70–75, 2017.
- [13] H. B. Kang, "Various approaches for driver and driving behavior monitoring: A review," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 616–623, 2013.
- [14] Bosch, "CES® 2017: Bosch is showing these smart solutions in Las Vegas - Bosch Media Service." [Online]. Available: <http://www.bosch-presse.de/pressportal/de/en/ces-2017-bosch-is-showing-these-smart-solutions-in-las-vegas-81792.html>
- [15] Nvidia, "NVIDIA DRIVE IX Self-Driving Vehicles — NVIDIA Automotive." [Online]. Available: <https://www.nvidia.com/en-us/self-driving-cars/drive-ix/>
- [16] CardioID, "CardioID — Every heart has a beat, but the way we use it is unique!" [Online]. Available: <https://www.cardio-id.com/>
- [17] Plessey, "Plessey Semiconductors WARDEN ECG monitoring system." [Online]. Available: <http://www.plesseysemiconductors.com/products/warden/>
- [18] StopSleep, "StopSleep : the best solution against drowsiness." [Online]. Available: <https://www.stopsleep.co.uk>
- [19] J. M. Mullington, C. A. Czeisler, N. Goel, J. M. Krueger, T. J. Balkin, M. Johns, and P. J. Shaw, "Panel discussion: current status of measuring sleepiness," *Journal of clinical sleep medicine*, vol. 7 5 Suppl, pp. S22–5, 2011.
- [20] S. Silveira, "Driver's Fatigue State Monitoring using Physiological Signals," Master's thesis, FEUP, 2017.
- [21] C. Silveira, J. S. Cardoso, A. Lourenco, and C. Ahlstrom, "The importance of subject-dependent classification and imbalanced distributions in driver sleepiness detection in realistic conditions," *IET Intelligent Transport Systems*, accepted for publication.
- [22] C. Fors, C. Ahlstrom, P. Sorner, J. Kovaceva, E. Hasselberg, M. Krantz, J.-F. Gronvall, K. Kircher, and A. Anund, "Camera-based sleepiness detection: final report of the project sleepeye," *Vip Publication*, 2011. [Online]. Available: <http://www.vti.se/en/publications/pdf/camera-based-sleepiness-detection.pdf>
- [23] T. Åkerstedt and M. Gillberg, "Subjective and objective sleepiness in the active individual," *The International journal of neuroscience*, vol. 52 1-2, pp. 29–37, 1990.
- [24] SmartEye, "SmartEye AB, Gothenburg, Sweden." [Online]. Available: <http://smarte.se/>
- [25] T. Åkerstedt, A. Anund, J. Axelsson, and G. Kecklund, "Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function," *Journal of Sleep Research*, vol. 23, no. 3, pp. 242–254, 2014.
- [26] P. S. Hamilton, "Open Source ECG Analysis Software Documentation," *Computers in Cardiology*, vol. 2002, IEEE, pp. 101–104, 2002. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1166717>
- [27] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, 1985.
- [28] Q. Ji, "Real-Time Eye, Gaze, and Face Pose Tracking for Monitoring Driver Vigilance," *Real-Time Imaging*, vol. 8, no. 5, pp. 357–377, 2002.
- [29] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," *Proceedings of the symposium on Eye tracking research & applications - ETRA '00*, pp. 71–78, 2000.