Faculdade de Engenharia da Universidade do Porto

Driver drowsiness detection using non-intrusive signal acquisition

Licínio Manuel França de Oliveira

MASTER'S THESIS

Integrated Master in Bioengineering

Internal Supervisor: Jaime dos Santos Cardoso, PhD
External Supervisor: André Ribeiro Lourenço, PhD

July 12, 2018

Abstract

Driving a car is a complex, multifaceted and potentially risky activity requiring full mobilization of physiological and cognitive resources. Drowsiness leads to a decline in cognitive performance, being the cause of many driving accidents. In multiple European countries drowsy driving accounts for 10 to 30 % of all road accidents. A solution to this problem is the inclusion of Advanced Driver Assistance Systems in vehicles that can warn the driver if sleepiness is detected.

In order to detect drowsiness, subjective, vehicle-based, behavioral (visual based), physiologic and hybrid methods are used. Hybrid methods consist in a fusion of the previous ones and are, arguably, the most promising approach to efficiently detect drowsiness as it combines the strengths of the different methods. However, the individual limitations of each still need to be overcome, as the intrusive conditions needed to properly acquire a physiologic signal, generally considered the more reliable to detect drowsiness.

This thesis focus on detecting drowsiness using non-intrusive measures, as it is the most promising solution to use on a real life application. As such, one of the goals was to assess if the performance of a drowsiness detector based in non-intrusive behavioral measures was comparable to a detector based in a physiologic intrusive signal (EOG).

Using the measures from a large validated dataset (SleepEye), acquired in an experimental field study in real traffic on real roads, and applying a supervised machine learning classification problem it was concluded that the extracted video features (head orientation, eyelid opening, pupil diameter and gaze direction) performance was similar to the obtained using only EOG. Specifically, the best results have an accuracy, with a total of 17 video features and for multiple classifiers, of roughly 84% for a binary problem categorization (‘awake’ vs ‘drowsy’) and 73% for a three class problem (‘awake’ vs ‘medium’ vs ‘drowsy’).

In view of the potential of a hybrid methodology, several experiments were done, using the same dataset, combining video (17 features) with ECG (16 features). The best accuracy result, similar for multiple classifiers, was close to 89 %. Again, the two class problem was better than the three class by around 10%. Other elaborated strategies as classifier combination, imbalanced data and sequential data approaches were also tested but no considerable

improvements were obtained, supporting that, likely, the obtained results are the limit of what is achievable with this dataset or with the selected features.

Due to the restrictions of the first dataset, that is, only measures already extracted were available without direct access to the raw video source, an experimental work was developed to allow an enriched measure extraction from video. Specifically, the developed framework allows the extraction of additional drowsiness related measures by analyzing the drivers face with a standard camera that can be easily implemented in a real life application. However, further generalization of the algorithm to different light conditions and camera positioning still need to be researched.

In summary, using a non-intrusive drowsiness detection system is the most promising direction to build a real life applicable solution. A solution that could be enhanced by taking advantage of different methods, i.e., a hybrid system.

Keywords: Drowsiness detection, physiologic measures, behavioral measures, non-intrusive hybrid system, machine learning, computer vision.

Resumo

Conduzir um carro é uma atividade complexa, multifacetada e potencialmente arriscada que requer a utilização de recursos fisiológicos e cognitivos na sua totalidade. A sonolência leva a um declínio no desempenho cognitivo, sendo a causa de muitos acidentes de viação. Em vários países europeus, a condução sonolenta é responsável por 10 a 30% de todos os acidentes rodoviários. Uma solução para este problema é a inclusão em veículos de sistemas de deteção de sonolência para alertar o motorista se esta for detetada.

Para detetar a sonolência, são utilizados métodos subjetivos, métodos baseados em veículos, no comportamento dos condutores, fisiológicos e híbridos. Os métodos híbridos consistem numa fusão dos restantes e são, questionavelmente, a abordagem mais promissora para detetar eficientemente a sonolência, uma vez que combina as vantagens dos diferentes métodos. No entanto, as limitações individuais de cada um necessitam de ser superadas, como é o caso das condições intrusivas necessárias para adquirir adequadamente sinais fisiológicos, geralmente considerados os mais fidedignos para detetar sonolência.

Esta tese foca-se na deteção de sonolência usando medidas não-intrusivas, visto ser a forma mais promissora para usar numa aplicação em condições reais. Como tal, um dos objetivos foi avaliar se o desempenho de um detetor de sonolência baseado em medidas comportamentais, não intrusivas, era comparável a um detetor baseado num sinal fisiológico (EOG), intrusivo.

Usando as medidas de um *dataset* autenticado (SleepEye), adquirido num estudo de condução naturalista, e aplicando um problema supervisionado de classificação de *machine learning*, foi possível concluir que o desempenho usando medidas visuais (orientação da cabeça, abertura das pálpebras, diâmetro da pupila e direção do olhar) é semelhante ao obtido usando apenas o sinal de EOG. Concretamente, os melhores resultados, obtidos com um total de 17 características de vídeo e para diferentes classificadores, tem uma precisão, aproximadamente, de 84% para uma divisão binária do problema ('acordado' vs 'sonolento') e 73% para um problema de três classes ('acordado' vs 'médio' vs 'sonolento').

Tendo em conta a potencialidade de uma metodologia híbrida, foram realizados diversos testes, utilizando o mesmo conjunto de dados, combinando vídeo (17 características) com ECG (16 características). O melhor resultado, semelhante para vários classificadores, foi cerca de

89%. Novamente, a divisão binária do problema foi cerca de 10% melhor do que a divisão em três classes. Outras estratégias como combinação de classificadores, *imbalancing approaches* e de dados sequenciais foram também testadas, mas nenhuma melhoria considerável foi obtida, concluindo-se que, provavelmente, os resultados obtidos são o limite do que é alcançável com este conjunto de dados ou com as características selecionadas.

Devido às restrições do primeiro *dataset*, ou seja, apenas medidas já extraídas estavam disponíveis, sem possibilidade de acesso direto aos videos originais, um trabalho experimental foi desenvolvido para permitir a extração de medidas adicionais indicadoras de sonolência. O *framework* desenvolvido permite a extração destas medidas, analisando a face dos motoristas com uma câmara *standard* que pode ser facilmente introduzida num veículo. Apesar de promissor, algumas limitações ainda estão presentes, nomeadamente, o algoritmo desenvolvido necessita de maior robustez face a diferentes condições de luminosidade e posicionamento da câmara.

Em suma, usar um sistema de deteção de sonolência baseado em métodos de aquisição não intrusivos é a direção mais promissora para construir uma solução aplicável em condições reais. Uma solução que poderá beneficiar da combinação de diferentes métodos, ou seja, um sistema híbrido.

Palavras-chave: Deteção de sonolência, medidas fisiológicas, comportamento condutor, sistema híbrido não intrusivo, *machine learning*, visão computacional.

Acknowledgements

First I would like to thank my supervisor, Professor Jaime Cardoso, for the continuous support and availability during my thesis development, for all the suggestions and advices. I would also like to thank him for promoting the collaboration and knowledge exchange between his students.

My sincere thanks also goes to my external supervisor Professor André Lourenço, from CardioID Technologies, who personally recorded a driving session to develop my research, and also to Professor Christer Ahlström, from Swedish National Road and Transport Research Institute, for providing the datasets that made this work possible. Their collaboration was very important and valuable in order to improve my work.

To the Faculty of Engineering of the University of Porto for the knowledge and excellent conditions provided during all my degree.

Last but not the least, I would like to thank to all my friends and my girlfriend for all the motivation and support. And, a final remark, to my family, especially to my parents, who always supported me, throughout my personal and academic life.

Contents

List of Figures

List of Tables

Abbreviations

ADAS	Advanced Driver Assistance System
ADS	Automated Driving Systems
AECS	Average Eye Closure Speed
ANN	Artificial Neural Networks
BN	Bayesian networks
CAN	Controller Area Network
CCD	Charge-Coupled Device
CES	Consumer Electronics Show
CMOS	Complementary Metal- Oxide Semiconductor
CNN	Convolutional Neural Network
DAC	Driver Alert Control
DWT	Discrete Wavelet Transform
ECG	Electrocardiogram
EDA	Electrodermal activity
EEG	Electroencephalogram
EOG	Electrooculogram
EMG	Electromyogram
ESS	Epworth Sleepiness Scale
FN	False Negative
FP	False Positive
FPS	Frames per Second
GW	Gabor Wavelets
HF	High Frequency
HMM	Hidden Markov Model
HOG	Histogram of Oriented Gradients
HR	Heart Rate
HRV	Heart Rate Variability
IR	Infrared
KNN	K-Nearest Neighbors
KSS	Karolinska Sleepiness Scale
LBP	Local Binary Pattern
LDA	Linear Discriminant Analysis
LED	Light Emitting Diode

LF	Low Frequency
MWT	Maintenance of Wakefulness Test
MSLT	Multiple Sleep Latency Test
NREM	Non-Rapid Eye Movement
NHTSA	National Highway Traffic Safety Administration
PERCLOS	PERcentage of Eyelid CLOsure
PnP	Perspective n Point
PPG	Photoplethysmogram
REM	Rapid Eye Movement
RF	Random Forest
RMSE	Root Mean Square Error
SDLP	Standard Deviation of Lane Position
SSS	Stanford Sleepiness Scale
SVM	Support Vector Machine
SWM	Steering Wheel Movement
TN	True Negative
TP	True Positive
VAS	Visual Analogue Scales
VLF	Very Low Frequency
VTI	Swedish National Road and Transport Research Institute
WPD	Wavelet Packet Decomposition
WT	Wavelet Transform

Chapter 1

Introduction

1.1 Context and motivation

Driving a car is a complex, multifaceted and potentially risky activity requiring full mobilization of physiological and cognitive resources to maintain performance over time. Any loss of these resources can have dramatic consequences, namely accidents [1].

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. Studies show that being awake for more than 20 hours results in an impairment equal to a blood alcohol concentration of 0.08%, the legal limit in the United States [2].

A report from U.S. National Highway Traffic Safety Administration (NHTSA) [3], stated that drowsy driving was reportedly involved in 2.3 to 2.5% of all fatal crashes nationwide from 2011 through 2015. In 2015, 2.3% (824) of the fatalities that occurred on U.S. roadways are reported to have involved drowsy driving. Also, according to data from Australia, England, Finland, and other European nations, all of whom have more consistent crash reporting procedures than the U.S., drowsy driving represents 10 to 30 percent of all crashes [4].

In summary, the impact of drowsy driving in traffic accidents 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 [5], the study of this problem could be considered unnecessary.

However, the following analysis to the car automation process needs to be considered. Car automation can be grouped into six levels, where the first level (zero) describes a car with no autonomy and the last level (five), a fully autonomous car [6].

First, currently, there are broadly available only vehicles with level two of automation, that is, cars that have advanced driver assistance systems (ADAS), the vehicle can itself actually control both steering and braking/accelerating simultaneously under some restricted circumstances.

Secondly, while level three and four automation systems, that is Automated Driving Systems (ADS) where the vehicle can itself perform all aspects of the driving task under some circumstances, are expected to be gradually introduced over the next few years, the driver may still need to take to control of the vehicle 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 [7], only empathizing the critical necessity to further explore drowsy driving.

Lastly, the level five of automation is only expected to be available, in the best scenario, at least after 2025 [6], meaning that the transition between manual and full autonomous driving will be slow and a substantial proportion of the population will still drive non full automated cars for a significant number of years.

1.2 Challenges and goals

Drowsiness is related to a natural physiological need, as so, it simply cannot be eliminated [8]. For this reason, the current developed works approach the problem by monitoring and alerting the driver when they are drowsy or in a distracted state [9].

Multiple researches have already been done in this field, common techniques include monitoring the vehicle state, analyzing driver behavior or driver physiologic signals [10].

Vehicle monitoring correlation with driving drowsy has been shown [11], [12], and some cars manufacturers use this technique to evaluate driver drowsiness, as Volkswagen [13], Mercedes (MERCEDES Attention Assist™) [14] and Volvo (Driver Alert Control) [15]. However, vehicle-based measures are too dependent on road conditions [10].

Driver behavior monitored by a camera provides efficient measures, yawning, blinking and head orientation are all good indicators of a driver's fatigue level [16]. Based on this technique, car suppliers like Bosch [17] and NVidia [18] are developing driver drowsiness detection systems. Typically, these method's disadvantages are the dependence on proper lighting conditions and high difficulty in analyzing some features when the driver uses some object as glasses [19].

A more efficient approach is analyzing driver physiologic features, as they have high accuracy and allow an early drowsiness detection (physiological signals start to change in earlier stages of drowsiness) [8]. The biggest limitation using physiological signals is that they are usually acquired using intrusive methods [19] and although some approaches have been made to acquire signal using non-intrusive systems, loss of signal quality is still significant [20]. CardioID Technologies [21], Plessey [22] and StopSleep UK [23] are examples of companies that develop non-intrusive solutions to detect driver drowsiness based on physiologic signals.

In order to overcome the disadvantages of the different methods, some researches present as solution a fusion of different measurement types, creating a hybrid driver drowsiness detection system. Although the number of studies combining several types of measurement types is still reduced, current studies suggest that a combination of the three types of methods—behavioral, physiological and vehicle-based—is a promising avenue worth pursuing in the development of real-world, vehicle-mounted, driver drowsiness detection solution [8].

The overall challenge is ambitious: not only detecting, but also predicting, degradation in the driver's operational state [20].

Considering all the above mentioned, the main purpose of this work is developing algorithms to create an accurate drowsiness detection system, with high relevance for methods that do not require intrusive acquisition systems, thus contemplating a valid introduction in a real world application. Combination of multiple drowsiness detection methods should also be assessed. Ultimately, the goal is leading to safer driving conditions and reduction in drowsy driving associated accidents.

1.3 Contributions

In summary, the developed work contributes to the driver drowsiness detection field in the following aspects:

- Through a detailed literature analysis, present a clear explanation of the drowsiness concept and its causes, a review of state of the art methods for drowsiness detection both in scientific and commercial solutions, as well as, used experimental setups.

- Compare, using a large and validated dataset, containing more than 40 hours of driving data collected under a naturalist driving study, the drowsiness detection performance between non-invasive behavioral measures and invasive physiologic signals, as the ECG and EOG.

- Analyze a series of machine learning strategies that could be used to improve the overall drowsiness detection capability independently of the used features, but focusing on a multimodal approach, which is still not properly studied.

- Developing an experimental procedure, to easily allow the extraction of a high amount of driver drowsiness related features using a standard camera.

1.4 Thesis outline

Besides the current chapter, this document has four chapters.

The next chapter (Chapter 2) is a revision of the state of the art of driver drowsiness detection. Within this chapter, first, a clarification of the drowsiness concept and the causes that most contribute to a drowsy state is presented, followed by a revision of the main techniques used to detect drowsiness in scientific and commercial solutions. A summary of the main experimental setups used to conduct drowsiness detection studies and performance measures is also presented.

The third chapter (Chapter 3) reports the developed work using a supervised machine learning approach to classify the driver drowsiness state. This includes extracted features, tested classifiers and several machine learning strategies that are used to improve the detection performance.

The fourth chapter (Chapter 4) explains the exploratory framework developed in order to extract drowsiness related measures using a standard camera.

Lastly, in the fifth chapter (Chapter 5), the project main conclusions and future work suggestions are discussed.

Chapter 2

Literature review

In this chapter, first a clarification of the drowsiness concept and its causes are addressed, followed by a review of state of the art methodologies for drowsiness detection, including commercial solutions. The experimental setups and typical evaluation methodologies are also discussed.

2.1 Driver drowsiness

Sleep is an active period in which a lot of essential processing, restoration, and strengthening occurs [24]. It is a ubiquitous biological imperative process that appears to be evolutionarily conserved across species [25].

This process takes place in repeating cycles, averaging 90 minutes, and in each cycle the body alternates between two distinct modes: rapid eye movement (REM) sleep, the last part of the sleep cycle where typically dreams occur, and non-REM sleep, which consists of three separate stages [10], [26]:

Stage I: transition from awake to asleep;
Stage II: light sleep;
Stages III: deep sleep.

These stages can be distinguished by measuring electrical activity in the brain with electroencephalography (EEG) or the eye movements via electrooculography (EOG) or muscle activity via electromyography (EMG).

Drowsiness is a transitional state between wakefulness and sleep (stage I) in which the "sleep onset process" has already begun, albeit intermittently, and is likely to proceed to sleep [27], [28], [29]. During this phase a reduction in vigilance (alertness) and performance is often observed.

In the context of the traffic safety and transportation field, the term drowsiness is often used interchangeably with sleepiness and fatigue, although having different meanings in research and clinical settings [4], [30], [31].

Specifically, sleepiness is a more subjective concept than drowsiness. Usually it refers to the physiological drive to sleep, that is, it can be considered as a measure of a subject's tendency at a particular time to doze or fall asleep, at least briefly [4], [27], [30].

Fatigue is a multidimensional construct that has been difficult for researchers to define [32], but typically is described as a global reduction in physical or mental arousal that results in a performance deficit and a reduced capacity of performing a task [30].

A person can be fatigued without being sleepy/drowsy, but a person cannot be sleepy without being fatigued [30]. It can be inferred that drowsiness/sleepiness is a consequence of fatigue [11]. Driver fatigue can be subcategorized into sleep-related (SR) and task-related (TR) fatigue based on the causal factors contributing to the fatigued state [32]. Figure 2.1, adapted from [32], illustrates this division, the types of fatigue, their causes, consequences and interactions.

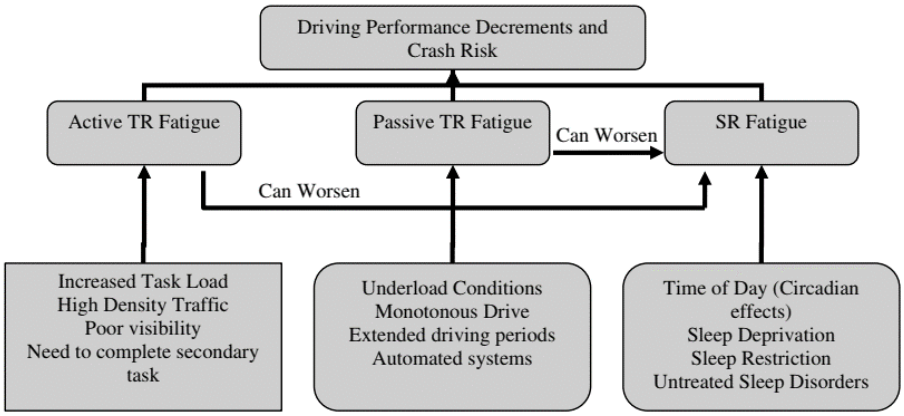

Figure 2.1 - Fatigue model.

SR fatigue is mainly influenced by circadian rhythms, sleep deprivation and sleep restriction.

Circadian rhythms are governed by an internal body clock that completes a cycle approximately every 24 h, circadian pacemaker is the name given to that clock. This internal clock is in the origin of sleepiness peaks at certain day times [8]. Moreover, a dysregulation of circadian rhythms results in increased drowsiness states, as it is the case when traveling to a different time zone. This happens because our circadian rhythms are slow to adjust and remain on their original biological schedule for several days, leading to sleepiness peaks during the middle of the afternoon or wakefulness peaks at late night. This experience is known as jet lag [33].

In relation to homeostatic factors, such as the duration of wakefulness and sleep deprivation, there is a correlation between the level of sleep deprivation and performance, that is, performance becomes worse the longer a person remains awake [34], [32].

Furthermore, drowsiness is affected by sleep quality or not obtaining adequate sleep (sleep restriction). Although the need for sleep varies among individuals, sleeping 8 h per 24-h period is common, and 7–9 h is needed to optimize performance. Experimental evidence shows that sleeping less than 4 consolidated hours per night impairs performance on vigilance tasks [8].

TR fatigue is caused by the driving task and driving environment. TR fatigue can be further subdivided into active or passive fatigue [32], [35], [36].

Active fatigue is generally related to mental overload (high demand) driving conditions being the most common form of TR fatigue that drivers experience. Examples of high task demand situations include high density traffic, poor visibility, or the need to complete an auxiliary or secondary task (e.g. searching for an address) in addition to the driving task [35], [32].

Passive fatigue, in contrast, develops by experiencing underload conditions, for example, when the roadway is monotonous and there is little traffic. In this scenario, the driver is mainly monitoring the driving environment, that is, the driving task is predictable, which may lead to a reduction in effort exerted on the task [36], [32].

Despite the different causes of fatigue, independently of being sleep related or not, the perceptible signs are essentially the same. From a drivers perspective, the key signs of drowsiness are: trouble focusing, keeping eyes open or head up; yawning or rubbing the eyes repeatedly; daydreaming and wandering thoughts; drifting from lane; tailgating and missing signs or exits; feeling restless, irritable or aggressive and overall slower reaction time [2]. Furthermore, the traffic safety consequences are also equal, that is, an impaired performance at the wheel, with increased crashing risk, which can ultimately result in falling asleep at the wheel [32], [37].

For these reasons, in this work, whose main objective is to generically detect drowsiness, no causal factors were taken into consideration. However, a distinction should be taken in account when designing a countermeasure system, as sleeping and/or caffeine can be considered the only effective countermeasure for drowsiness/sleepiness, but the countermeasure for task related fatigue could be other [32], [30], for example, the use of automation technologies.

In addition, it should be noted that, in the following sections, no further distinction is made between fatigue, sleepiness and drowsiness.

2.2 Drowsiness detection methods

The main approaches for detecting driver drowsiness can be divided into four categories: subjective measures, vehicle-based measures, behavioral measures and physiological [8], [10], [38], [39]. A fifth category, hybrid methods, can also be considered, consisting in a combination of the methods of the previous categories.

2.2.1 Subjective measures

This type of measures is based on questionnaires which are performed either by the driver, that is, the driver assesses its own state of drowsiness, or by a supervisor evaluation. Some of the most used questionnaires include Karolinska Sleepiness Scale (KSS), Stanford Sleepiness Scale (SSS), Visual Analogue Scales (VAS) and Epworth Sleepiness Scale (ESS).

The Karolinska Sleepiness Scale is the most common used drowsiness scale [8], [40]. It 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 [41]. Several studies have already focused in validating the KSS, showing, for example, that there is a relatively strong intra-individual correlation between the KSS and electroencephalographic measurements [42], [43]. KSS scores are presented next, in Table 2.1.

Table 2.1 - Karolinska Sleepiness Scale.

Level	Sleepiness description
1	Extremely alert
2	Very alert
3	Alert
4	Rather alert
5	Neither alert nor sleepy
6	Some signs of sleepiness
7	Sleepy, but no difficulty remaining awake
8	Sleepy, some effort to keep alert
9	Extremely sleepy, fighting sleep

The Stanford Sleepiness Scale (SSS) is similar to the KSS, it is a Likert scale, having seven states [40], [44]. On the original test, typically, subjects were asked to rate their alertness level every 2 hours throughout the day by choosing a single number associated with specific alertness description.

Visual Analogue Scales (VAS) [8], [45] rating is performed by asking subjects to rate their "sleepiness" using a scale spread along a 100 mm wide line. Suggestions for sleep deprivation state range from "just about asleep" (left end) to "as wide awake as I can be" (right end). Subjects place a mark on the line expressing how sleepy they feel they are. Sleepiness level is measured by the distance in millimeters from one end of the scale to the mark placed on the line. The VAS is convenient since it can be rapidly administered as well as easily repeated.

The Epworth Sleepiness Scale (ESS) [46] measures a person's general level of daytime sleepiness, or their average sleep propensity in daily life. It is a simple questionnaire based on retrospective reports of the likelihood of dozing off or falling asleep in a variety of different situations.

Other tests like the Maintenance of Wakefulness Test (MWT) and Multiple Sleep Latency Test (MSLT) may also be used to describe sleepiness levels [8].

A major drawback of this measurement type is the difficulty of implementation in real world driving conditions due to their subjective nature. Asking a driver to rate their arousal level may stimulate alertness, thus biasing the ratings. Variations in self-rated drowsiness can also be caused by stress or the use of drug substances [39]. In addition, the average estimates have to be interpreted with caution, as there are considerable individual differences in the relationship between subjective ratings [40].

2.2.2 Vehicle-based measures

Vehicle-based measures, as the name suggests, are obtained by monitoring the vehicle, that is, a correlation between the driver's alertness state and a series of vehicle parameters is made [47]. The two most commonly used vehicle-based measures for driver drowsiness

detection are: the steering wheel movement (SWM) and the standard deviation of lane position (SDLP) [48]. Other measures include monitor use of pedals (gas, break), usage of in-vehicle devices (switching radio, air conditioning) and cruising speed [15].

Steering wheel movement is measured using a steering angle sensor mounted on the steering column [10]. When drowsy, the number of micro-corrections on the steering wheel reduces compared to normal driving [11], [49]. To eliminate the effect of lane changes, the researchers consider only small steering wheel movements (between 0.5° and 5°), which are needed to adjust the lateral position within the lane [50].

Standard deviation of lane position is an established indicator of sleepiness related performance [12], [51]. SDLP consists in monitoring the car's relative position within its lane with an externally-mounted camera. Specialized software is used to analyze the data acquired by the camera and compute the car's position relative to the road's middle lane [12]. Typically, a drowsy driver, or a driver who has fallen asleep will tend to leave the designated lane and crossing into a lane of opposing traffic or going off the road.

However, vehicle-based measures have a set of limitations. They tend to be too dependent on the geometric characteristics of the road (road marking, climatic and lighting conditions), and to a lesser extent the kinetic characteristics of the vehicle [10]. These measures can also be influenced by other driver's states that not drowsiness, as cognitive distraction and visual distraction (texting task) [52].

Also, every driver has a different style of driving, there are drivers who can drive calmly because their long experience of driving and there are drivers who are still amateur and often drive reckless.

Regardless, all these disadvantages could possibly be overcome by developing robust enough strategies to adapt to the behavior of the driver and car type in order to avoid false alarms [19]. The biggest constraint, is possibly, that 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 will no longer be a feasible detection method [53].

2.2.3 Behavioral measures

Behavioral measures consist in detecting driver visual features using a camera [19], [20]. Source of visual information can include facial expression, eye movement and head movement.

Figure 2.2, adapted from [16], depicts the main behavioral measures.

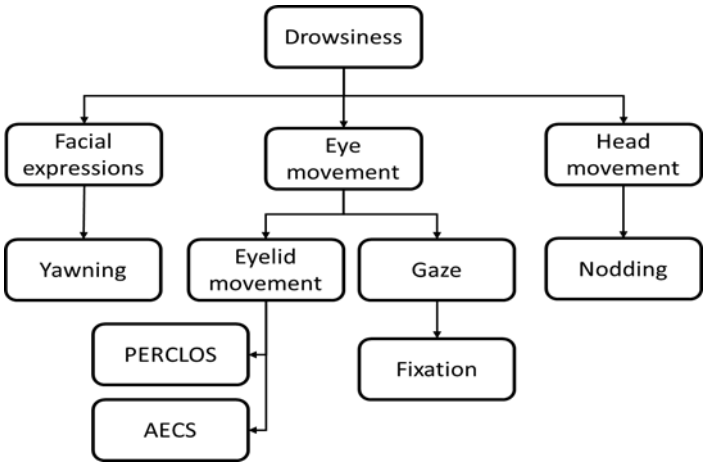

Figure 2.2 - Scheme of most used behavioral measures.

Regarding facial expression features, several have been proposed in order to detect drowsiness, as wrinkles, chin, nose, lip, nasolabial fold, lid tightener and yawn [20], [19], [54], [55], [56], [57]. For example, yawn, is a common fatigue symptom and happens to every human being, furthermore, it is relatively easy to detect (for example, frequency of yawns), as such, yawn is one of the most used fatigue indicators from facial expression related features [58], [59].

Eye movement can be described by the eyelid and the gaze. Eyelid movement is often characterized by one measure, the Percentage of Eyelid Closure (PERCLOS).

PERCLOS is a robust feature used in drowsiness detection and one of the most used [19], [59], [60], [61], [62], [63], [64], [65]. It is obtained by the proportion of time that the subject's eyes are closed over a specific period of time, allowing a correlation between drowsiness and frequency of blinking (Figure 2.3, adapted from [66]). In subjects experiencing drowsiness, PERCLOS measurements will be higher than in alert drivers, as the eyes are closed for longer periods of time and more often than in alert drivers [67].

Figure 2.3 - Example of different states of eye opening.

As mentioned PERCLOS is a strong drowsiness measure, however in most cases, proper lighting conditions are needed and use of glasses and sunglasses limit this feature accuracy. Also, it might happen that a driver who is trying to stay awake is able to fall asleep with his eyes open [20].

Examples of other eyelid related drowsiness measures are blink frequency and eye closure/opening speed, as Average Eye Closure Speed (AECS) [30], [68]. Eye closure/opening speed represents the amount of time needed to fully close the eyes or to fully open the eyes. A drowsy person will blink distinctly slower than an alert person [69].

The movement of a person's pupil (gaze) may have the potential to indicate one's intention and mental condition. For example, for a driver, the nominal gaze is frontal. Looking at other

directions for an extended period of time may indicate fatigue or inattention. In addition, when people are drowsy, their visual awareness cannot cover a wide enough area, concentrating on one direction. Hence, gaze (deliberate fixation) and saccade eye movement may contain information about the one's level of alertness [64], [70], [71], [72].

When a driver is sleepy, some of the muscles in the body begin to relax, leading to nodding. For example, when drowsy, a person´s head tend to lean forward due to its weight. However, then, it becomes difficult to breath and the head returns to its normal state, which results in nodding [71].

Head pose, estimation of the position and orientation of a driver's head, is a strong indicator of a driver's field of view and current focus of attention. It is intrinsically linked to visual gaze estimation. Intuitively, it might seem that looking at the driver's eyes might provide a better estimate of gaze direction, but in the case of lane-change intent prediction, for example, head dynamics were shown to be a more reliable cue [73], [74].

Usually, the framework for computing behavioral measures, from the captured drivers video, includes using Viola–Jones algorithm, Connected Component Analysis, Cascade of Classifiers, Hough Transform or Gabor Filter to detect the face, eye or mouth [10], [63], [75], [76]. Features are then extracted using a feature extraction technique, most used techniques are geometric constraints, Wavelet Packet Decomposition (WPD), Gabor Wavelets (GW) or Discrete Wavelet Transform (DWT). Recently, some approaches using Convolutional Neural Networks (CNNs) to extract drowsiness-related features from the face region have also been developed as in [77] and [78] works.

Having the interest features, some projects classify drowsiness state based on threshold levels, while other employ more elaborated classifiers, such as Support Vector Machine (SVM), fuzzy classifier, Bayesian Networks (BN), neural classifiers and Linear Discriminant Analysis (LDA) [10], [16], [58], [59], [60], [63], [64], [75], [79], [80].

Table 2.2, adapted from [10], summarizes some of the works using behavioral methods.

Table 2.2 - List of previous works on driver drowsiness detection using behavioral measures.

Source	Drowsiness measure	Detection techniques	Feature Extraction	Classification
[80]	Pupil, Head pose estimation	Viola-Jones Algorithm	Geometric constraints, Lucas–Kanade optical flow method	SVM
[63]	Blink duration	Viola-Jones algorithm, Neural network-based eye detector	Eye Horizontal Symmetry	Threshold based
[65]	PERCLOS, Head pose	Viola-Jones Algorithm	Active Appearance Model	Threshold based

Source	Drowsiness measure	Detection techniques	Feature Extraction	Classification
[16]	PERCLOS, AECS, Gaze, Facial Expression, Head pose	SVM classifier (eye detection) Kalman filter	Gabor wavelet Generalized regression neural networks	BN
[81]	PERCLOS, AECS, Mouth opening	Viola-Jones Algorithm	Correlation coefficient template matching	SVM
[82]	Head pose, Yawning	Geometric method	Steerable filters, Histogram of oriented gradients (HOG), Haar features	SVM
[75]	PERCLOS	SVM classifier Harr features and Adaboost classifier	HOG, Maximum-Likelihood algorithm Spectral Regression	Threshold based
[76]	PERCLOS	Haar-like features and Kalman filter	Local binary pattern (LBP)	SVM
[83]	Eye closure duration, Frequency of eye closure	Hough Transform	DWT	Neural Classifier
[84]	PERCLOS	Haar Algorithm	Kalman filter	SVM
[85]	Yawning	Viola-Jones Algorithm	Viola-Jones Algorithm	SVM

As stated, proper lightning conditions affect visual features detection performance, being challenging to get visual data which is independent from all types of illumination, for example, the contrast between day and night luminosity conditions [19]. Normal cameras do not perform well at night [10], [69] and, in order to overcome this limitation, some researchers have tried active illumination using an infrared (IR) Light Emitting Diode (LED) [10], [69]. Although working

fairly well at night, LEDs are considered less robust during the day [10], [86]. The two most popular technologies for imaging sensors used in visible light cameras are CCD (Charge - Coupled Device) and CMOS (Complementary Metal - Oxide Semiconductor). Typically, CCD based cameras obtain images with higher resolution and lower noise level, on the other hand, CMOS based cameras are cheaper and more battery efficient [8]. Some works have also used depth cameras, for example, the Microsoft Kinect which captures 3D data combining a CMOS sensor with an IR laser to obtain image depth [87].

One of the major benefits of the visual measures is that they can be acquired non-intrusively [72].

2.2.4 Physiological measures

Physiological methods offer an objective, precise way to measure sleepiness. They are based upon the fact that physiological signals start to change in earlier stages of drowsiness, which could allow a potential driver drowsiness detection system some extra time to alert a drowsy driver and, thereby, preventing road accidents [8]. Electrocardiogram (ECG), electroencephalogram (EEG), electrooculogram (EOG), photoplethysmogram (PPG) and electrodermal activity (EDA) are physiological signals from where we can extract measures to detect drowsiness [20].

The ECG records the electric activity of the heart. This activity, more precisely the heart rate, is controlled by the balance between the two branches of the autonomic nervous system (ANS), the sympathetic nervous system and the parasympathetic nervous system. Wakefulness states are characterized by an increase in sympathetic activity and/or a decrease of para-sympathetic activity, while a decrease in mental workload/ a relaxation state, which can occur in sleepy drivers over prolonged monotonous driving, is characterized by an increase in parasympathetic activity and/or a decrease in sympathetic activity [20], [88], [89], [90].

ANS activity can be indirectly measured by the heart rate variability (HRV), a measure of the beat-to-beat (R-R Intervals, Figure 2.4, adapted from [91]). For example, an increased HRV indicates a higher parasympathetic activity, as the HR lowers resulting in more time between heartbeats, thus more variability.

Conventional approaches to analyze HRV include time and frequency domain methods.

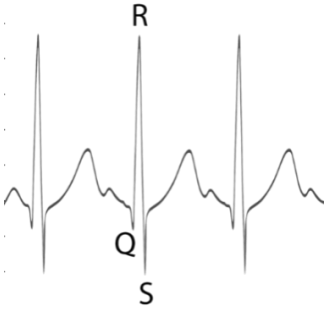

Figure 2.4 - ECG waveform, with QRS complex identification example. The three waves of the QRS complex represent ventricular depolarization.

Typical time domain HRV measures are the mean RR intervals, standard deviation of RR intervals (SDNN), root mean square of the differences between consecutive RR intervals (RMSSD), frequency of successive differences of RR intervals that spanned more than 50 ms (NN50) and percentage value of NN50 (pNN50) [89], [92], [90].

Frequency domain analysis can be performed using power spectral density (PSD) of the RR interval series. RR series may be classified in three frequency bands: very low frequency (VLF) (0-0.04 Hz), low frequency (LF) (0.04-0.15 Hz), and high frequency (HF) (0.15-0.4 Hz) [89]. For example, LF and HF are reliable measures, respectively, of sympathetic and parasympathetic activity, thus when a driver progresses from an awake to a drowsy state there is a decrease in the LF/HF ratio [10], [88], [93], [94], [95].

To obtain these measures, first a pre-processing step needs to be applied to the ECG signal, as band pass filters, followed by algorithms for R-peak detection like Pan-Tompkins algorithm [96]. For the frequency domain analysis the signal domain conversion is done typically with Fast Fourier Transform (FFT) [97].

Even though the reliability and accuracy in detecting driver's drowsiness based on ECG is high when compared to other methods, an important limitation of physiological signal measurement is its intrusive nature [19]. One possible way to overcome this problem is trying to use a non-intrusive system, however, it is less accurate compared to intrusive systems due to improper electrode contact [20].

Electroencephalogram (EEG) records electrical activity of the brain providing information represented by waves (frequency bands). It is a standard technique used in sleep studies [98].

The most relevant waves for drowsiness and sleepiness experiments are: alpha waves (8 – 12 Hz), they play an important part in the transition from wakefulness to sleep, although they are not directly related to sleepiness itself, but to 'relaxed wakefulness', which leads to a reduced readiness to react to stimuli; beta waves (13 – 30 Hz) are associated with increased alertness, arousal and excitement; theta waves (4 – 7 Hz) are commonly regarded as a clear indicator of lack of attention and the onset of sleep and delta waves (< 4 Hz) which originate during deep sleep [15], [98].

A schematization of the described waves is depicted next, in Figure 2.5, adapted from [99].

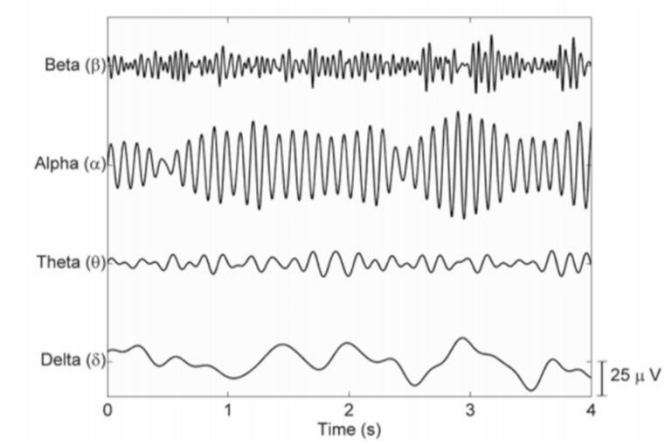

Figure 2.5 – Representation of the main EEG signal rhythms used from drowsiness detection.

Variations in the EEG signal, such as increases in alpha and theta rhythms and reduction of beta waves, are interpreted as an indication of weariness and sleepiness [100]. To assess these changes, typically FFT or Wavelet Transform (WT) [101] are applied to the signal. In the same way as the ECG and other physiologic signals, to extract features from the signal, first, a pre-processing step is required, as applying a band pass filter [102], [103].

Although the EEG signal in most cases discriminates accurately between alert and drowsy states, there are occasions where the states are not clearly differentiable, being difficult to distinguish them. For example, during the transition from wakefulness to sleep, there may be several minutes where the EEG looks like that of wakefulness, but awareness of the environment is already lost [104].

EEG signal acquisition also requires very specific conditions for being measured properly (intrusive methods). In a real-world driving scenario, the required acquisition system is of huge inconvenience, as it can hinder driving capabilities and potentially increase the chances of an accident happening [8].

The Electrooculogram (EOG) is a method used for measuring the potential difference between the front and back of the eye ball. The eye is a dipole with the positive cornea in the front and the negative retina in the back, this electric potential between cornea and retina lies in the range 0.4 - 1.0 mV. When the eyes are fixated straight ahead a steady baseline potential is measured by electrodes placed around the eyes. When moving the eyes a change in potential is detected as the poles come closer or farther away from the electrodes. The sign of the change depends on the direction of the movement [10], [72]. The EOG can therefore be used for detection of eye movements and blinks. In sleepiness studies, for instance, it can be used to differentiate rapid eye movements and non-rapid eye movements which are characteristic of an awake and drowsy state respectively [100].

Photoplethysmogram (PPG) is an optical obtained signal which measures volumetric changes in blood in peripheral circulation. It can be acquired by using a pulse oximeter that makes uses of low-intensity IR light which is proportional to the quantity of blood flowing through the blood vessels [9], [105]. PPG can be used to estimate, for example, heart and respiration rate.

The Electrodermal activity refers to the phenomena of skin conductance. Specifically, it is based on the fact that the electrical characteristics of the skin (skin conductivity) mainly change due to the level of sweat produced by sweat glands. In turn, the level of activity of the sweat glands directly reflects the mental state of the person [106]. Although EDA can be influenced by multiple other factors that not drowsiness, for example, by stress, it has already been explored in some driver drowsiness related studies as a complement to other physiologic signals [107], [108].

As in behavioral methods, more recently, some deep learning approaches, for instance, using CNNs, have also been used to efficiently extract features from the physiologic signals [109], [110]. However, most of the developed approaches extract features using the previously described methods, which are then used to evaluate the driver state applying a model as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Linear Discriminant Analysis (LDA), Decision Trees or K-Nearest Neighbors algorithm (KNN).

Table 2.3, adapted from [10], [92], reviews some of the approaches used to detect driver drowsiness based on physiologic measures.

Table 2.3 - List of previous works on driver drowsiness detection using physiological measures.

Source	Signal	Preprocessing	Feature Extraction	Classification
[90]	ECG	Band-pass filter	FFT	Neural Network
[88]	ECG	Signal qualification block based on standard deviation of ECG amplitude	Frequency domain HRV features	LDA
[101]	EEG	Band-pass filter	DWT	K-means clustering
[102]	EEG	Band-pass filter	FFT	Fuzzy logic system
[103]	EEG	Band-pass filter	FFT	SVM
[93]	ECG and respiration	Low-pass Butterworth filter, smoothing filters	PSD using FFT for frequency HRV and respiration features	SVM, Decision Trees, Naive Bayes
[94]	ECG, EEG and respiration	Averaging and low-pass filters	PSD for frequency HRV, skin conductivity and respiration features	LDA, Naive Bayes
[111]	ECG, EEG, EOG and respiration	Second order Butterworth low-pass filter	Breathing variability features	Threshold based
[112]	ECG, EEG and EOG	None	Frequency domain HRV features (FFT)	BN
[113]	EOG and EMG	Low-pass filter and visual inspection	DWT	ANN
[114]	EOG, EEG and EMG	High-pass filter and thresholding	Neighborhood search	SVM
[115]	EEG	Independent component analysis decomposition	Power Spectrum Analysis using FFT	Self-organizing Neural Fuzzy Inference Network
[104]	EEG and EMG	Band-pass filter and visual inspection	DWT	ANN
[116]	EEG	Low-pass filter	FFT	Mahalanobis distance

Source	Signal	Preprocessing	Feature Extraction	Classification
[117]	EEG	Band-pass filter	FFT	Linear Regression Model
[95]	EEG and ECG	None	Mean power frequency EEG features and frequency domain HRV features (PSD) using FFT	Threshold based

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. For example, while EEG is excellent to distinguish between awake and sleep states, there might be some limitations when identifying the transitions state. Heart rate measures, as stated, generally have differences between alert and drowsy states, nonetheless heart rate is also affected by other factors.

2.2.5 Hybrid methods

As described over the previous sections, each measurement method has its strengths and weaknesses. Subjective measures are important as a support or ground truth but are not feasible for a real time monitoring system.

Vehicle-based measures provide satisfactory results, however, they tend to be too dependent on the geometric characteristics of the road (road marking, climatic and lighting conditions), and to a lesser extent the kinetic characteristics of the vehicle [10]. These measures can also be influenced by other driver states that not drowsiness as visual distraction (texting task).

Behavioral measures, although non-intrusive and easy to use, still have some drawbacks that must be considered in visual data acquisition from driver, as illumination and items used by driver such as glasses [19].

Physiologic measures generally provide the most accurate values and the biggest limitations is their intrusive nature. Although less intrusive measurements of some signal as the ECG have already been developed, signal quality is still lost, and for other methods as EEG and EOG, electrodes still need to be placed on the scalp or eye area in an intrusive manner [20].

A summary of advantages and limitations of each method is presented in Table 2.4 (adapted from [38]).

Table 2.4 – Comparison between different drowsiness measurement types.

Measures	Advantages	Limitations
Subjective	Takes personal feeling into account	Not feasible in real time
Vehicle-based	Non-intrusive	Interpersonal accuracy Dependent on external conditions
Behavioral	Non-intrusive Ease of use	Lighting condition Background
Physiological	Earlier detection	Mostly intrusive Low signal quality in non-intrusive solutions

To obtain the best possible results, a promising option is to use several of the referred methods, creating a hybrid system from visual, physiological and driving behavior measures.

Although some works have already combined different drowsiness detection methods, there is still a big room to improvements [19]. An example of an hybrid drowsiness detection system is the work of [105] which included a system combining facial features, eye movement, and a physiologic signal, photoplethysmogram. Drowsiness analysis with driver's facial data and steering wheel data was also performed by [118]. Vehicle-based and behavioral measures are also combined in [119]. Works as [1] have combined all the measurement types, using indicators such as heart rate variability, respiration rate, head and eyelid movements (blink duration, frequency and PERCLOS), time-to-lane-crossing, speed, steering wheel angle and position on the lane.

The proposed thesis goal fits in hybrid system methods group, as it will focus on the fusion of information obtained by physiologic signals (ECG) and from behavioral measurements.

2.3 Commercial solutions

From the above mentioned measures used for driver drowsiness detection, several systems have already been developed and commercialized.

Most systems use vehicle-based measures and/or behavioral measures. For example, Bosch [120], [121] develops solutions as a steering-angle sensor that is sold to car manufacturers. In addition, from driver's steering behavior, the system algorithm evaluates approximately 70 signals received via the vehicle's Controller Area Network (CAN), among them the length of a trip, use of turn signals, and the time of day. The function calculates the driver's level of fatigue. If that level exceeds a certain value, an icon such as a coffee cup flashes on the instrument panel to warn drivers that they need a rest. This solution is used for example in Volkswagen's Passat alltrack [13].

Bosch is also developing a camera system, presented in Consumer Electronics Show (CES) 2017 in Las Vegas [17], that will allow face recognition, intelligent personalization and driver drowsiness detection.

NVidia, a chip supplier to Audi, Mercedes, Tesla and others also presented at CES 2017, an artificial-intelligence tool (the Co-Pilot) that can learn the behaviors of individual drivers and determine when they are operating outside their norms [18]. The system will evaluate driver's standard posture, head position, eye-blink rate, facial expression and steering style, among other indexes. Based on a vehicle's capabilities, the driver will be warned or automatically driven to a safe spot when conditions warrant.

Valeo, a French supplier of automotive technology, also developed a camera [122], [123] that analyses eyelid movement, pupil position, head angle and other key features processing them in real time by an embedded computer. This data is then combined with vehicle trajectory and tracking information to determine how alert or how distracted the driver is, sending a warning if necessary.

MERCEDES Attention Assist™ [14] checks as many as 90 indexes, such as steering wheel angle and lane deviation and external factors as weather and road surface influences, for example, pothole avoidance that may be causing the irregular driving behavior. If the system determines drowsiness is the cause, it will send an audible and visible alert letting the driver know it's time to take a break.

Driver Alert Control (DAC) by Volvo is a camera-based vehicle system that detects ideal road trajectory and compares it to steering wheel movements. The system provides sound alert and visual notification on instrument cluster (Figure 2.6 from [124]) when drowsiness is detected. Notification is repeated if driver behavior does not improve. The system is designed to function on highways and is activated when the speed exceeds 65 km/h [15], [125].

Figure 2.6 - Driver Alert Control (DAC) warning signal.

The brasillian division of Ford recently proposed the Safe Cap [126], made with special consideration for truck drivers. The Safe Cap includes an accelerometer and a gyroscope that measure drivers head movement. If drowsiness is detected light, sound and vibration warnings are activated to alert the driver.

Some companies have also developed drowsiness solutions based on driver physiological methods.

For example Plessey [22] uses a capacitive sensor that enables ECG signal acquisition without the need of skin contact (WARDEN™ system, Figure 2.7, adapted from [22]) .The system uses an array of sensors to detect changes in electric potential in the human body without

direct skin contact and returns an accurate R peak signal from the users ECG, this in turn can be used to calculate already mention features as heart rate variance (HRV). The WARDEN™ system can be easily and discretely incorporated inside vehicle seatbacks to access the necessary biometric signals and provide earlier warning of drowsiness comparatively to non-physiological methods.

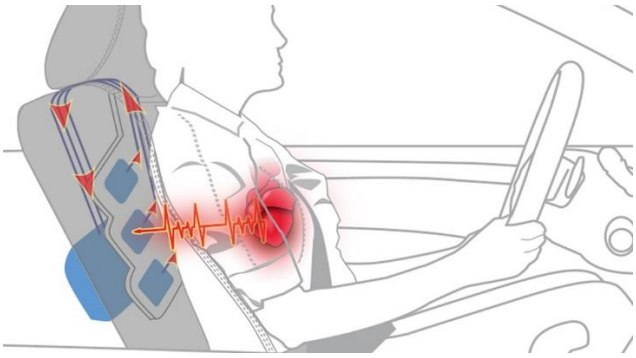

Figure 2.7 - WARDEN™ capacitive system, capable of acquiring ECG without skin contact.

Harken [127], a public-private European consortium, is developing a fatigue detector by monitoring cardiac and respiratory rhythms, using an embedded, nonintrusive, acquisition system in the seat cover and safety belt of a car.

StopSleep UK [23] developed an anti-sleep ring, which using cutaneous sensors measures electrodermal activity, evaluating fatigue and alertness levels. As soon as levels of concentration start to decrease, StopSleep will send an alert to the driver (vibration and sound warnings).

Wearvigo [128], produces a wearable, a smart headset, that tracks over 20 parameters, such as blinks related measures, blink rates, blink durations and drooping eyelids, by analyzing biosignal patterns. In addition, with an accelerometer and gyroscope, movements of the head are also measured, as head nods.

CardioID Technologies, a Portuguese company, developed the cardiowheel (Figure 2.8) [129]. Cardiowheel is an Advanced Driver Assistance System that acquires the ECG from the driver's hands to continuously detect drowsiness. This is done using a custom steering wheel cover equipped with special conductive elements, thus capturing the electrical impulses generated by the heart. In addition, from the measured ECG, it is possible to monitor the heart activity detecting multiple cardiac pathologies. Driver's identity can also be extracted from the ECG signal.

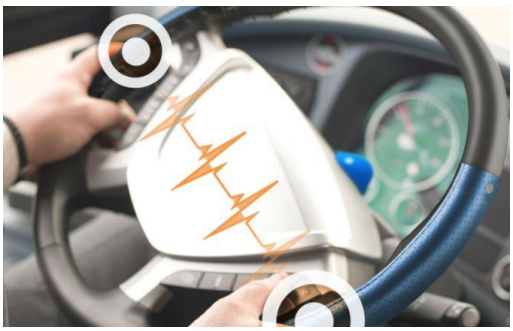

Figure 2.8 – Cardiowheel system developed by CardioID Technologies.

2.4 Experimental setups

Acquiring data to study driver drowsiness is a complex process. Pragmatically, this process can be performed in essentially four types of experimental setups: driving simulators, test tracks, road studies and naturalistic driving studies [130], each of whom presents advantages and disadvantages. A thorough description of the introduced topics is presented next.

2.4.1 Platforms and settings

Driving simulators are middle to high fidelity simulators that provide the driver with an at least somewhat genuine feeling of sitting in a real car. The environment is computer generated, and it is possible to log a host of variables [130].

Tongji advanced driving simulator (Figure 2.9, adapted from [131]), is an example of one the most advanced simulators. It simulates a fully instrumented Renault 23 Megane III vehicle cab in a dome mounted on an 8 degree-of-freedom motion system 24 with an X-Y range of 20 × 5 meters. An immersive 5 projector system provides a front 25 image view.

Figure 2.9 - Tongji advanced driving simulator.

The Swedish National Road and Transport Research Institute (VTI) also has several advanced driving simulators, being the newest the Sim IV. This simulator also has high realistic simulation capacities providing the driver a 210-degree forward field of vision, both longitudinal and lateral acceleration and can simulate a passenger car and a truck compartments.

The NADS-1 advanced driving simulator from university of Iowa [132] and the TruckSIM from University of Leads driving simulator [133] are examples of other equally high realistic simulators.

Test track studies are performed in instrumented vehicles on a controlled environment, closed to public traffic, which can easily be adjusted to the desired needs [130]. Examples of drowsiness and also distraction related studies performed on test track include [134] and [135] works.

Similarly to a test track study, an on-road study is often quite limited in time, the driven routes are pre-determined, and there may be an experimenter in the car. However, the study is conducted in real traffic [130].

For example, in the study performed by [136] fatigue level was evaluated in an expressway from Beijing to Qinhuangdao analyzing driver behavior and vehicle based measures.

A large field study is the SleepEYE project, a collaborative project between Smart Eye, Volvo Cars and VTI which recorded vehicle, behavior and physiologic based measures. The project included 18 participants which conducted one alert and one sleepy driving session on a motorway [30].

Naturalistic data collection is also conducted in real traffic, but no experimenter is present in the car, and the studies are usually long-term, lasting for a month or more. The drivers have free choice of route and use the vehicles for their daily lives.

A typical example of a naturalistic data study is the 100-Car Naturalistic Driving Study [137]. This study tracked 100 drivers across a period of one year, collecting data for each vehicle with a highly capable instrumentation system that included five channels of video and vehicle kinematics. The collected data includes approximately 2 000 000 vehicle miles and almost 43 000 hours of data containing many extreme cases of driving behavior and performance, as severe fatigue, impairment, judgment error, risk taking, willingness to engage in secondary tasks, aggressive driving and traffic violations.

Australian Naturalistic Driving Study also collected data from multiple driver and vehicle-based parameters aiming to understand what people do when driving their cars in normal and safety-critical situations [138]. Likewise, the strategic highway research program, whose main objective is understanding driver performance and behavior in traffic safety, has retrieved roadway data that is been used to study issues as drivers inattention, rear-end crashes and safety on curves [139].

Car companies as Mercedes-Benz also tested their vehicle based fatigue detection system (Attention Assist system) in real work driving conditions, driving over two million kilometers and using more than one thousand drivers [140]. Ford in 2011 [141], made a dataset, obtained from naturalistic studies but also simulators, which was provided during a competition to develop a classifier to detect driver alertness.

2.4.2 Benefits and limitations

Each discussed experimental setup implies some advantages but also limitations. In particular, two key factors that must be considered are the setup validity and the degree of control [130].

The setup validity (ecological validity), refers to a question of generalization, that is, the possibility to generalize the results to be valid for situations in real life. Having a high degree of ecological validity implies to have high terms of realism of the setting for the driver, taking

into account aspects as obtrusive/unobtrusive instrumentation, the scenarios realism and how well the collected data correspond to the research question asked.

In most scientific work it is of great importance to have a high degree of control of the experimental setting in order to reduce the risk for confounded results. The control is not only related to the scenario used for driving, but also to the selection of the participants, their preparation before the study and how they are treated during and after the study.

Control is often necessary for reliable repeatability within a reasonable time frame. With a high degree of control unusual scenarios, like a moose crossing the street, can be tested within a short time frame, but also rather more frequent scenarios, like a lead vehicle braking, can be reproduced any number of times under completely similar circumstances.

Within traffic research, control is also crucial in order to minimize the consequences of dangerous events. In an uncontrolled setting a driver who falls asleep can kill both himself and others, while in a controlled study nobody will be harmed. For ethical reasons many studies can only be conducted when a high level of control over the possible consequences is guaranteed, and even though the external validity might be reduced, the only other alternative would be not to conduct the study at all.

A high degree of external validity and a high degree of control over the study are often difficult to reconcile with each other. The relation between ecological validity and study control for the previous described experimental setup types, can be organized according to the following scheme, Figure 2.10, adapted from [130].

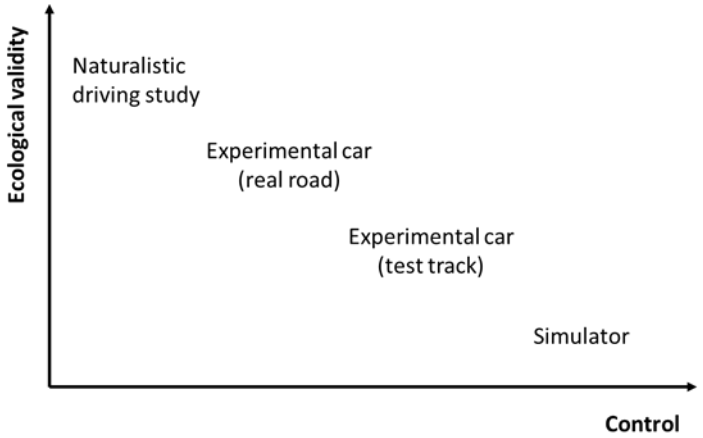

Figure 2.10 – Comparison of the relation between ecological validity and control for the different experimental setups.

A setup as the simulator has a high degree of control, it is possible to control test conditions, dissuading and stimulating drowsiness. Reducing the need to change lanes and gear, ask drivers to drive at constant speed and not introducing environmental disturbances as cross winds are all examples of factors that stimulate drowsiness. However, in opposition, there is a lower degree of external validity (the drivers' behavior will not be the same under simulated driving since there is not any risk involved).

An experiment on a test track or on a road with an experimental vehicle will still provide a high degree of control of the participants, but a decrease in the control of events like interactions with other road users, animals, but also of the weather and road constructions.

Naturalistic driving studies will have a low degree of control. For example, it is not possible to use electrodes to acquire physiologic signals, to use a subjective measure as the KSS and

there is no prior information about the participants state, as the number of hours slept. By contrast it has a high degree of external validity, as it depicts real life situations.

Regardless of the experimental setting the quality of the results is highly dependent on the data quality.

It is also worth mention that, although large databases have already been created acquiring data either in simulated or real world driving conditions, most of them are private.

2.5 Performance assessment

An important step in any research is to evaluate the system performance. In drowsiness related studies, in most cases, a drowsiness label is set, that is, a ground truth measure that is then used to compare with the developed system predictions, thus evaluating the performance.

However, as it was already briefly discussed, defining an accurate drowsiness ground truth label it is not a straightforward process and there is no uniformly agreed measure [142]. Some labels are obtained based on the time of drive, others according to the occurrences of crashes, namely in simulator studies, but the most common approach is obtaining the labels based in subjective ratings, as the previous discussed in section 2.2.1.

Generally, labels based in subjective measures, are obtained either with the driver self-reporting [143] or by a supervisor asking the driver his status [70]. Other strategies include experts review and rating fatigue in video segments [80]. The most used subjective measure is the KSS, previously mentioned, being used in several works as a drowsiness state label [19], [40], [60], [70], [94], [144], [145]. Other subjective scales like SSS and VAS are also used [40].

When evaluating driver drowsiness, some researches classify driver sleepiness state as a binary problem [70], [81], [146], where the driver is either alert or drowsy, while others make a three level incremental division [62], [114], into alert, drowsy and very drowsy states.

Predicted results are then compared against given labels using several measures. A prevailing measure is the accuracy [62], [64], [146], [114], [147], [148]. Accuracy expresses the percentage of samples correctly classified.

To estimate the accuracy, the proportion of true positive and true negative in all evaluated cases needs to be calculated. For a binary drowsy driving problem, true positive (TP) 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. Table 2.5, summarizes this explanation in a binary confusion matrix.

Table 2.5 - Confusion matrix of a binary alert/drowsy problem.

		Predicted	
		Alert	Drowsy
Given	Alert	TN	FP
	Drowsy	FN	TP

Mathematically, accuracy can then be stated as:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \times 100\ \% \tag{2.1}$$

And for a three class problem [92],

$$\text{Accuracy} = \frac{\sum_{i=1}^{3} M_{i,i}}{\sum_{i=1}^{3}\sum_{j=1}^{3} M_{i,j}} \times 100\ \% \tag{2.2}$$

Where M represents the confusion matrix $M = [M\ i,j]_{3x3}$. More precisely, $M_{i,j}$ is the number of samples predicted as i with true label j. Each label number (1,2,3) represents, respectively, for both i and j, the classes awake, medium and drowsy.

Other measures also used include precision, specificity, sensitivity, F1-score and mean square error [81], [92], [149], [150], [151]. Precision measures the proportion of correctly classified drowsy states of all the drowsy states classification.

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\ \% \tag{2.3}$$

Specificity measures the proportion of alert states which are correctly identified as alert, that is, the classifiers ability to correctly identify alert states.

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100\ \% \tag{2.4}$$

Sensitivity, or recall rate, measures the proportion of drowsy states that are correctly classified as drowsy, that is, the classifiers ability to correctly identify drowsy states.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100\ \% \tag{2.5}$$

F1-score is a measure of a test's accuracy. It is a weighted average of the precision and sensitivity.

$$\mathrm{F_1} = 2 \times \frac{Precision \times \text{Sensitivity}}{Precision + \text{Sensitivity}} \tag{2.6}$$

The mean square error (MSE) is a measure of the quality of an estimator, it measures the average of the squares of the errors.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i^{predicted} - Y_i^{real})^2 \tag{2.7}$$

Where n is the total number of samples and Y_i is the label (predicted or real) of sample i.

Chapter 3

A framework for driver drowsiness monitoring

This chapter focuses in presenting a set of machine learning methods used for classifying the driver state. Following the overall method overview, there is a detailed description of each of the applied steps and lastly a report and discussion of the obtained results.

All the performed experiments of this chapter are based on the measures obtained from the SleepEye dataset [30].

3.1 Methods overview

The overall workflow used in this section follows an approach similar to [92] work. Using the same dataset here used, the cited work focused on evaluating driver drowsiness using physiologic signals, considered a robust method to the driver drowsiness evaluation problem. As so, this work replicates the mentioned workflow, first using physiologic signals, particularly ECG, obtaining a baseline for performance comparison, and then developing a new approach using video signal. Thus, assessing the performance of the video signal, acquired non-intrusively, both alone and in a hybrid approach with the ECG signal, a signal that has the potential to be acquired non-intrusively, to evaluate driver drowsiness.

This approach corresponds to a standard supervised machine learning classification problem workflow. Classification implies that a categorical description of the problem is performed and supervised learning that the algorithms are trained using labeled data.

Specifically, a binary and a three level multiclass problem divisions were considered, consistent with [92] approach to the problem. 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 were obtained from the KSS provided with the dataset and according to the following scheme, Figure 3.1, as in [92] work.

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

Figure 3.1 – KSS rating and their corresponding states for 2-class and 3-class problems.

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 in the test dataset using a performance measure, in particular, accuracy values and confusion matrixes, for a more detailed view, previously described in section 2.5 are used. For a better generalization of the results, the entire process is repeated 10 times (10-fold validation).

In order to improve the performance, different strategies were tested as classifier combination methods, data imbalance and sequential data approaches. Throughout this chapter, a more detailed explanation of the described steps is presented.

3.2 Feature engineering

In the machine learning field, the process of deriving values from a raw input, for example an ECG signal, is defined as feature extraction. Those values, the features, intend to be informative, characterizing the dataset in a non-redundant way, thus, facilitating the learning process. Extracting meaningful features from the information sources is an essential step of any machine learning framework, which may involve further refinement, applying strategies as feature selection and transformation. Bellow, the extracted features and respective acquisition strategies are reviewed.

3.2.1 Feature extraction

As previously indicated, all the measures used in this experiment are obtained from the SleepEye dataset. This dataset contains physiologic data, ECG and EOG signals, and behavioral data, obtained by a SmartEye Pro 5.7 system [152], acquired in an experimental field study in real traffic on real roads. As already implied, a KSS label characterizing each driver drowsiness level is also available in the dataset.

The features were extracted from the ECG and video signals in non-overlapping 2 minutes windows, the same window length used in [92] work.

Regarding the ECG signals, a preprocessing similar to the described in [92] work was applied. A modified version [153], [154] of the Pan-Tompkins QRS detection algorithm [96] was used for R peak detection. As an example, Figure 3.2 illustrates detected R peaks for a 2 minute sample of an ECG signal.

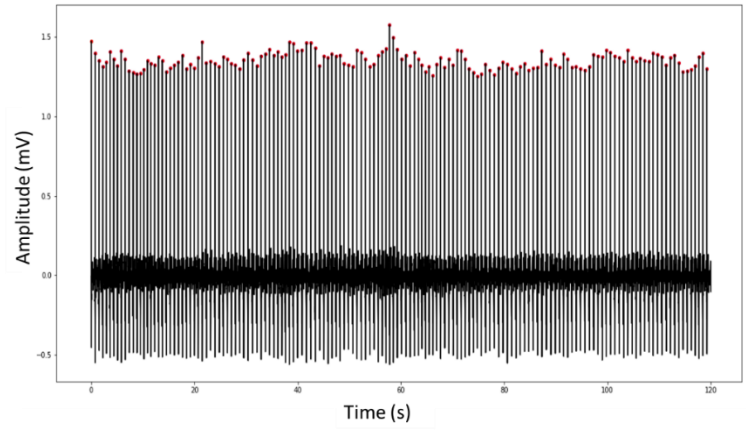

Figure 3.2 - ECG signal with detected R peaks marked in red circles.

A detailed view of the Figure 3.2 ECG signal, is also shown next, in Figure 3.3.

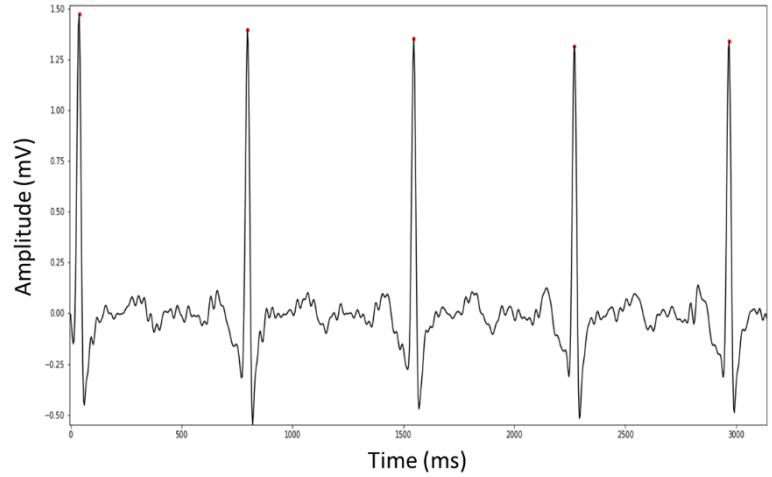

Figure 3.3 – Detailed view of detect R peaks in ECG signal.

After R peak detection, time and frequency domain features were computed. Those features are summarized next, in Table 3.1.

Table 3.1 - Features extracted from the ECG.

Nr	Feature	Description
1	HR	Heart rate
2	SDNN	Standard deviation of NN intervals (HRV_i)
3	SDSD	Standard deviation of differences between adjacent NN intervals ($HRV_i - HRV_{i-1}$)
4	RMSSD	Square root of the mean of the sum squares of differences between adjacent NN intervals
5	NN50	Number of pairs of successive NNs that differ by more than 50 milliseconds

Nr	Feature	Description
6	pNN50	NN50 divided by total number of NNs
7	NN20	Number of pairs of successive NNs that differ by more than 20 millisecond
8	pNN20	NN20 divided by total number of NNs
9	HF	Total energy in the high frequency band (0.15 to 0.4 Hz)
10	LF	Total energy in the low frequency band (0.04 to 0.15 Hz)
11	VLF	Total energy in the very low frequency band (0.003 to 0.04 Hz)
12	TP	Total power in all bands (0.003 to 0.4 Hz)
13	HFnu	HF normalized (HF divided by the difference between TP and VLF)
14	LFnu	LF normalized (LF divided by the difference between TP and VLF)
15	VLFnu	VLF normalized (VLF divided by TP)
16	LF/HF	Ratio between LF and HF

As for the behavioral data, only already extracted measures from the SmartEye Pro system were available, namely drivers head orientation, gaze direction, eyelid opening and pupil diameter. Following the same two minutes time window approach used in ECG feature extraction, drowsiness related features were computed from the available measures. Due to its more elaborated processing, the blink and fixation analysis methods, obtained, respectively, from the eyelid opening signal and gaze direction, are summarized next.

For blink detection, a similar approach to [67] work was considered, with 4 points (t1,t2,t3,t4) characterizing a blink. To compute these points, first, the signal valleys (closed eyes) are located based on the signal first order difference and an adaptive threshold. Then, for each detected location, the signal is searched before and after the valley until it stops increasing within an established range, obtaining, respectively, the t1 and t4 point's positions. Lastly, t2 is defined by the position where the signal is approximately 60% of the difference between t1 and the valley opening value. The analog process is performed for t3 using t4 point and valley values.

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 next, in Figure 3.4.

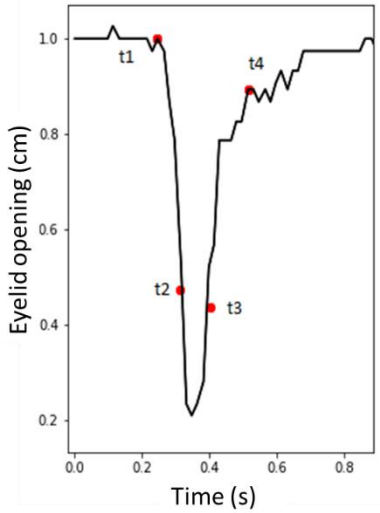

Figure 3.4 – Scheme of points (t1 to t4 points marked in red) characterizing an eye blink.

An example of the detected blinks for a 2 minute eyelid opening signal is also shown next, in, Figure 3.5.

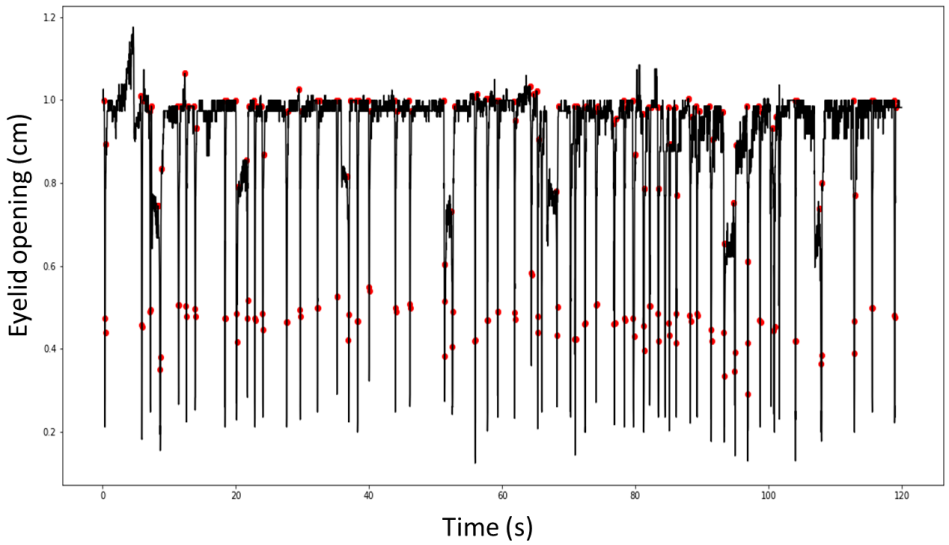

Figure 3.5 – Detected eye blinks in eyelid opening signal. Each detected blink characterized for 4 points, previously described, in red.

Regarding the fixation detection, several algorithms have already been developed either based on spatial characteristics, as velocity, dispersion and area based methods, or on temporal characteristics, as duration sensitive and locally adaptive methods [155].

In this work, a simple velocity based algorithm (Velocity-Threshold Identification) was implemented. Using this method, fixations can be identified based on two main velocity distributions that can be found on eye movements: low velocities for fixations (< 100 deg/s) and high velocities for saccade eye movements (> 300 deg/s) [155].

The complete list of the features extracted from the video measures is summarized in Table 3.2.

Table 3.2 - Features extracted from the available SmartEye measures.

Group	Nr	Feature	Description
Eyelid distance	1	Blink duration	Average blink duration
	2	Median blink duration	Median blink duration
	3	Blink frequency	Number of blinks per time interval
	4	fracBlinks15	Fraction of blinks longer than 0.15 seconds
	5	Blink duration25	Average blink duration of 25% longest blinks
	6	Blink opening/closing speed	Average of blink opening and closing speed
	7	PERCLOS	Percentage of eyelid closure over time
Pupil diameter	8	PupilD	Average pupil diameter
	9	SPupilD	Standard deviation of pupil diameter
Yaw, pitch and roll head angles	10	HP	Average head pose of combined yaw, pitch and roll head angles
	11	SHP	Standard deviation of combined head pose
	12	Nodding frequency	Number of nods per time interval
	13	Nodding duration	Average nodding duration
	14	Total nodding duration	Total nodding duration
Yaw and pitch gaze angles	15	Gaze	Average of combined yaw and pitch gaze angles
	16	SGaze	Standard deviation of combined gaze
	17	Total fixation	Total duration of fixation time

As already briefly mentioned, using the listed ECG and video features three different approaches were tested. A first approach using only the ECG features (16 features) in order to establish a baseline, comparing with the performance of [92] work. A second approach using only the video features (17 features) with the main goal of evaluating if non-invasive video

measurements could replace the EOG features, also in comparison with [92] work. The last strategy involved a multimodal procedure, combining the ECG and video features (33 features), aiming to obtain the best possible results.

3.2.2 Window size sensitivity

As denoted above, this work followed a similar approach to the [92] work in order to make valid comparisons, this included using the same data segmentation method. That is, the discretization process of the ECG and video signals into smaller segments from which the features are computed.

The applied segmentation method, as previously explained, consists of a fixed size non-overlapping window. Although multiple other segmentation methods could also be tried as the fixed-size Overlapping Sliding Window, Dynamic Windowing and Variable-size Sliding Window (VSW) [156], a key factor to evaluate using the same method is the impact of the chosen window size in the classifier accuracy.

For example, an overly wide window can lead to information overload which causes the features to be mixed up with other information, while a small window could split an activity, both leading to suboptimal information decreasing the process performance [156], [157].

In order to evaluate the window size impact, a window sensitivity analysis is performed, comparing the algorithm performances using windows of different sizes.

3.3 Classifiers

Along with the discriminative capacity of the extracted features to correctly characterize the dataset, a decisive factor determining the performance of the process is the classifier used. 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, the best approach is to evaluate different models.

In this work, five typical 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). The classifiers used are the same tested by [92], so that a valid comparison of the results could be performed, additionally the KNN classifier, already used in similar approaches, was also tested to evaluate if a different classifier could provide better results. A brief description of each classifier and respective trained hyperparameters is presented next.

3.3.1 Support Vector Machine

Support vector machine is one of the most used machine learning algorithms for driver drowsiness detection [76], [80], [81], [103], [82]. SVMs revolve around the notion of a “margin”, that is, the distance between the separating hyperplane (decision boundary) and the samples (support vectors) [158]. In SVM training, using the support vectors, the margin is maximized and, thereby, the largest possible distance between the separating hyperplane and the instances on either side of it is obtained [158], [159]. An illustration, adapted from [160], for a two features SVM example is displayed next, in Figure 3.6.

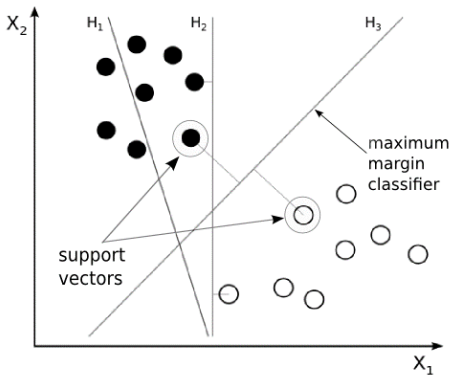

Figure 3.6 – Example of a set of hyperplanes (H_1,H_2) and respective best separating hyperplane (H_3) for a two features problem.

Generally, the performance of SVM will depend on factors as the type of kernels used. In this work, the optimized hyperparameters are: the type of kernel, the kernel coefficient and the penalty parameter.

3.3.2 Random Forest

A decision tree (DT) is a hierarchical data structure that represents data through a divide and conquer strategy. That is, a decision tree uses a branching method that will recursively split the dataset into binary partitions. The key concept of the Random Forest classifier is to grow an ensemble of trees and letting them vote for the most popular class [161].

One of the biggest problems of decision trees, is that they are prone to overfitting, particularly when many features are considered leading to a large number of splits (deep trees).

This algorithm uses techniques as bootstrap, that is, the classification trees are constructed simultaneously using random data samples from the dataset (forms new training sets independently), and random feature vectors generation, meaning that, the growth of each tree in the ensemble is based in different feature vectors, which limits the trees growth [162] [161]. With these techniques, the previously mentioned decision tree disadvantage is overcome, thus creating a much more robust model.

An example of a random forest classifier architecture is shown next, in Figure 3.7, adapted from [92], [163].

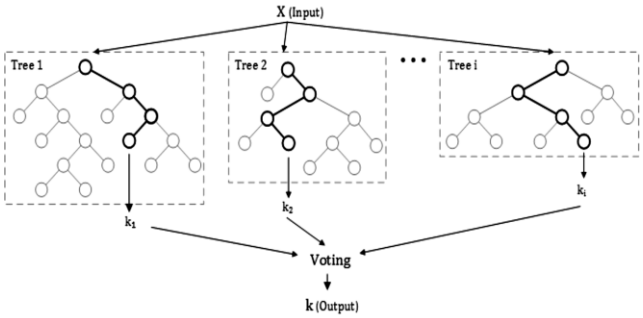

Figure 3.7 – Example of random forest classifier architecture.

In this work, the RF algorithm training was done by controlling the total number of trees.

3.3.3 Artificial Neural Networks

Artificial neural network architectures are motivated by models of our own brains and nerve cells. They are formed by a network of simple elements called neurons, interconnected and

arranged in a series of layers: the input layer (receives the features feed to the network), the output layer (the provided data classes) and, in between, the hidden layers [90].

Each neuron is composed by a set of inputs (received from some other neuron or from an external source), a weight associated to each input and an activation function (for example, a step function). The neuron translates these inputs into a single output, which is then feed to another layer of neurons. The weights as well as the functions that compute the activation are obtained during the training stage, using techniques as backpropagation [90].

An illustration of a ANN architecture is shown next in Figure 3.8, adapted from [104].

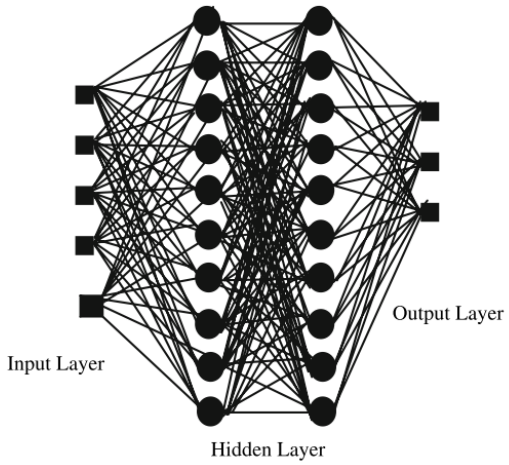

Figure 3.8 – Example of ANN architecture with 5 inputs, 2 hidden layers with 10 nodes each and 3 outputs.

A neural network can be designed by selecting the architecture, activation function and learning rule. In this work, the number of neurons and hidden layers, activation function and learning rate were the trained hyperparameters.

3.3.4 Gradient Boosting Tree

Boosting is an ensemble technique in which the predictors are not made independently, but sequentially [164]. Similarly to the previous mentioned RF algorithm, the gradient boosting tree uses an ensemble of decisions trees. However, while RF involves a parallel ensemble of trees, GBT generates a sequence of trees where each tree is grown on the residuals of the previous trees [165]. Each weak classifier in the iterative process will try to complement the existing trees based on the reconstructed residual between the target function and the current ensemble prediction, minimizing the training error of the ensemble and optimizing a loss function [92]. Prediction is accomplished by weighting the ensemble outputs of all the trees [165].

Due to its sequential nature, GBT tends to be more susceptible to overfitting than RF, however, with the adequate training, it usually outperforms the RF. Figure 3.9 adapted from [92], [165], illustrates an example of a gradient boosted decision tree ensemble.

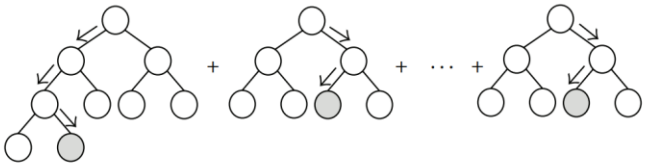

Figure 3.9 - Gradient boosted decision tree ensemble.

The number of boosting stages and the loss functions were the GBT hyperparameters trained during this work in order to obtain the best performance.

3.3.5 K-Nearest Neighbors

Neighbors-based classification is a type of instance-based learning or non-generalizing learning, as it does not attempt to construct a general internal model, but simply stores instances of the training data. In order to assign a label to a query point, a simple majority vote of the K-nearest neighbors is computed, that is, the assigned label is given according to the data class which has the most representatives within the nearest neighbors of the point. To obtain the nearest neighbors, multiple search algorithms and metric distances can be used, as the Ball Tree and KD Tree search algorithms and the standard Euclidean distance, one of the most common metrics [166].

An illustration of K-nearest neighbor algorithm classification process is shown next, in Figure 3.10, adapted from [167].

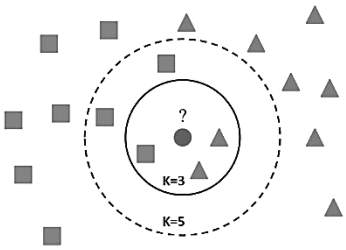

Figure 3.10 – Example of KNN algorithm. The sample (circle) is classified according to the majority of classes in the selected neighborhood (K).

For the KNN algorithm, the trained hyperparameters were the number of neighbors and the search function.

3.3.6 Classifier combination

There are several strategies that can be implemented in order to improve the results, one of those strategy is classifier combination.

The basic idea behind classifier combination is that when making a decision, one should not rely only on a single classifier, but rather, classifiers need to participate in decision making by combining or fusing their individual opinions [168]. If the classification results of different classifiers can be fused in an efficient way, then the outcome of such classifier combinations can be superior to all the individual classifiers. In essence, a generalization of methods as the ones applied in GBT and RF with the decision trees.

In order to combine classifiers, there are two main aspects that need to be considered first, the level of fusion and the combination strategy.

In general, information fusion can be achieved at two different levels: fusion prior to classification/matcher stage, that is, at the sensor and feature levels and fusion after classification/matcher [168], [169].

Fusion levels after classification can be divided in fusion at score level, rank level and decision level. In score level fusion, each classifier provides an opinion on the possible decisions, where non-homogeneous metrics can be used, for example, one classifier provides its opinion in terms of distances while another in terms of a likelihood measure [169]. Rank level fusion is used when the output of each classifier is a subset of possible measures stored in decreasing order of confidence. In fusion at decision/abstract level each classifier makes the decision independently, being then combined [169].

Regarding the existing classifier combination strategies, based on the execution order of the combination, three approaches can be considered: sequential, parallel (Figure 3.11, adapted from [168]) and hybrid combination.

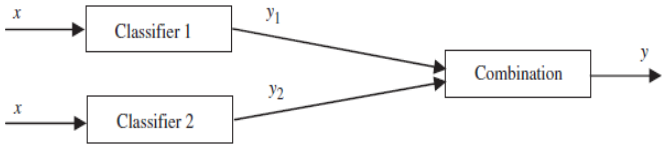

Figure 3.11 - Parallel classifier combination.

In sequential combination, the result of one classifier is used as input to another classifier. The execution order of the classifiers is important and if it is changed, the final result can be different [170]. Parallel combination consists of exploiting several classification algorithms by merging their results. The execution order of the classifiers is not important.

Hybrid combination consists of combining the sequential and parallel techniques. Other strategies can also be considered as combination with looping criterion, that is, the combination result can be used by the classifiers and thus modify their outputs, and combination with interaction between classifiers [168].

In this work, the tested classifier combinations were mainly based on a parallel combination at decision level. Many method have already been developed for this approach [168] as the Bayesian combination [171], fuzzy integrals-based method [172], [173], methods using the Dempster–Shafer theory [174] and majority voting methods [175].

The majority voting is the most commonly used approach for decision level fusion due to its simple theoretical analysis [168], being the chosen method. More detailed, two variations of the major voting technique were tested: hard and soft voting. The first approach can be defined according to Equation (3.1):

$$\hat{Y} = mode\{C_1(x), C_2(x), \dots, C_n(x)\} \tag{3.1}$$

That is, the class label $\hat{Y}$ is chosen via majority (plurality) voting of each classifier (C) (hard voting). A slightly different version can be obtained if the individual classifiers produce probability estimates (soft voting), Equation (3.2):

$$\hat{Y} = arg\ max \sum_{j=1}^{n} w_j p_{ij} \tag{3.2}$$

Where W_j is the weight that can be attributed to the classifier j and p_{ij} is the probability of label i computed by classifier j.

3.4 Imbalance approach

Any dataset that exhibits an unequal distribution between its classes can be considered imbalanced [176].

From the SleepEye dataset, following the mentioned approach of feature extracting in non-overlapping 2 minutes windows, there is simultaneous available from ECG and Video data 1209 samples. For both 2 and 3 class dataset division, the classes' proportions are imbalanced, more precisely the class distribution is the following, Figure 3.12.

Figure 3.12 - Data proportion for 2 and 3-classes problem division.

One of the biggest problems of having an imbalanced dataset is that classifiers tend to provide a severely imbalanced degree of accuracy, with the majority class having close to 100 percent accuracy and the minority class having smaller accuracies.

As in other typical cases, for example, cancerous/non-cancerous medical image classification, there are different misclassification costs [176], [177]. Regarding the current problem, the consequence of misclassifying a drowsy driver as awake is potentially higher than the opposite.

Some of the most used techniques used to deal with imbalanced datasets include cost-sensitive and sampling-based methods. Other popular methods include Kernel-Based and Active Learning Methods [176].

Cost-sensitive methods attempt to balance distributions by considering the costs associated with misclassifying examples [49], [50]. That is, a cost matrix is used to represent the cost associated with each misclassification. If any minority sample gets misclassified in majority class, then its cost will be higher than misclassification of a majority class sample.

Sampling methods consists of the modification of an imbalanced dataset by some mechanism, typically oversampling of the minority class or undersampling of the majority class, in order to provide a balanced distribution.

In this work, a cost-sensitive and a sampling method were applied. Regarding the cost-sensitive method, cost matrix weights were computed inversely proportional to the class weights distribution. As for the sampling method technique, naive random oversampling method was used.

3.5 Models for sequential data

As mentioned in [92] work, typically, drowsiness gradually increases over time, as so, predicting the driver state based on the features in the actual moment may be less accurate than predicting it looking at the pattern also in the moments before. Although [92] tried some simple approaches involving sequential data, that is, considering the sequential nature of the signals, a more detailed studied was needed.

For this purpose, a technique based on hidden Markov models (HMMs) was used, a method most commonly used in speech and language technology fields that recently has started to be applied in driver behavior modeling [178].

The foundation of HMM is a stochastic Markov process that consists of a number of states with corresponding transitions. For example, as illustrated next, in Figure 3.13, adapted from [179].

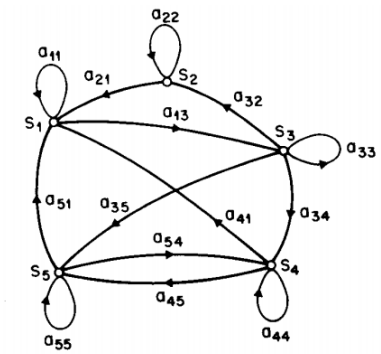

Figure 3.13 – A markov chain with 5 states (S1 to S5) with selected state transitions.

The HMM is always in one of those states, moving from one state to another according to a set of transition probabilities. State changes in the Markov process are hidden from the user [180], [179].

In this work, similar to [180] and [181] work, the overall workflow consisted of two main phases. First, training one HMM (with sets of sequences) for each considered data label, for example, drowsy and awake, and then, giving a new sequence, estimate which HMM presents the maximum likelihood. The training phase consists essentially in adjusting the HMM transition parameters, using methods as the Baum-Welch algorithm, and for the prediction phase, likelihood estimation is done applying the Viterbi algorithm [182].

Using the above mentioned framework, two approaches were considered to obtain data sequences.

A first approach, consisted of computing the mentioned features in 3.2 on smaller windows, for example, 15 seconds windows. Adjacent windows with identical labels were then grouped together, thus obtaining several sets of sequential windows.

For the second approach, no features were extracted, and a raw signal was used by dividing it in smaller sets of windows.

3.6 Subject-independent classification

So far, the discussed approaches involved random splitting all the samples in training and test datasets which are used to train and evaluate the classifiers respectively. This implies that, it is likely to have samples of a specific subject in both the training and test datasets,

increasing the similarity between those two groups and, thus, the classifier will more easily achieve good performances.

Although some intermediate scenarios can also be tested, controlling the number of samples from a certain subject used in the training and then testing the classifier in that subject, the scenario of a completely independent subject classification, is the ideal for a real life classification. Pre-training a classifier that can then be directly applied to detect drowsiness in an unseen driver.

Idealizing this scenario of subject independent classification, an evaluation of the performance that could be achieved using a classifier trained in n-1 subjects and then tested in the n-th subject was also done.

3.7 Results and discussion

3.7.1 Classifiers and features comparison

The results detailed in this section were obtained using the overall workflow described in section 3.1, Table 3.3, shown next, contains a collection of the average accuracies obtained with the previous described classifiers using different features (ECG, Video and ECG and Video) and different dataset categorizations (binary and multiclass). For comparison, the results obtained by [92] with the same dataset and similar methodology, here verified, using EOG features are also displayed next.

Table 3.3 – Accuracy results (average and standard deviation) for combined classifiers using ECG and video features.

Classifiers	Nº Classes	ECG	EOG	Video	ECG + Video
SVM	2	72.1 ± 1.3	86.3 ± 1.3	84.4 ± 0.9	88.8 ± 1.9
	3	55.2 ± 2.9	73.4 ± 1.1	73.4 ± 1.3	78.0 ± 1.9
RF	2	72.0 ± 0.9	86.6 ± 1.4	86.4 ± 1.1	87.7 ± 1.4
	3	54.2 ± 1.2	75.6 ± 1.2	77.1 ± 1.5	78.8 ± 1.2
ANN	2	70.3 ± 1.4	84.5 ± 1.2	82.8 ± 1.1	87.7 ± 1.1
	3	53.8 ± 2.3	73.0 ± 1.4	73.3 ± 1.3	77.4 ± 1.7
GBT	2	72.2 ± 1.1	87.4 ± 0.8	86.6 ± 1.5	89.2 ± 0.9
	3	54.8 ± 2.4	75.8 ± 2.4	75.3 ± 1.2	76.9 ± 1.8
KNN	2	70.0 ± 1.8	81.8 ± 2.1	83.0 ± 2.0	86.4 ± 1.5
	3	51.8 ± 2.4	70.3 ± 1.8	71.6 ± 1.0	74.5 ± 1.6

For each of the performed experiments, further detailed results are presented next, from Table 3.4 to Table 3.13.

Support Vector Machine

Table 3.4 – Normalized confusion matrix using SVM classifier with 2 labels.

Features			predicted	
			awake	drowsy
ECG	**given**	awake	88.8 ± 2.4	11.2 ± 2.4
		drowsy	60.8 ± 6.3	39.2 ± 6.3
			predicted	
			awake	drowsy
Video	**given**	awake	87.7 ± 1.7	12.3 ± 1.7
		drowsy	22.1 ± 3.1	77.9 ± 3.1
			predicted	
			awake	drowsy
ECG+Video	**given**	awake	92.0 ± 2.1	8.0 ± 2.1
		drowsy	17.6 ± 3.6	82.4 ± 3.6

Table 3.5 – Normalized confusion matrix using SVM classifier with 3 labels.

Features			predicted		
			awake	medium	drowsy
ECG	**given**	awake	76.8 ± 6.2	22.4 ± 5.5	0.8 ± 0.8
		medium	53.6 ± 5.8	42.6 ± 5.5	3.8 ± 1.4
		drowsy	49.6 ± 7.1	36.5 ± 6.9	13.8 ± 7.4
			predicted		
			awake	medium	drowsy
Video	**given**	awake	84.6 ± 3.0	14.4 ± 2.3	1.0 ± 1.3
		medium	22.8 ± 2.5	70.3 ± 3.9	6.8 ± 3.6
		drowsy	15.3 ± 4.7	41.6 ± 8.3	43.2 ± 9.5
			predicted		
			awake	medium	drowsy
ECG+Video	**given**	awake	89.4 ± 2.9	10.0 ± 2.7	0.7 ± 0.4
		medium	21.5 ± 2.8	72.9 ± 3.8	5.6 ± 1.7
		drowsy	5.7 ± 4.3	41.3 ± 7.9	52.9 ± 8.1

Random Forest

Table 3.6 – Normalized confusion matrix using RF classifier with 2 labels.

Features			predicted	
			awake	drowsy
ECG	**given**	awake	87.8 ± 1.5	12.2 ± 1.5
		drowsy	59.3 ± 3.1	40.7 ± 3.1
			predicted	
			awake	drowsy
Video	**given**	awake	89.7 ± 1.7	10.3 ± 1.7
		drowsy	19.9 ± 1.5	80.1 ± 1.5
			predicted	
			awake	drowsy
ECG+Video	**given**	awake	91.2 ± 1.6	8.8 ± 1.6
		drowsy	19.3 ± 2.6	80.7 ± 2.6

Table 3.7 – Normalized confusion matrix using RF classifier with 3 labels.

Features			predicted		
			awake	medium	drowsy
ECG	**given**	awake	72.6 ± 2.0	25.1 ± 1.8	2.3 ± 1.1
		medium	52.2 ± 3.2	44.4 ± 3.1	3.4 ± 1.7
		drowsy	46.3 ± 5.3	36.5 ± 3.9	17.2 ± 4.2
			predicted		
			awake	medium	drowsy
Video	**given**	awake	90.0 ± 3.2	9.5 ± 3.2	0.5 ± 0.6
		medium	23.6 ± 3.0	71.7 ± 2.4	4.7 ± 2.3
		drowsy	8.4 ± 3.2	44.7 ± 5.0	46.9 ± 4.4
			predicted		
			awake	medium	drowsy
ECG+Video	**given**	awake	90.9 ± 2.6	8.7 ± 2.4	0.5 ± 0.3
		medium	22.6 ± 2.9	74.1 ± 3.3	3.3 ± 1.8
		drowsy	7.0 ± 2.9	43.2 ± 5.7	49.8 ± 3.8

Artificial Neural Networks

Table 3.8 – Normalized confusion matrix using ANN classifier with 2 labels.

Features			predicted	
			awake	drowsy
ECG	**given**	awake	90.9 ± 1.3	9.1 ± 1.3
		drowsy	70.5 ± 2.9	29.5 ± 2.9
			predicted	
			awake	drowsy
Video	**given**	awake	86.5 ± 2.6	13.5 ± 2.6
		drowsy	24.5 ± 2.8	75.5 ± 2.8
			predicted	
			awake	drowsy
ECG+Video	**given**	awake	90.3 ± 1.5	9.7 ± 1.5
		drowsy	17.5 ± 1.0	82.5 ± 1.0

Table 3.9 – Normalized confusion matrix using ANN classifier with 3 labels.

Features			predicted		
			awake	medium	drowsy
ECG	**given**	awake	77.2 ± 5.7	21.0 ± 4.7	1.8 ± 1.4
		medium	57.3 ± 7.6	40.7 ± 6.5	2.0 ± 1.6
		drowsy	50.8 ± 9.7	41.8 ± 7.4	7.4 ± 5.6
			predicted		
			awake	medium	drowsy
Video	**given**	awake	85.4 ± 4.3	13.8 ± 4.0	0.8 ± 0.6
		medium	21.9 ± 3.4	69.0 ± 4.4	9.1 ± 3.5
		drowsy	13.8 ± 3.1	43.3 ± 4.9	42.9 ± 5.2
			predicted		
			awake	medium	drowsy
ECG+Video	**given**	awake	86.0 ± 2.3	13.1 ± 2.4	0.9 ± 0.5
		medium	20.0 ± 3.6	72.2 ± 3.9	7.8 ± 1.9
		drowsy	5.4 ± 2.1	33.0 ± 6.3	61.6 ± 7.1

Gradient Boosting Tree

Table 3.10 – Normalized confusion matrix using GBT classifier with 2 labels.

Features			predicted	
			awake	drowsy
ECG	**given**	awake	88.3 ± 1.5	11.7 ± 1.5
		drowsy	59.4 ± 3.1	40.6 ± 3.1
			predicted	
			awake	drowsy
Video	**given**	awake	90.1 ± 2.0	9.9 ± 2.0
		drowsy	20.3 ± 3.1	79.7 ± 3.1
			predicted	
			awake	drowsy
ECG+Video	**given**	awake	91.0 ± 1.5	9.0 ± 1.5
		drowsy	14.4 ± 1.8	85.6 ± 1.8

Table 3.11 – Normalized confusion matrix using GBT classifier with 3 labels

Features			predicted		
			awake	medium	drowsy
ECG	**given**	awake	71.3 ± 2.8	26.5 ± 2.9	2.2 ± 1.0
		medium	48.7 ± 4.8	45.8 ± 4.9	5.5 ± 1.5
		drowsy	41.7 ± 6.6	36.7 ± 3.9	21.6 ± 5.2
			predicted		
			awake	medium	drowsy
Video	**given**	awake	86.5 ± 2.6	12.4 ± 2.7	1.0 ± 0.6
		medium	23.2 ± 3.4	71.3 ± 4.3	5.5 ± 1.8
		drowsy	10.4 ± 3.0	41.9 ± 5.2	47.6 ± 4.6
			predicted		
			awake	medium	drowsy
ECG+Video	**given**	awake	88.0 ± 2.5	11.3 ± 2.4	0.7 ± 0.4
		medium	22.0 ± 2.7	71.0 ± 3.5	7.0 ± 1.7
		drowsy	3.6 ± 2.5	41.6 ± 4.0	54.8 ± 4.4

K-Nearest Neighbors

Table 3.12 – Normalized confusion matrix using KNN classifier with 2 labels.

Features			predicted	
			awake	drowsy
ECG	**given**	awake	90.3 ± 2.1	9.7 ± 2.1
		drowsy	70.0 ± 3.6	30.0 ± 3.6
			predicted	
			awake	drowsy
Video	**given**	awake	86.5 ± 2.2	13.5 ± 2.2
		drowsy	23.9 ± 3.6	76.1 ± 3.6
			predicted	
			awake	drowsy
ECG+Video	**given**	awake	90.4 ± 1.9	9.6 ± 1.9
		drowsy	21.4 ± 2.1	78.6 ± 2.1

Table 3.13 – Normalized confusion matrix using KNN classifier with 3 labels.

Features			predicted		
			awake	medium	drowsy
ECG	**given**	awake	79.6 ± 5.7	19.3 ± 5.0	1.1 ± 1.0
		medium	63.6 ± 6.5	33.6 ± 4.3	2.8 ± 2.8
		drowsy	60.3 ± 6.8	34.9 ± 6.3	4.9 ± 5.5
			predicted		
			awake	medium	drowsy
Video	**given**	awake	86.7 ± 3.5	11.7 ± 2.9	1.6 ± 0.9
		medium	27.6 ± 3.5	65.7 ± 4.2	6.7 ± 2.4
		drowsy	16.4 ± 4.5	47.4 ± 8.5	36.1 ± 9.0
			predicted		
			awake	medium	drowsy
ECG+Video	**given**	awake	88.4 ± 1.6	10.9 ± 1.6	0.7 ± 0.6
		medium	26.9 ± 6.0	68.3 ± 5.0	4.8 ± 2.4
		drowsy	13.3 ± 3.1	43.6 ± 6.7	43.1 ± 6.3

In summary, analyzing the results, when comparing between different classifiers, their performance is similar. Although SVM, RF and GBT classifiers seem to perform better in some cases, the overall differences are not significant.

Inspecting the detailed results, the drowsy class reveals the hardest to identify. In particular, all the classifiers tend to predict as awake, drivers that are drowsy.

When analyzing different features, video features perform considerable better than ECG, roughly 14% for the 2 class problem and 20% for the 3 class problem, with ECG performance being similar to [92] work. The multimodal approach, combining ECG and video features increases the performance by approximately 3% when comparing to the video features alone.

For all the classifiers and features, the 2 class problem performance is better than the 3 class.

One of the goals of this project was to evaluate if non-invasive behavioral measures could replace EOG maintaining the same performance. This proved valid, as video features performed equally well or better to EOG features tested in [92] work.

3.7.2 Classifiers combination

As discussed, all the classifiers performed similarly, as so, a potential approach to increase the performance was to combine the decisions of multiple classifiers. Using simultaneously ECG and video features, and following the previous approach of data preprocessing, multiple classifiers were combined at the decision level using both soft and hard voting, the techniques previously mentioned in the methods section 3.3.6.

Initially SVM, RF and GBT classifiers combination was tested, due to their slightly better performance in the binary multimodal approach, other combinations were also tried, but in general no noteworthy results were obtained. For the soft voting approach each classifier has an equal weight. The obtained accuracy results are shown next in Table 3.14.

Table 3.14 – Accuracy results (average and standard deviation) for combined classifiers using ECG and video features.

Classifiers	Nº Classes	Accuracy (soft voting)	Accuracy (hard voting)
SVM+RF	2	87.5 ± 1.1	-
	3	79.5 ± 1.5	-
SVM+GBT	2	88.1 ± 2.2	-
	3	78.9 ± 2.5	-
RF+GBT	2	88.6 ± 1.4	-
	3	78.0 ± 1.7	-
SVM+RF+GBT	2	89.6 ± 1.7	88.7 ± 1.4
	3	80.2 ± 1.9	78.8 ± 1.5
SVM+RF+ANN	2	88.0 ± 1.4	88.3 ± 1.2
	3	79.0 ± 1.3	78.7 ± 1.8

Detailed normalized confusion matrix for all the results are also displayed next, from Table 3.15 to Table 3.19.

Table 3.15 – Normalized confusion matrix using soft fusion of SVM and RF classifiers.

		predicted				predicted		
		Awake	Drowsy			Awake	Medium	Drowsy
given	Awake	90.7 ± 1.3	9.3 ± 1.3		Awake	90.9 ± 2.1	8.4 ± 2.0	0.7 ± 0.7
	Drowsy	18.5 ± 3.8	81.5 ± 3.8	given	Medium	19.3 ± 1.5	75.8 ± 2.1	4.9 ± 1.7
					Drowsy	6.0 ± 2.8	43.0 ± 7.4	51.1 ± 6.5

Table 3.16 – Normalized confusion matrix using soft fusion of SVM and GBT classifiers.

		predicted				predicted		
		Awake	Drowsy			Awake	Medium	Drowsy
given	Awake	92.1 ± 1.9	7.9 ±1.9		Awake	89.5 ± 2.2	10.0 ± 2.1	0.5 ± 0.7
	Drowsy	19.1 ± 4.8	80.9 ± 4.8	Given	Medium	19.9 ± 3.7	75.4 ± 3.6	4.6 ± 1.4
					Drowsy	5.5 ± 3.2	41.0 ± 7.8	53.5 ± 7.7

Table 3.17 – Normalized confusion matrix using soft fusion of GBT and RF classifiers.

		predicted				predicted		
		Awake	Drowsy			Awake	Medium	Drowsy
given	Awake	90.5 ± 1.7	9.5 ± 1.7		Awake	89.8 ± 1.9	9.4 ± 2.0	0.8 ± 0.4
	Drowsy	14.9 ± 4.0	85.1 ± 4.0	given	Medium	22.6 ± 3.2	72.2 ± 3.1	5.2 ± 1.7
					Drowsy	8.2 ± 3.2	37.6 ± 7.3	54.3 ± 6.4

Table 3.18 – Normalized confusion matrix using soft and hard fusion of SVM, RF and GBT classifiers.

Soft voting								
		predicted				predicted		
		Awake	Drowsy			Awake	Medium	Drowsy
given	Awake	92.1 ± 2.0	7.9 ± 2.0		Awake	91.5 ± 2.1	7.8 ± 2.0	0.6 ± 0.6
	Drowsy	15.3 ± 3.0	84.7 ± 3.0	Given	Medium	18.6 ± 3.0	76.8 ± 2.4	4.6 ± 1.9
					Drowsy	8.2 ± 4.9	39.3 ± 7.8	52.5 ± 7.0

Hard voting								
		predicted				**predicted**		
		Awake	Drowsy			Awake	Medium	Drowsy
given	Awake	91.8 ± 1.3	8.2 ± 1.3	**Given**	Awake	91.0 ± 2.3	8.5 ± 2.0	0.5 ± 0.6
	Drowsy	17.2 ± 2.9	82.8 ± 2.9		Medium	21.8 ± 2.7	74.1 ± 3.4	4.2 ± 1.8
					Drowsy	10.2 ± 5.0	39.9 ± 9.4	49.9 ± 6.6

Table 3.19 – Normalized confusion matrix using soft and hard fusion of SVM, RF and ANN classifiers.

Soft voting								
		predicted				**predicted**		
		Awake	Drowsy			Awake	Medium	Drowsy
given	Awake	91.8 ± 2.0	8.2 ± 2.0	**Given**	Awake	89.0 ± 1.4	10.4 ± 1.4	0.6 ± 0.4
	Drowsy	18.6 ± 3.3	81.4 ± 3.3		Medium	17.3 ± 2.5	76.7 ± 2.7	6.0 ± 2.2
					Drowsy	7.7 ± 4.0	40.5 ± 7.2	51.8 ± 6.9
Hard voting								
		predicted				**predicted**		
		Awake	Drowsy			Awake	Medium	Drowsy
given	Awake	91.6 ± 1.6	8.4 ± 1.6	**Given**	Awake	89.4 ± 1.6	10.1 ± 1.3	0.5 ± 0.6
	Drowsy	17.7 ± 2.6	82.3 ± 2.6		Medium	20.5 ± 4.0	74.4 ± 3.9	5.1 ± 1.6
					Drowsy	8.2 ± 4.2	38.2 ± 6.9	53.6 ± 4.4

After reviewing the results, the applied combinations of classifiers did not increase the overall performance. This might be explained by the high performance that the classifiers were already obtaining, high correlation between the classifiers and, in addition, the obtained results may already be on the limit of what is achievable with this dataset.

Nonetheless, other attempts to increase the performance were tested. One of the experiments consisted in combining two SVM classifiers, with each classifier using different information sources, namely one using only ECG features and another using video features. The obtained accuracy results and normalized confusion matrixes and shown next, in Table 3.20 and Table 3.21 respectively.

Table 3.20 – Accuracy (average and standard deviation) results obtained with two SVM classifiers using different information sources.

Classifier (features)	Nº Classes	Accuracy
SVM(ECG)+SVM(Video)	2	83.9 ± 2.2
	3	73.8 ± 2.6

Table 3.21 – Normalized confusion matrix for the two SVM classifiers using different information sources.

		predicted				predicted		
		Awake	Drowsy			Awake	Medium	Drowsy
given	Awake	94.2 ± 1.5	5.8 ± 1.5	given	Awake	88.0 ± 3.1	11.2 ± 3.0	0.8 ± 0.5
	Drowsy	34.6 ± 6.1	65.4 ± 6.1		Medium	25.4 ± 4.3	71.5 ± 4.4	3.0 ± 1.2
					Drowsy	10.8 ± 5.6	56.7 ± 5.8	32.4 ± 5.2

The combination of the two SVM classifiers one using EGG features and another with video features did not perform well, decreasing the performance when comparing with the results obtained using the same classifier with both ECG and video features. The accuracy lowered by approximately 3% for both the 2 and 3 classes' problem.

3.7.3 Imbalanced data

As can be verified from the previous results, although the overall obtained accuracies are reasonably high, when analyzing the performances per class, the majority class is always obtaining better results. To balance these variations, the cost-sensitive and oversampling approaches previous explained in the methods section were applied for an SVM classifier using ECG and video features. The same workflow of data division and processing was used. The results are presented next in Table 3.22.

Table 3.22 – Accuracy (average and standard deviation) results for SVM classifier using ECG and video features.

Classifier	Balancing approach	Nº Classes	Accuracy
SVM	Cost-sensitive	2	87.3 ± 1.5
		3	76.1 ± 1.6
	Oversampling	2	87.2 ± 0.9
		3	76.2 ± 1.7

Further detailed results are displayed next in Table 3.23 and Table 3.24.

Table 3.23 - Normalized confusion for imbalanced approaches for 2 class problem.

Balancing approach			**predicted**	
			awake	drowsy
Cost-sensitive	**given**	awake	88.4 ± 3.8	11.6 ± 3.8
		drowsy	14.2 ± 7.4	85.8 ± 7.4
			predicted	
			awake	drowsy
Oversampling	**given**	awake	91.1 ± 2.8	8.9 ± 2.8
		drowsy	20.5 ± 3.4	79.5 ± 3.4

Table 3.24 - Normalized confusion for imbalanced approaches for 3 class problem.

Balancing approach			**predicted**		
			awake	medium	drowsy
Cost-sensitive	**given**	awake	86.7 ± 3.0	11.2 ± 3.2	2.1 ± 1.5
		medium	18.3 ± 3.1	69.8 ± 4.8	11.9 ± 4.5
		drowsy	8.5 ± 5.5	32.8 ± 7.6	58.8 ± 10.5
			predicted		
			awake	medium	drowsy
Oversampling	given	awake	86.3 ± 1.8	13.1 ± 1.9	0.6 ± 0.4
		medium	20.1 ± 4.9	71.8 ± 3.8	8.1 ± 2.2
		drowsy	9.2 ± 3.1	36.9 ± 7.2	53.8 ± 7.3

The overall performance for both methods was identical to the performance obtained without using any imbalance data method for both 2 and 3 class division. Analyzing the detailed confusion matrixes, although only minor differences are noticed, two conclusions can be made. First, the cost-sensitive method performs better than the oversampling and, secondly, for this approach, the accuracy gap between classes slightly decreases, however with a bigger variability (standard deviation).

3.7.4 Window size sensitivity

The window size sensitive (section 3.2.2) was evaluated analyzing the performance of an SVM classifier trained with both ECG and Video features. The features were extracted in different time windows ranging from 15 seconds to 5 minutes. The obtained accuracy results are presented next in Figure 3.14.

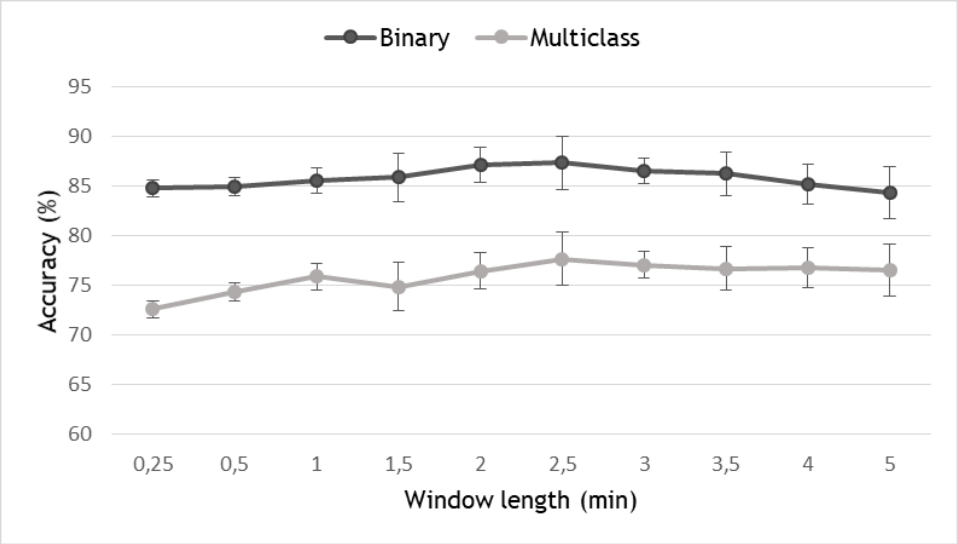

Figure 3.14 – Accuracy results (average and standard deviation) for two and three class problem for features extracted in different time windows.

Analyzing the results, we can observe that, although there are some slight differences in the performances from the different windows lengths, the results are fairly similar. As so, it is possible to conclude that, most likely, choosing a different time window length would not provide better results.

3.7.5 Sequential data

According to the procedure explained in section 3.5 a sequential data approach was tested. More precisely, regarding the first strategy, that is, extracting features in small windows and creating sequences with the adjacent windows, multiple windows lengths were evaluated, as well as, the total number of windows in each sequence (sequence length). Most of the results did not show good accuracies. The best performance was obtained using windows of 15 seconds organized in 2 minutes sequences.

Table 3.25 shows the obtained accuracy results using ECG features for both 2 and 3 labels problem categorization, using 10-fold validation.

Table 3.25 - HMM accuracy (average and standard deviation) results using ECG features.

Features	Sequence lenght	Nº classes	Accuracy (%)
ECG	2 minutes (8 x 15 seconds windows)	2	75.5 ± 5.6
		3	53.9 ± 3.7

A more detailed view of obtained results is showed next, in Table 3.26

Table 3.26 - Normalized confusion matrix using HMM with 15 (seconds) x 8 sequences.

		predicted	
		Awake	Drowsy
given	Awake	88.7 ± 5.4	11.3 ± 5.4
	Drowsy	74.1 ± 15.1	25.9 ± 15.1

		predicted		
		Awake	Medium	Drowsy
given	Awake	30.7 ± 11.9	67.5 ± 11.8	1.8 ± 1.6
	Medium	18.6 ± 8.4	79.0 ± 8.4	2.4 ± 1.6
	Drowsy	14.1 ± 11.9	72.7 ± 15.2	13.1 ± 8.6

Analyzing the overall accuracy results for the HMM approach and in comparison with the respective results obtained with the classifiers in the previous section, particularly for the 2 class problem, the HMM seems to be more efficient. However, one a more detailed look, inspecting the values per class there is a high variability between the results, indicating a lack of consistency.

The same procedure was also tested with video features and the combination of both, but it revealed ineffective with results roughly 10% below the obtained with a classifier as SVM.

Furthermore, to analyze if the overall unsatisfactory results obtained with the HMM approach were because of the features and signal discretization used, the second discussed approach was tested. Specifically, the raw ECG signal was used for comparison. The results are the following, Table 3.27.

Table 3.27 – HMM accuracy (average and standard deviation) results obtained using the raw ECG signal.

Input	**Sequence length (minutes)**	**Accuracy (%)**
ECG signal	1	61.6 ± 11.1
	2	50.9 ± 19.6
	3	52.3 ± 13.7

This approach was considerably inferior to the first described. The best result was obtained using sequences of 1 minute length, but even so, it is more than 10% inferior to the HMM approach using features. The variability between the results is also very high, with some of the validations folds obtaining performances close to zero in the drowsy class.

3.7.6 Subject-independent classification

For this experiment both ECG and Video features were used. An SVM classifier was trained in n-1 subjects and then evaluated in n-th subject. This process was repeated for all the subjects, which represents a total of 17 persons. The average, minimum and maximum accuracy values from the predictions of all subjects are presented next, in Table 3.28.

Table 3.28 - Accuracy results for subject independent classification.

Classifier	Nº Classes	Average	Min	Max
SVM	2	75.5	28.8	92.9
	3	58.8	23.8	77.4

As concluded in the previous work [92], the obtained results with subject independent tests, for both two and three class problems were considerably worse than when using all the data mixed. There is also a substantial difference between the average results and the worst result. Partially, this difference might be due to the fact that there is only available data from the day session for that subject (minimum result) in comparison with others were day and nights sessions are available. Experimenting only for subjects were both day and night data is available (15 subjects), increases the worst case by approximately 15%, but the performance is still far below the required for a functional drowsiness detector.

However, the main reason for the overall low performances obtained with the subject independent experiment is, most likely, the subjective nature of drowsiness. First, there might be individual subject differences in the KSS scale used as ground truth, as one subject feeling of drowsiness can be considerable different from another subject. Indeed, the subjective nature of the KSS scale has already been reported [12]. It is also probable that this drowsiness subjectiveness may also extend to physiological levels, meaning that each driver may manifest drowsiness in slightly different ways [40], [134], [178].

Consequently, in order to obtain an accurate sleepiness detector, further research needs to be performed accounting for this individual subject differences, using, for example, personalized algorithms, that can be adjusted between individuals.

Chapter 4

A method for non-invasive acquisition of drowsiness measures

In this chapter, a methodology developed to extract drowsiness related measures non-intrusively is described.

While the approaches discussed in the previous chapter consist of a series of techniques that attempt to maximize sleepiness detection from a set of already extracted measures (SleepEye dataset), this chapter focus on the measure acquisition process.

In the SleepEye dataset, as stated, all the visual measures are obtained using a Smart Eye Pro system. Although it provides accurate measures, the system was not specifically designed to detect drowsiness, which may be limiting sleepiness detecting performance using video. In particular, there are other potential drowsiness related measures that could be analyzed and used to better evaluate drowsiness.

As such, an exploratory work was developed, to allow the extraction of additional drowsiness related measures using video from the drivers face, particularly yawn related measures and heart rate estimation. The exploratory nature of this work should be empathized and the obtained results, although promising, should only be regarded as a proof of concept rather than a final validation.

All the algorithms implemented in this chapter were developed using videos obtained from YouTube, as well as, a video provided by Professor André Lourenço.

4.1 Framework description

To achieve the outlined objective of extracting drowsiness related measures from the drivers face, an algorithm was developed based on a review of computer vision state of the art methodologies used in face analysis.

The algorithm was implemented in Python, as it is a high-level programming language with extensive support libraries, and using videos obtained from standard cameras in a single viewing angle, which has the advantage that it is more practical to implement in a real life application.

The developed framework can be divided in three main steps. First a face detection stage, which identifies the driver face in a given video frame, secondly, using the obtained face boundary, a set of facial landmarks locations are determined, lastly, with the estimated face

and landmarks locations several drowsiness related measures are computed. The described workflow is outlined in Figure 4.1.

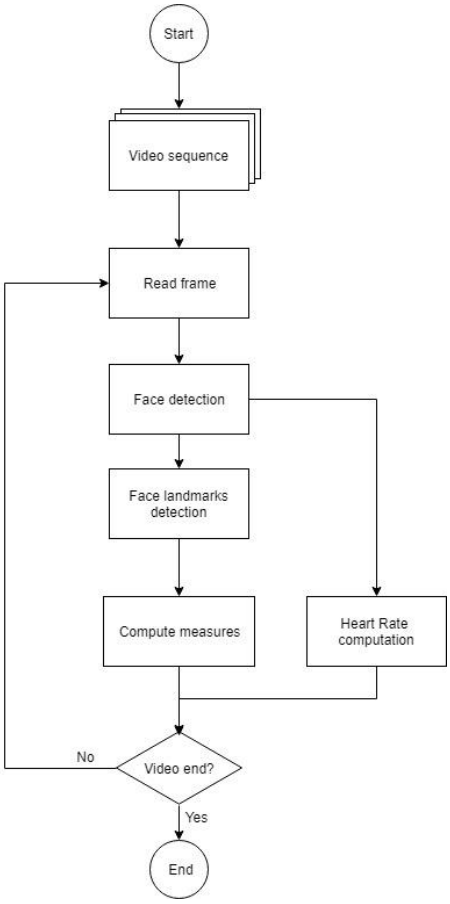

Figure 4.1 – Overall workflow of the drowsiness measures extraction process.

A more detailed description of the methods used in each phase of the framework is further explained next.

4.2 Face detection

As indicated, the first step in the workflow is to correctly identify the drivers face. Because all the subsequent steps are dependent on its correct detection, a robust face detector is crucial. In the Literature review chapter, some of the methods typically used for face detection in drowsiness related problems are discussed, as the Viola-James detector [80], [63], [65], [81], [85]. Although those more traditional algorithms perform well in controlled environments, this kind of detector may degrade significantly in real-world applications with larger visual variations of human faces or illumination conditions.

Consequently, and with more recent and efficient face detection strategies already being used in other projects [183], [184], in this work, a robust cascaded Convolutional Neural Network (CNN) face detector was implemented based in [183] work.

Identically to the ANN discussed in the previous chapter, a CNN is a multi-layered non-fully-connected neural network biologically inspired [185], [186]. Although a CNN may have many different architectures, in general, they consist of convolutional, pooling (or subsampling) and

fully connected layers. Convolutional layers work as feature extractors, and thus they learn the feature representations of their input images. Pooling layers reduce the spatial resolution of the feature maps, achieving spatial invariance to input distortions and translations.

These layers are usually stacked on top of each other, to extract more abstract feature representations, constituting modules, and either one or more fully connected layers, as in a standard feedforward neural network, follow these modules interpreting these feature representations and performing the function of high-level reasoning

Figure 4.2, adapted from [186], illustrates this overall CNN architecture.

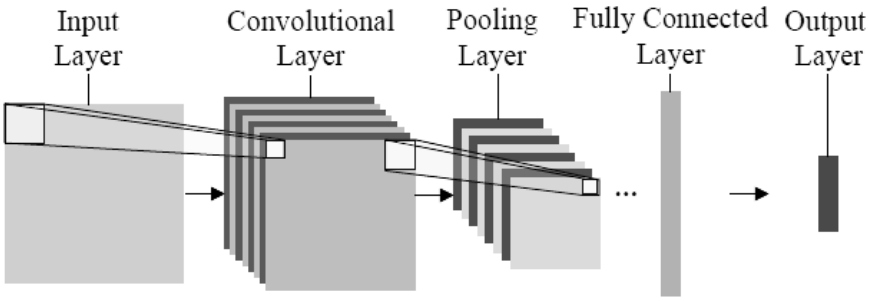

Figure 4.2 – Overall CNN architecture.

The cascade of CNNs implemented in this work consists of three sequential stages. In the first stage, a shallow CNN quickly produces candidate windows for the face locations (P-Net). Then, it refines the windows by rejecting a large number of non-faces windows through a more complex CNN (R-Net). Finally, it uses a more powerful CNN to refine the result again and output five facial landmarks positions (O-Net).

A more detailed view of each of the CNN architectures is illustrated in Figure 4.3, obtained from [183].

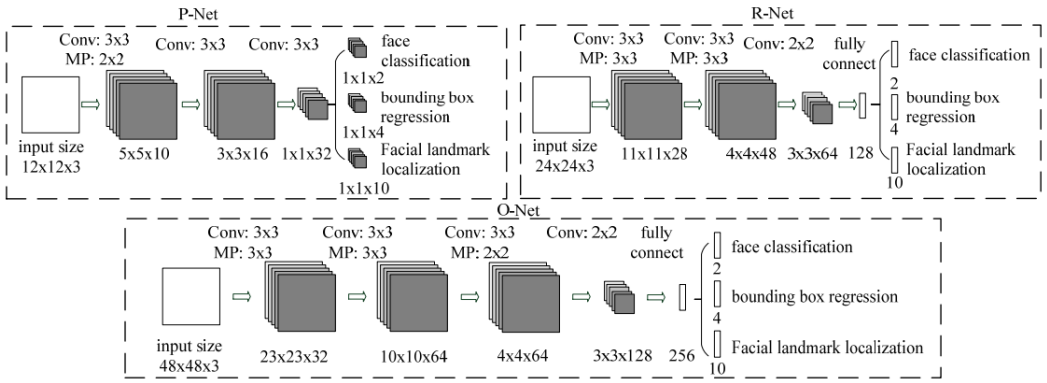

Figure 4.3 - The architectures of P-Net, R-Net, and O-Net, where "MP" means max pooling and "Conv" means convolution.

More precisely the used algorithm is adapted from [187] repository, a TensorFlow implementation of the described workflow, which provides an already trained model.

4.3 Face landmarks detection

Following the face detection, which, for practical purposes, provides a rectangle outlining the drivers face, a face landmark detection step is applied. Here, the provided landmarks will

serve as the structure for the drowsiness related measures extraction, but the facial landmark detection is an essential process for many tasks as face recognition, face tracking, face animation and 3D face modeling [188].

In summary, the facial landmarks detection (or face alignment) consists in locating characteristic points of the face, such as eyes, nose, mouth and chin. This is performed using a face shape mark-up, consisting of several landmarks, that is fitted to the face image. In order to estimate the alignment there are two main categories of approaches that can be considered, optimization-based, which fits the shape minimizing an error function, and regression-based methods, which learn a regression function that directly maps image appearance to the target output [188].

In this project, the adopted alignment strategy consists of a cascade of regression functions [189]. Considering a shape (S) denoting the (x, y) coordinates of all the facial landmarks in the image (I), each regressor (r) in the cascade predicts an estimation of the shape based on the current shape estimation ($\hat{S}$) and a set of features extracted from the image, such as intensity, as represented in Equation (4.1).

$$\hat{S}^{(t+1)} = \hat{S}^{(t)} + r_t(I, \hat{S}^{(t)}) \tag{4.1}$$

The initial shape can simply be chosen as the mean shape of the training data centered and scaled according to the bounding box output of the face detector [189]. An illustration of the used facial shape mark-up is shown next, in Figure 4.4, adapted from [190].

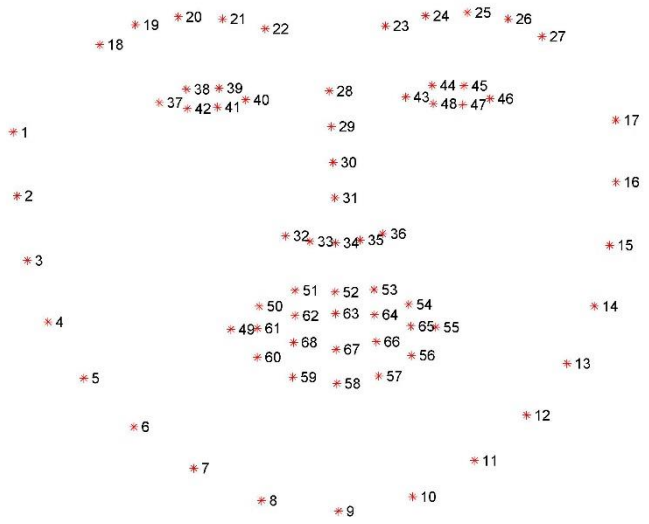

Figure 4.4 – Facial landmarks 68 points mark-up.

The implementation of the described method was done using Dlib-ml toolikit [191], which provides an already trained facial landmark detector.

4.4 Measure extraction

As stated, the main objective in developing this framework was to extract from the drivers face a vast amount of drowsiness related characteristics. This includes obtaining the same

measures used in Chapter 3 plus some additional measures. However, the pupil average diameter available from the SleepEye dataset could not be obtained, as it requires specific conditions to be properly acquired, for example, using an infrared light source.

Nonetheless, eyelid opening, eye gaze and head pose are acquired and additionally, the mouth opening (lips distance), eyebrow raise and heart rate are also computed. A brief explanation of the methods used for the measures computation is presented next.

4.4.1 Facial expressions

Multiple researches have approached the problem of drowsiness detection by analyzing facial expressions. In this project three measures related to facial expressions are extracted, the eyelid opening, mouth opening and eyebrows raise.

All these measures are computed based on the same principle, which is, using a Euclidian distance metric between some of the detected landmarks positions.

To compute the eyelid opening the distance between the upper and lower eyelid is used. For example, for the left eye, illustrated in Figure 4.5, following the previous introduce landmarks scheme, the distance is obtained averaging two distances, d1 and d2, respectively the distance between point 38 and point 42 and the distance between point 39 and 41. The distance between the eye corners is used for normalization (d3).

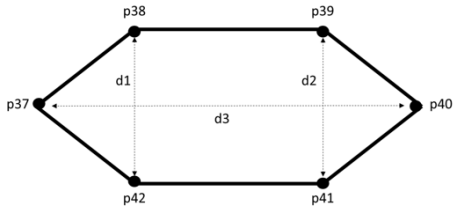

Figure 4.5 - Scheme of left eyelid distance calculation method.

The final eyelid opening measure is obtained averaging the computed eyelid distances from both eyes.

Regarding the mouth opening, an equivalent strategy is applied, using the average distance between the upper and lower lips contours. The measure is normalized with the mouth corner distance.

The eyebrow rising measure is subdivided in inner and outer eyebrow rise. In order to obtain this measure, the average eyebrow distance to a reference point is calculated. As schematized in Figure 4.6, for the left face region, the inner eyebrow distance is computed by the average of distance d1 and d2 and for the outer eyebrow the distance d3 and d4.

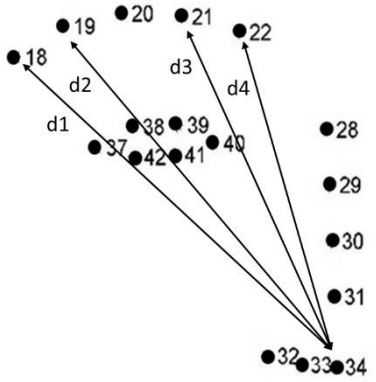

Figure 4.6 – Detailed view of the 68 mark up, with schematized references used for the eyebrow distance calculation method.

The symmetric process is done with the right face region and the final results is the average of both. The measure is normalized with the distance between eye corners.

4.4.2 Head pose

Head pose estimation, as already introduced, is the process of determining the 3 dimensional orientation and position of the head (rotation and translation). There are multiple approaches to this problem, as appearance templates, nonlinear regression and geometric methods [192]. In this project, taking advantage of the knowledge of multiple facial key points location, a geometric method is used.

The geometric relationship between a 3D point in the world coordinate space and its projection in the 2D image can be mathematically expressed according to Equation (4.2):

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K\ [R|t] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{4.2}$$

Where s is a scale factor, (u, v) denotes the 2D coordinates of the projection point and (X, Y, Z) denotes the coordinates of a 3D point in the world coordinate space. K is a 3x3 matrix which contains the camera calibration parameters such as the focal length and the optical center point coordinates. [R|t] is a 3x4 matrix which corresponds to the Euclidean transformation from a world coordinate system to the camera coordinate system. R is the rotation matrix and t the translation vector [193].

Knowing the position of multiple 3D points and respective 2D image projections, it is possible to estimate the orientation (R), relatively to the camera coordinate system, solving a Perspective n Point (PnP) problem. The described problem scheme can be visualized in the following image, Figure 4.7, adapted from [193].

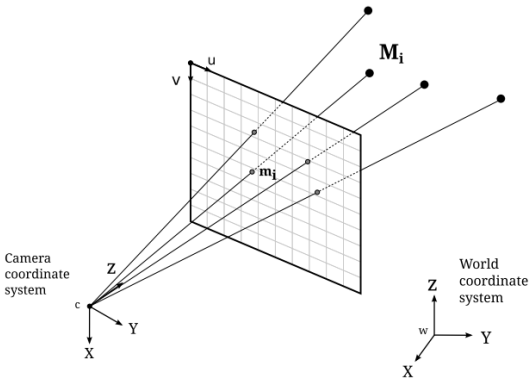

Figure 4.7 – The PnP Problem scheme: Given a set of 3D points (Mi) expressed in a world reference frame, and their 2D projections (mi) onto the image, we seek to retrieve the pose relatively to the camera coordinate system.

More precisely, the 3D location of a set of reference facial points in a world coordinate system (obtained using 3D anatomical reference points [194], [195]), the respective 2D projections in the image coordinates (facial landmarks) and the camera intrinsic parameters (estimated from the image size) are used to estimate the head pose using OpenCV library PnP solver [196].

The estimated head pose orientation is given in the form of a matrix that can be viewed as a multiplication of three rotations: one about each principle axis, as in Equation (4.3).

$$R = R_z(\gamma)R_y(\theta)R_x(\phi) \tag{4.3}$$

Using this relation between the rotation matrix (R) and the three principle axis, head pose yaw (γ), pitch (θ), and roll (φ) angles are calculated [193].

4.4.3 Gaze

As discussed, eyes and their movements play an important role in expressing a person's desires, needs, cognitive processes and emotional states. In fields as human-computer interaction, the development of robust gaze tracking methods is crucial, and, often, complex methods involving more than one camera or IR lights are used to accurately track gaze [197].

In this work, however, a simpler approach was implemented, as, for the needed purpose, the fundamental gaze information can be obtained by the relative pupil-eye corners displacement. In order to compute this displacement (d), first the pupil location is determined. This process, based in [198] work, uses a shape-based approach that exploits the predicted dark circular appearance of eye features, combining image intensity and gradient vectors to accurately estimate the pupil location. The method is applied directly on the eye region determined by the facial landmarks location.

After obtaining the pupil location, the displacement is computed based on the distance to the eye center, which is estimated using the eye corner landmarks. Displacement in both horizontal and vertical directions is obtained which is then normalized using the distance between the eyes. A schematization of the described process is illustrated next, in Figure 4.8.

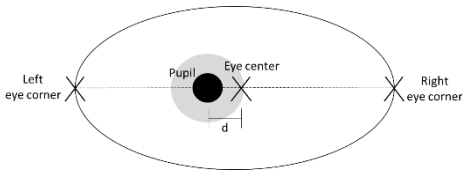

Figure 4.8 – Scheme for pupil displacement calculation.

Nonetheless, an approach using a simple approximation to a 3D gaze was also tried. This estimation is done assuming a fixed radius (R) of the eyeball, based on anatomic reference values, and applying a straightforward trigonometric formula to solve the angle (Θ) between the eyeball center/eye center line and the computed displacement (d), as illustrated in Figure 4.9.

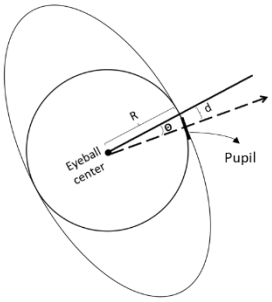

Figure 4.9 – Scheme of the 3D gaze approximation.

The angle computed using the horizontal displacement is equivalent to the yaw angle and the computed from vertical displacement the pitch angle. The final gaze is obtained averaging the results from both eyes and respectively adding the obtained angles to the head pose yaw and pitch angles.

It should be noted that this method accuracy is limited, and if a true 3D gaze is to be used, further research should be performed.

4.4.4 Heart rate

Remotely measuring the heart rate using a camera has already been successfully achieved, although in controlled environments, using more than one method, namely methods based in subtle head motions [199] and average optical intensity based methods [200], [201]. In this work, the technique described in [201] was followed, based on [202] implementation.

This technique uses the cardio-vascular pulse wave that travels through the body in order to compute the heart rate. This principle is the same as of the Photo-plethysmography (PPG), that is, blood absorbs light more than surrounding tissue so variations in blood volume affect transmission or reflectance correspondingly [201].

Typically, the PPG is performed with dedicated light sources in the red and/or infra-red (IR) wavelengths. However, the same procedure can still be applied with normal ambient light as the source.

As so, in this work, a heart rate estimation is performed by analyzing the average optical intensity extracted from the forehead drivers region. The forehead region is computed from the boundaries of the detected face region.

4.5 Validation

In the analyzed driver videos, it was not available any driver sleepiness ground truth. Due to this lack of a reference label, evaluation the process applying a supervised machine learning classification task is not possible.

Nonetheless, as an alternative, a validation of the key steps that establish the algorithm performance was done, specifically, the landmark detection and heart rate steps were tested. In addition, in spite of its limited proof of validation, the overall process was evaluated adopting the trained models described in the previous chapter.

A more detailed description of the executed validation methods and its discussion is presented below.

4.5.1 Landmarks validation

In the description of the measures extraction process it is possible to conclude that most of the extracted measures are dependent on the accuracy of the detected facial landmarks location. As such, evaluating the error of its detection, although not directly confirming the extracted measures correlation with drowsiness, it can confirm, at least with some degree, that they are accurate.

As already stated, the algorithm development was done using videos obtained from YouTube. In particular, for the landmarks evaluation, frames from three videos ([203], [204], [205]) of self-recorded driving sessions were selected and manually annotated. Figure 4.10 illustrates a comparison of the detected and manually annotated landmarks for two of the evaluated frames, obtained from [203] (left frame) and [204] (right frame).

Figure 4.10 - Example of detected (blue) and manually annotated (red) landmarks.

The comparison of the detected and annotated landmarks was done using a normalized root mean square error (RMSE) [206], as depicted in Equation (4.4).

$$RMSE = \frac{\sum_{i=1}^{N} \sqrt{(x_i^f - x_i^g)^2 + (y_i^f - y_i^g)^2}}{d_{outer} N} \tag{4.4}$$

Where x, y are the landmarks location in the image coordinates, f and g, denote detected and ground truth respectively, N the total number of landmark and d_{outer} the Euclidian distance between the inner eye corners. For the chin region as the exact spacing between points was very hard to mark consistently, the error was computed between the predicted landmark and the closest point belonging to a line defined by the linked ground truth chin points.

The average RMSE obtained from a total of 100 frames (captured from the cited videos), is listed next, in Table 4.1.

Table 4.1 – Average landmark detection RMSE values discriminated by face region.

Face region	RMSE
Eyes	0.042 ± 0.006
Mouth	0.033 ± 0.006
Eyebrows	0.077 ± 0.017
All landmarks	0.048 ± 0.040

Analyzing the results the obtained average error seems satisfactory. Indeed, when visual inspecting the predicted landmarks locations in several frames, although some slight deviations might be observed in relation to the manually annotated, the detected landmarks are correctly depicting the interest shapes, as the eyelids, lips and face contours. As so, most likely, the measures computed using the landmarks are sufficiently accurate for the purposed objectives.

However, it should be emphasized that all the tested frames are obtained from videos with a favorable camera positioning, that is, the driver face is frontally recorded. Moreover, all the videos were recording during day light. Further generalization of the algorithm to be satisfactorily robust to typical variations, as well as a faster processing speed, should be researched.

It is also noteworthy that on this work the different tasks, that is, face location, landmark location and measure extraction were approached as separate problems. However, there are already some projects, as the HyperFace [207] and All-In-One Face [107], that explore the face analysis problem using a different strategy, an integrated approach which seems to boost the performance of the individual tasks. For example, using a deep CNN that can simultaneous perform face detection, face alignment, pose estimation, gender recognition, smile detection, age estimation and face recognition. Implementing a similar, integrated, strategy might increase the robustness of the non-invasive drowsiness measures acquisition process, as well as, potentially allow a better evaluation of the driver state by including contextual features.

For illustrative purposes, a frame with depicted eyes and mouth contours, pupil center and head orientation (represented by the green cube) is shown next, Figure 4.11.

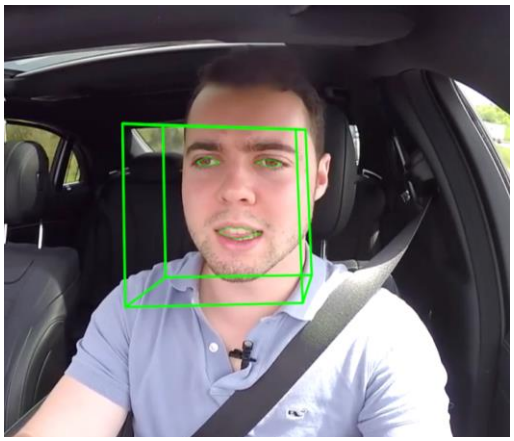

Figure 4.11 – Illustration of detect eye and mouth contours and head orientation.

An example of the eyelid opening distance signal obtained using the developed framework is presented next in Figure 4.12. For comparison, a sample of the eyelid opening signal obtained with the Smart Eye Pro system is also displayed next in Figure 4.13.

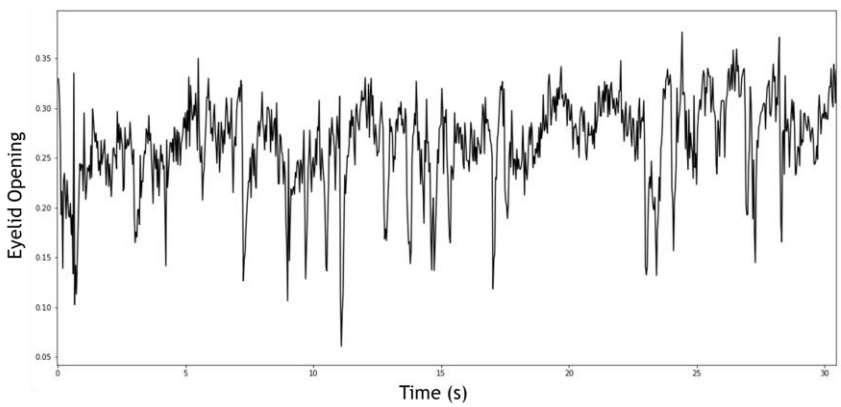

Figure 4.12 – Example of measured eyelid opening distance.

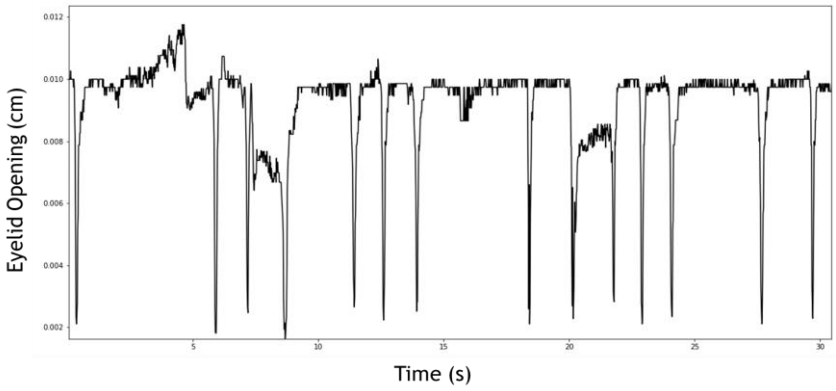

Figure 4.13 – Example of eyelid opening distance obtained by the Smart Eye Pro system.

Although the samples are from different subjects in different conditions, thus not being the most accurate comparison, the signal measured using the developed framework has considerably more noise.

Additionally, as an example, samples of measured head pose and mouth opening signals are also shown next in Figure 4.14 and Figure 4.15 respectively.

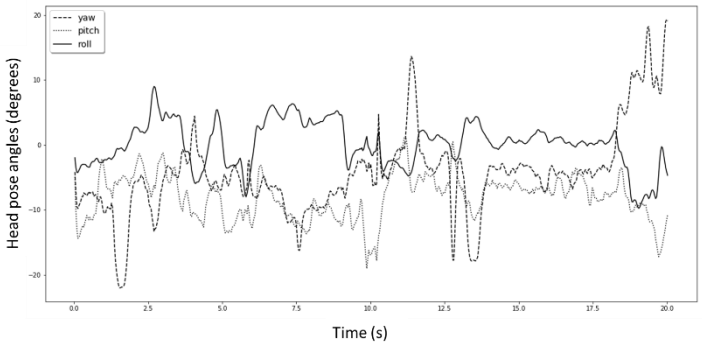

Figure 4.14 – Example of measured head pose angles.

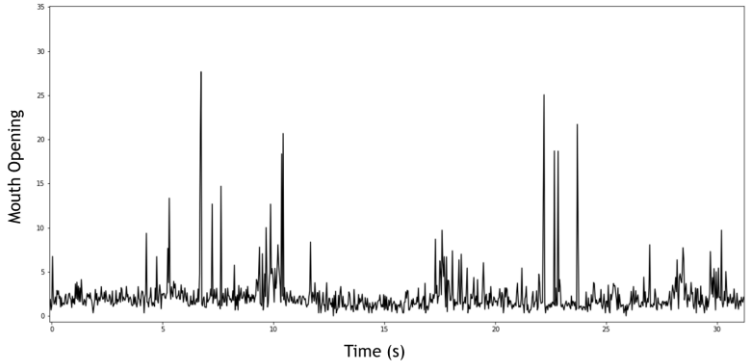

Figure 4.15 – Example of measured mouth opening.

4.5.2 Heart rate validation

In an attempt to validate the heart rate acquisition method, a video recorded during a driving session provided by Professor André Lourenço was used. The computed heart rate was compared with a reference signal, simultaneously acquired during the video recording with a chest strap. The result, obtained averaging both signals in 2 minutes windows, is displayed next in Figure 4.16.

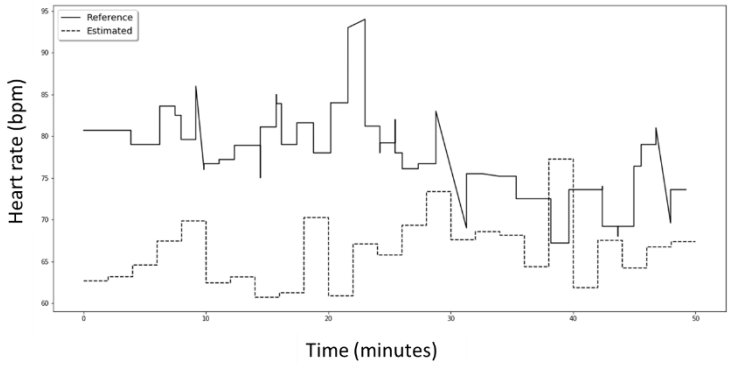

Figure 4.16 – Comparison between estimated and chest strap acquired heart rate signal.

The results were unsatisfactory as it does not seem to be a valid correlation between the obtained measure and the heart rate provided by the chest strap. This is implicitly a method

extremely dependent on the light conditions, and in real driving conditions, shadows and high peaks of light create too much noise. And, it is possible that this method cannot achieve satisfactory results under such conditions.

Nonetheless, the recorded drivers face viewing angle was not the most favorable, which may also have affected the results. Moreover, when testing the algorithm in a controlled environment with a steady pose and light, at least a difference between periods of low and high heart rate could be observed. Indeed, if some degree of distinction between different levels of heart rate could be obtained, the signal could, for example, provide some complementarity to a noisy, non-intrusive, ECG acquisition method. As so, further investigation in this or similar approaches should not be disregarded.

4.5.3 Proof of concept

Despite the importance of having an effective algorithm extracting accurate measures from the driver face, the main goal of this thesis is to detect drowsiness, as so, it is important to understand if the extracted measures are valid indicators of driver drowsiness.

The desired validation process could not be applied in the available data, both for the lack of ground truth already mentioned, but also, because the amount of adequate video samples available was not enough to train and evaluate a classifier.

In alternative, the strategy used was to train a classifier with the data available from the SleepEye dataset and then evaluating the performance in samples extracted using the developed algorithm. In particular, an SVM classifier was trained using all the video features described in Chapter 3, except for pupil related measures. A total of 15 samples, computed on 2 minute windows were evaluated, 12 samples were considered examples of awake drivers and 3 samples were specifically recorded simulating a drowsy driver behavior, for example, with slow eyelid movement and some representation of nods.

All the samples were rightfully predicted (100% accuracy). The obtained results are satisfactory, nonetheless, as expressed, it is always a very limited validation.

Foremost, the additional features obtained using the developed algorithm (mouth opening, eyebrow raising, heart rate) should be included to evaluate drowsiness. Thereby, an adequate validation of the developed method could be performed, training the classifiers using all the available features in a large dataset, namely with several hours of labeled driving sessions.

Chapter 5

Conclusions and future work

5.1 Conclusions

Driver drowsiness presents as an actual and relevant problem that is the cause of many driving accidents. The number of causalities due to drowsy driving accidents is high and has been constant over the last years with associated high economic impact. The high impact of drowsy driving related issues results in a strong motivation for developing a countermeasure for this problem.

The natural physiological need associated with drowsing driving cannot be altered, being the most valid solution the implementation of a drowsiness detection system in vehicles. A system that could detect drowsiness, warning the driver.

There are already several developed projects focusing on driver sleepiness detection and a growing interest from vehicle manufacturers to include such systems in their vehicles.

Most of the implemented solutions are based on features extracted from the vehicle or from driver monitoring, that is, analyze multiple key indicators of drowsiness as steering wheel patterns, eye blinks, head nodding and yawning. Other studies have implemented drowsiness detection based on physiologic signals analysis which is typically the most reliable drowsiness indicator. Most of the methods offers satisfactory results, even so there are different disadvantages associated to each method. A more robust approach is detecting drowsiness based in the fusion of several of the described methods, an option not properly explored. However, regardless of the potential of the hybrid approach, it should also be taken in account that in order to implement a drowsy detection system in a real life application the used methods must be non-intrusive.

Considering this, one of the proposed work main goals was to evaluate if intrusive drowsiness detection using a physiologic measure (EOG), considered a reliable solution, could be replaced by non-intrusive behavioral methods maintaining the detection performance. A second and more ambitious goal, to experiment new approaches that could lead to an overall improvement of drowsiness detection capacity.

Using a validated database, obtained in the scope of an experimental field study in real traffic on real roads, it was concluded that drowsiness detection performance using non-intrusive measures was equal to the performance obtained with EOG intrusive signals, thus verifying one of the enumerated goals.

As for the second objective, a first strategy involved using the previous mentioned database signals, combining behavioral methods with physiologic measures (ECG). Although the ECG alone it not very accurate, failing to detect several sleepy cases, when combining both methods an increment in performance was verified in comparison to the results obtained using only visual features, further corroborating that the most valid solution to detect drowsiness is by combining different methods. Independently of the used features, the binary class problem categorization had a substantial better accuracy than the multiclass.

An analysis of the time window length used for feature extraction also revealed that the performance of the system is not being substantially affected by this, neither significantly increasing nor decreasing the obtained results.

Applying several strategies it was possible to compare the most and least promising approaches, but no considerable improvement was obtained, leading to believe that, likely, the obtained results are the limit of what is achievable with this dataset or with the selected features.

Regarding the second strategy, as one of the limitations of the previous datasets was that there was no direct access to the raw video source, an experimental framework was developed to allow the extraction of additional drowsiness related features using a single camera recording the drivers face. The implemented non-invasive acquisition method revealed satisfactory accuracy, however, only a limited validation of the procedure as a drowsiness detector could be performed. In addition, the developed approach still needs further improvement, for example, to better generalize in different light conditions, as during night, which might only be feasible with a specific camera.

5.2 Future work

Although the proposed goals were achieved, demonstrating that non-intrusive measures have a comparable performance to physiologic signals and experimenting hybrid methods that improved the overall accuracy, there is still a big margin to improve. This work can also be considered one initial study of the AUTOMATIVE (AUTOmatic multiMOdal drowsiness detecTIon for smart VEhicles) project. A recently approved project which will focus on the research of signal processing and machine learning techniques for driver-specific drowsiness detection in smart vehicles. Considering all the above mentioned some suggestions of future work are depicted next.

First, the detection process needs to be completely non-intrusive in order to be applied in an actual solution. Within this setting, a next step would be the complementarity of the already non-intrusive behavioral measures with a physiologic non-intrusive signal. Although with less emphasis, vehicle-based measures could also be incorporated.

Other information that also should be considered is contextual information. These factors may include the type of road, the time of the day or for how long the driver has been continuously driving. This information may further increase the performances and allow a better discrimination of what is causing the drowsiness state, a useful feature for designing a countermeasure system.

In this project, as in previous works, non-independent subject drowsiness detection systems were tested but the obtained performances are far from what is needed for a real life application. Solving this problem is a nontrivial task due to the intrinsic physiologic differences in drowsiness manifestation, but it is crucial to put an extra effort into solving this issue. For

example, assessing the use of a different ground truth method for validation, as the KSS scale was the only validation method used. Personalized algorithms to account for individual subject differences should also be considered.

In relation to the implemented method for non-invasive acquisition of drowsiness measures, as indicated, its application should be considered as it allows an easy extraction of multiple drowsiness related measures, however, further development is needed and more important validating the framework using a large dataset.

Moreover, an integrated approach of the implemented method should also be considered, as it might improve the individual accuracy of each step and overall process performance. For example, using a deep CNN to simultaneously perform face detection, pose estimation and age estimations among other measures that could potentially allow a better evaluation of the driver state.

References

[1] C. Jacobé de Naurois, C. Bourdin, A. Stratulat, E. Diaz, and J.-L. Vercher, “Detection and prediction of driver drowsiness using artificial neural network models,” *Accid. Anal. Prev.*, no. July, pp. 0–1, 2017.

[2] National Sleep Foundation, “Facts About Drowsy Driving.” 2007.

[3] N. C. for S. and Analysis, “Drowsy Driving 2015,” *Natl. Highw. Traffic Saf. Adm.*, vol. DOT HS 812, 2017.

[4] DrowsyDriving, “Facts.” [Online]. Available: http://drowsydriving.org/about/. [Accessed: 13-Jan-2018].

[5] VentureBeat, “Autonomous cars could end drowsy driving, but don’t hold your breath.” [Online]. Available: https://venturebeat.com/2017/11/19/autonomous-cars-could-end-drowsy-driving-but-dont-hold-your-breath/. [Accessed: 14-Jan-2018].

[6] N. H. T. S. Administration, “Automated Vehicles for Safety.” [Online]. Available: https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety. [Accessed: 22-May-2018].

[7] M. Walch, K. Mühl, J. Kraus, T. Stoll, M. Baumann, and M. Weber, “Automotive User Interfaces,” pp. 273–294, 2017.

[8] A. Colic, O. Marques, and B. Furht, *Driver Drowsiness Detection Systems and Solutions*. 2014.

[9] J. Kim and H. Shin, *Algorithm & SoC Design for Automotive Vision Systems*. 2014.

[10] A. Sahayadhas, K. Sundaraj, and M. Murugappan, “Detecting driver drowsiness based on sensors: A review,” *Sensors (Switzerland)*, vol. 12, no. 12, pp. 16937–16953, 2012.

[11] R. Feng, G. Zhang, and B. Cheng, “An on-board system for detecting driver drowsiness based on multi-sensor data fusion using dempster-shafer theory,” *Proc. 2009 IEEE Int. Conf. Networking, Sens. Control. ICNSC 2009*, pp. 897–902, 2009.

[12] M. Ingre, T. Åkerstedt, B. Peters, A. Anund, and G. Kecklund, “Subjective sleepiness, simulated driving performance and blink duration: Examining individual differences,” *J. Sleep Res.*, vol. 15, no. 1, pp. 47–53, 2006.

[13] TelematicsNews, “Bosch supplies Drowsiness Detection System for VW Passat Alltrack.” [Online]. Available: http://telematicsnews.info/2012/03/30/bosch-supplies-drowsiness-detection-system-for-vw-passat-alltrack_m3301/. [Accessed: 16-Nov-2017].

[14] Mbusa, “Attention Assist - Vehicle Safety Video | Mercedes-Benz.” [Online]. Available: https://www.mbusa.com/mercedes/technology/videos/detail/title-safety/videoId-710835ab8d127410VgnVCM100000ccec1e35RCRD. [Accessed: 16-Nov-2017].

[15] A. Mashko, “Review of approaches to the problem of driver fatigue and drowsiness,” *2015 Smart Cities Symp. Prague, SCSP 2015*, 2015.

[16] Q. Ji, Z. Zhu, and P. Lan, “Real-time nonintrusive monitoring and prediction of driver

fatigue," *IEEE Trans. Veh. Technol.*, vol. 53, no. 4, pp. 1052–1068, 2004.

[17] 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. [Accessed: 16-Nov-2017].

[18] nvidia, "NVIDIA DRIVE IX Self-Driving Vehicles | NVIDIA Automotive." [Online]. Available: https://www.nvidia.com/en-us/self-driving-cars/drive-ix/. [Accessed: 16-Nov-2017].

[19] B. G. Pratama, I. Ardiyanto, and T. B. Adji, "A review on driver drowsiness based on image, bio-signal, and driver behavior," *Proceeding - 2017 3rd Int. Conf. Sci. Technol. ICST 2017*, pp. 70–75, 2017.

[20] H. B. Kang, "Various approaches for driver and driving behavior monitoring: A review," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 616–623, 2013.

[21] "Cardio-id." [Online]. Available: https://www.cardio-id.com/. [Accessed: 19-Jan-2018].

[22] plessey, "Plessey Semiconductors WARDEN ECG monitoring system." [Online]. Available: http://www.plesseysemiconductors.com/products/warden/. [Accessed: 16-Nov-2017].

[23] "stopsleep." [Online]. Available: https://www.stopsleep.co.uk. [Accessed: 15-Jan-2018].

[24] SleepFoundation, "Why Do We Need Sleep?" [Online]. Available: https://sleepfoundation.org/excessivesleepiness/content/why-do-we-need-sleep. [Accessed: 13-Feb-2018].

[25] J. M. Siegel, "Sleep viewed as a state of adaptive inactivity," vol. 10, pp. 747–753, 2009.

[26] V. Brodbeck *et al.*, "EEG microstates of wakefulness and NREM sleep," *Neuroimage*, vol. 62, no. 3, pp. 2129–2139, 2012.

[27] M. Johns, "Rethinking the assessment of sleepiness," *Sleep Med. Rev.*, vol. 2, no. 1, pp. 3–15, 1998.

[28] T. Bojić, A. Vuckovic, and A. Kalauzi, "Modeling EEG fractal dimension changes in wake and drowsy states in humans-a preliminary study," *J. Theor. Biol.*, vol. 262, no. 2, pp. 214–222, 2010.

[29] A. Chowdhury, R. Shankaran, M. Kavakli, and M. M. Haque, "Sensor Applications and Physiological Features in Drivers' Drowsiness Detection: A Review," *IEEE Sens. J.*, vol. 18, no. 8, pp. 3055–3067, 2018.

[30] C. Fors *et al.*, "Camera-based sleepiness detection: final report of the project SleepEYE.," *Vip Publ.*, 2011.

[31] D. F. DINGES, "An overview of sleepiness and accidents," *Journal of Sleep Research*, vol. 4. pp. 4–14, 1995.

[32] J. F. May and C. L. Baldwin, "Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 12, no. 3, pp. 218–224, 2009.

[33] SleepFoundation, "Jet-Lag and sleep." [Online]. Available: https://sleepfoundation.org/sleep-topics/jet-lag-and-sleep. [Accessed: 13-Jan-2018].

[34] E. Jewett, D.-J. Dijk, E. Kronauer, and D. F. Dinges, "Dose-response relationship between sleep duration and human psychomotor and subjective alertness," *Sleep*, vol. 20, no. 1, pp. 171–179, 1997.

[35] P. A. Desmond and P. A. Hancock, "Active and passive fatigue states," *Hum. factors Transp. Stress. Workload. fatigue (pp. 455-465)*, 2001.

[36] P. T. Gimeno, G. P. Cerezuela, and M. C. Montanes, "On the concept and measurement of driver drowsiness, fatigue and inattention: implications for countermeasures," *Int.*

J. Veh. Des., vol. 42, no. 1/2, p. 67, 2006.

[37] W. Vanlaar, H. Simpson, D. Mayhew, and R. Robertson, "Fatigued and drowsy driving: A survey of attitudes, opinions and behaviors," *J. Safety Res.*, vol. 39, no. 3, pp. 303–309, 2008.

[38] T. Kundinger, A. Riener, N. Sofra, and A. Ag, "A Robust Drowsiness Detection Method based on Vehicle and Driver Vital Data Sate of the Art and Research Challenges," no. September, 2017.

[39] M. Awais, N. Badruddin, and M. Drieberg, "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.

[40] C. C. Liu, S. G. Hosking, and M. G. Lenné, "Predicting driver drowsiness using vehicle measures: Recent insights and future challenges," *J. Safety Res.*, vol. 40, no. 4, pp. 239–245, 2009.

[41] C. A. A. I Elaine and Seaman, "Likert Scales and Data Analyses," *Quality Progress*, vol. 40, no. 7. pp. 64–65, 2007.

[42] K. Kaida, M. Takahashi, A. Nakata, Y. Otsuka, T. Haratani, and K. Fukasawa, "Validation of the Karolinska sleepiness scale against performance and EEG variables," *Clin. Neurophysiol.*, vol. 117, no. 7, pp. 1574–1581, 2006.

[43] J. A. Horne and S. D. Baulk, "Awareness of sleepiness when driving," *Psychophysiology*, vol. 41, no. 1, pp. 161–165, 2004.

[44] "Stanford Sleepiness Scale." [Online]. Available: http://www.stanford.edu/~dement/sss.html. [Accessed: 12-Jan-2017].

[45] M. E. Wewers and N. K. Lowe, "A critical review of visual analogue scales in the measurement of clinical phenomena," *Res. Nurs. Health*, vol. 13, no. 4, pp. 227–236, 1990.

[46] M. W. Johns, "The Epworth Sleepiness Scale." [Online]. Available: http://epworthsleepinessscale.com/. [Accessed: 12-Jan-2018].

[47] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, "Intelligent driver drowsiness detection system using uncorrelated fuzzy locality preserving analysis," *IEEE Int. Conf. Intell. Robot. Syst.*, pp. 4608–4614, 2011.

[48] B. Keal and T. Caldwell, "Harman demonstrates pupil-based driver monitoring system," *Biometric Technol. Today*, vol. 2016, no. 2, p. 2-, 2016.

[49] S. H. Fairclough and R. Graham, "Impairment of Driving Performance Caused by Sleep Deprivation or Alcohol: A Comparative Study," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 41, no. 1, pp. 118–128, 1999.

[50] S. Otmani, T. Pebayle, J. Roge, and A. Muzet, "Effect of driving duration and partial sleep deprivation on subsequent alertness and performance of car drivers," *Physiol. Behav.*, vol. 84, no. 5, pp. 715–724, 2005.

[51] G. O'Hanlon, J. and Kelly, "A psycho-physiological evaluation of devices for preventing lane drift and run-off-road accidents," *Tech. Rep. 1736-F, Hum. Factors Res. Inc., St. Barbar. Res. Park. Goleta, Calif.*, 1974.

[52] P. Choudhary and N. R. Velaga, "Analysis of vehicle-based lateral performance measures during distracted driving due to phone use," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 44, pp. 120–133, 2017.

[53] 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," *Behav. Res. Methods*, pp. 1–14, 2017.

[54] E. Vural, M. Bartlett, G. Littlewort, M. Cetin, A. Ercil, and J. Movellan, "Discrimination of moderate and acute drowsiness based on spontaneous facial expressions," *Proc. - Int.*

Conf. Pattern Recognit., pp. 3874–3877, 2010.

[55] T. Nakamura, A. Maejima, and S. Morishima, “Driver Drowsiness Estimation from Facial Expression Features,” *Proc. 9th Int. Conf. Comput. Vis. Theory Appl.*, pp. 207–214, 2014.

[56] S. Hachisuka, “Human and vehicle: Driver drowsiness detection by facial expression,” *Proc. - 2013 Int. Conf. Biometrics Kansei Eng. ICBAKE 2013*, pp. 320–326, 2013.

[57] E. Vural, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett, and J. Movellan, “Drowsy Driver Detection Through Facial Movement Analysis,” *Human–Computer Interact.*, pp. 6–18.

[58] S. Abtahi, B. Hariri, and S. Shirmohammadi, “Driver Drowsiness Monitoring Based on Yawning Detection,” *IEEE Int. Instrum. Meas. Technol. Conf.*, pp. 1–4, 2011.

[59] B. N. Manu, “Facial features monitoring for real time drowsiness detection,” *Proc. 2016 12th Int. Conf. Innov. Inf. Technol. IIT 2016*, pp. 78–81, 2017.

[60] I. G. Daza *et al.*, “Drowsiness monitoring based on driver and driving data fusion,” *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC*, pp. 1199–1204, 2011.

[61] W. Sheng, Y. Ou, D. Tran, E. Tadesse, M. Liu, and G. Yan, “An integrated manual and autonomous driving framework based on driver drowsiness detection,” *IEEE Int. Conf. Intell. Robot. Syst.*, pp. 4376–4381, 2013.

[62] W. Han, Y. Yang, G. Bin Huang, O. Sourina, F. Klanner, and C. Denk, “Driver Drowsiness Detection Based on Novel Eye Openness Recognition Method and Unsupervised Feature Learning,” *Proc. - 2015 IEEE Int. Conf. Syst. Man, Cybern. SMC 2015*, pp. 1470–1475, 2016.

[63] T. Danisman, I. M. Bilasco, C. Djeraba, and N. Ihaddadene, “Drowsy driver detection system using eye blink patterns,” *2010 Int. Conf. Mach. Web Intell. ICMWI 2010 - Proc.*, pp. 230–233, 2010.

[64] S. Darshana, D. Fernando, S. Jayawardena, S. Wickramanayake, and C. De Silva, “Efficient PERCLOS and gaze measurement methodologies to estimate driver attention in real time,” *Proc. - Int. Conf. Intell. Syst. Model. Simulation, ISMS*, vol. 2015–Septe, pp. 289–294, 2015.

[65] J. F. Xie, M. Xie, and W. Zhu, “Driver fatigue detection based on head gesture and PERCLOS,” *2012 Int. Conf. Wavelet Act. Media Technol. Inf. Process. ICWAMTIP 2012*, pp. 128–131, 2012.

[66] B. Akrout and W. Mahdi, “Hypovigilance Detection Based on Eyelids Behavior Study,” *iJES*, vol. 1, no. 1, pp. 39–45, 2013.

[67] 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.

[68] R. J. C. Lawrence Barr, Heidi Howarth, Stephen Popkin, “A review and evaluation of emerging driver fatigue detection measures and technologies,” pp. 1–5, 2004.

[69] L. M. Bergasa and J. Nuevo, “Real-time system for monitoring driver vigilance,” *IEEE Int. Symp. Ind. Electron.*, vol. III, no. 1, pp. 1303–1308, 2005.

[70] I. G. Daza, L. M. Bergasa, S. Bronte, J. Javier Yebes, J. Almazán, and R. Arroyo, “Fusion of optimized indicators from advanced driver assistance systems (ADAS) for driver drowsiness detection,” *Sensors (Switzerland)*, vol. 14, no. 1, pp. 1106–1131, 2014.

[71] I. H. Choi and Y. G. Kim, “Head pose and gaze direction tracking for detecting a drowsy driver,” *Appl. Math. Inf. Sci.*, vol. 9, no. 2, pp. 505–512, 2015.

[72] “Motivation For Drowsiness Detection Information Technology Essay,” *Essays, UK. (November 2013)*. [Online]. Available: https://www.ukessays.com/essays/information-technology/motivation-for-drowsiness-detection-information-technology-essay.php?cref=1.

[73] E. Murphy-Chutorian and M. M. Trivedi, “Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness,” *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 300–311, 2010.

[74] A. Doshi and M. M. Trivedi, “On the roles of eye gaze and head dynamics in predicting driver’s intent to change lanes,” *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 3, pp. 453–462, 2009.

[75] B. Mandal, L. Li, G. S. Wang, and J. Lin, “Towards Detection of Bus Driver Fatigue Based on Robust Visual Analysis of Eye State,” *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 3, pp. 545–557, 2017.

[76] V. K. Diddi and S. B. Jamge, “A Vision Based System for Monitoring the Loss of Attention in Automotive Drivers,” vol. 4, no. 7, pp. 2013–2016, 2015.

[77] K. Y. Huynh XP., Park SM., “Detection of Driver Drowsiness Using 3D Deep Neural Network and Semi-Supervised Gradient Boosting Machine,” *Comput. Vis. – ACCV 2016 Work.*, vol. 10118, 2017.

[78] H. C. Shih TH., “MSTN: Multistage Spatial-Temporal Network for Driver Drowsiness Detection,” *Comput. Vis. – ACCV 2016 Work.*, vol. 10118, 2017.

[79] S. D. Lin, J.-J. Lin, and C.-Y. Chung, “Sleepy Eye’s Recognition for Drowsiness Detection,” *2013 Int. Symp. Biometrics Secur. Technol.*, pp. 176–179, 2013.

[80] R. O. Mbouna, S. G. Kong, and M. G. Chun, “Visual analysis of eye state and head pose for driver alertness monitoring,” *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1462–1469, 2013.

[81] M. Sacco and R. Farrugia, “Driver fatigue monitoring system using support vector machines,” *5th Int. Symp. Commun. Control Signal Process.*, pp. 1–5, 2012.

[82] N. Alioua, A. Amine, A. Rogozan, A. Bensrhair, and M. Rziza, “Driver head pose estimation using efficient descriptor fusion,” *Eurasip J. Image Video Process.*, vol. 2016, no. 1, 2016.

[83] T. D’Orazio, M. Leo, C. Guaragnella, and A. Distante, “A visual approach for driver inattention detection,” *Pattern Recognit.*, vol. 40, no. 8, pp. 2341–2355, 2007.

[84] Z. Zhang and J. Zhang, “A new real-time eye tracking based on nonlinear unscented Kalman filter for monitoring driver fatigue,” *J. Control Theory Appl.*, vol. 8, no. 2, pp. 181–188, 2010.

[85] M. Saradadevi, “Driver Fatigue Detection Using Mouth and Yawning Analysis,” *Int. J. Comput. Sci. Netw. Secuirity*, vol. 8, no. 6, pp. 183–188, 2008.

[86] L. Hartley, T. Horberry, N. Mabbott, and G. P. Krueger, “Review of Fatigue Detection and Prediction Technologies,” no. September, pp. 1–41, 2000.

[87] L. Zhang, F. Liu, and J. Tang, “Real-Time System for Driver Fatigue Detection by RGB-D Camera,” *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 2, pp. 1–17, 2015.

[88] J. Vicente, P. Laguna, A. Bartra, and R. Bailón, “Drowsiness detection using heart rate variability,” *Med. Biol. Eng. Comput.*, vol. 54, no. 6, pp. 927–937, 2016.

[89] M. Mahachandra, Yassierli, I. Z. Sutalaksana, and K. Suryadi, “Sensitivity of heart rate variability as indicator of driver sleepiness,” *2012 Southeast Asian Netw. Ergon. Soc. Conf. Ergon. Innov. Leveraging User Exp. Sustain. SEANES 2012*, pp. 0–5, 2012.

[90] M. Patel, S. K. L. Lal, D. Kavanagh, and P. Rossiter, “Applying neural network analysis on heart rate variability data to assess driver fatigue,” *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7235–7242, 2011.

[91] imotions, “Measuring the Heart – How do ECG and PPG Work?” [Online]. Available: https://imotions.com/blog/measuring-the-heart-how-does-ecg-and-ppg-work/. [Accessed: 13-Jan-2018].

[92] S. Silveira, “Driver’s Fatigue State Monitoring using Physiological Signals,” FEUP, 2017.

[93] G. Rigas, Y. Goletsis, P. Bougia, and D. I. Fotiadis, “Towards driver’s state recognition on real driving conditions,” *Int. J. Veh. Technol.*, vol. 2011, 2011.

[94] J. A. Healey and R. W. Picard, “Detecting stress during real-world driving tasks using physiological sensors,” *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, 2005.

[95] A. Murata and Y. Hiramatsu, “Evaluation of Drowsiness by HRV Measures : Proposal of prediction method of low arousal state,” *5th Int. Work. Comput. Intell. Appl. Proc. IWCIA 2009*, no. 11, pp. 348–353, 2009.

[96] J. Pan and W. J. Tompkins, “A Real-Time QRS Detection Algorithm,” *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, 1985.

[97] M. Tasaki, M. Sakai, M. Watanabe, H. Wang, and D. Wei, “Evaluation of drowsiness during driving using electrocardiogram - A driving simulation study,” *Proc. - 10th IEEE Int. Conf. Comput. Inf. Technol. CIT-2010, 7th IEEE Int. Conf. Embed. Softw. Syst. ICESS-2010, ScalCom-2010*, no. Cit, pp. 1480–1485, 2010.

[98] H. De Rosario, J. S. Solaz, N. Rodríguez, and L. M. Bergasa, “Controlled inducement and measurement of drowsiness in a driving simulator,” *IET Intell. Transp. Syst.*, vol. 4, no. 4, p. 280, 2010.

[99] P. Campisi, D. La Rocca, and G. Scarano, “EEG for automatic person recognition,” *Computer (Long. Beach. Calif).*, vol. 45, no. 7, pp. 87–89, 2012.

[100] A. Lal, S. K. L. and Craig, “A critical review of the psychophysiology of driver fatigue. Biological Psychology,” vol. 55, no. 2001, pp. 173–194, 2001.

[101] N. Gurudath and H. Bryan Riley, “Drowsy driving detection by EEG analysis using Wavelet Transform and K-means clustering,” *Procedia Comput. Sci.*, vol. 34, pp. 400–409, 2014.

[102] M. Ben Dkhil, A. Wali, and A. M. Alimi, “Drowsy Driver Detection by EEG Analysis Using Fast Fourier Transform,” pp. 313–318, 2015.

[103] M. Rivera, “Monitoring of Micro-sleep and Sleepiness for the Drivers Using EEG Signal,” 2013.

[104] M. Akin, M. B. Kurt, N. Sezgin, and M. Bayram, “Estimating vigilance level by using EEG and EMG signals,” *Neural Comput. Appl.*, vol. 17, no. 3, pp. 227–236, 2008.

[105] B. G. Lee and W. Y. Chung, “Driver alertness monitoring using fusion of facial features and bio-signals,” *IEEE Sens. J.*, vol. 12, no. 7, pp. 2416–2422, 2012.

[106] D. Malathi, J. D. Dorathi Jayaseeli, S. Madhuri, and K. Senthilkumar, “Electrodermal Activity Based Wearable Device for Drowsy Drivers,” *J. Phys. Conf. Ser.*, vol. 1000, no. 1, 2018.

[107] R. Ranjan, S. Sankaranarayanan, C. D. Castillo, and R. Chellappa, “An All-In-One Convolutional Neural Network for Face Analysis,” *Int. Conf. Autom. Face Gesture Recognit.*, pp. 17–24, 2017.

[108] M. Rimini-doering, T. Altmueller, U. Ladstaetter, and M. Rossmeier, “Effects of lane departure warning on drowsy drivers’ performance and state in a simulator,” *Proc. Third Int. Driv. Symp. Hum. Factors Driv. Assessment, Train. Veh. Des.*, pp. 88–95, 2005.

[109] Y. Cui and D. Wu, “EEG-Based Driver Drowsiness Estimation Using Convolutional Neural Networks,” *Neural Inf. Process.*, vol. 10635, 2017.

[110] X. Zhu, W. L. Zheng, B. L. Lu, X. Chen, S. Chen, and C. Wang, “EOG-based drowsiness detection using convolutional neural networks,” *2014 Int. Jt. Conf. Neural Networks*, pp. 128–134, 2014.

[111] 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,” *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 6055–6058,

2011.

[112] M. Miyaji, “Method of Drowsy State Detection for Driver Monitoring Function,” *Int. J. Inf. Electron. Eng.*, vol. 4, no. 4, 2014.

[113] M. B. Kurt, N. Sezgin, M. Akin, G. Kirbas, and M. Bayram, “The ANN-based computing of drowsy level,” *Expert Syst. Appl.*, vol. 36, no. 2 PART 1, pp. 2534–2542, 2009.

[114] S. Hu and G. Zheng, “Driver drowsiness detection with eyelid related parameters by Support Vector Machine,” *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7651–7658, 2009.

[115] F. C. Lin, L. W. Ko, C. H. Chuang, T. P. Su, and C. T. Lin, “Generalized EEG-based drowsiness prediction system by using a self-organizing neural fuzzy system,” *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 59, no. 9, pp. 2044–2055, 2012.

[116] C. T. Lin, C. J. Chang, B. S. Lin, S. H. Hung, C. F. Chao, and I. J. Wang, “A real-time wireless brain-computer interface system for drowsiness detection,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 4, no. 4, pp. 214–222, 2010.

[117] C. T. Lin *et al.*, “Development of wireless brain computer interface with embedded multitask scheduling and its application on real-time driver’s drowsiness detection and warning,” *IEEE Trans. Biomed. Eng.*, vol. 55, no. 5, pp. 1582–1591, 2008.

[118] D. Tran, E. Tadesse, W. Sheng, Y. Sun, M. Liu, and S. Zhang, “A driver assistance framework based on driver drowsiness detection,” *2016 IEEE Int. Conf. Cyber Technol. Autom. Control. Intell. Syst.*, pp. 173–178, 2016.

[119] and X. Z. Bo Cheng, Wei Zhang, Yingzi Lin, Ruijia Feng, “Driver Drowsiness Detection Based on Multisource Information,” *wileyonlinelibrary.com/journal/hfm*.

[120] Bosch, “Driver drowsiness detection.” [Online]. Available: http://www.bosch-mobility-solutions.com/en/products-and-services/passenger-cars-and-light-commercial-vehicles/driver-assistance-systems/driver-drowsiness-detection/. [Accessed: 16-Nov-2017].

[121] R. B. GmbH, “Chassis Systems Control Driver Drowsiness Detection.” 2010.

[122] Valeo, “Driver Monitoring: a camera to monitor driver alertness.” [Online]. Available: http://www.valeo.com/en/driver-monitoring/. [Accessed: 16-Nov-2017].

[123] V. Safran, “Press kit Valeo - Safran,” pp. 1–20, 2015.

[124] Volvo, “V40 cross country.” [Online]. Available: https://www.volvocars.com/diplomat/cars/new-models/v40-cross-country/specifications. [Accessed: 14-Jan-2018].

[125] Volvo, “Volvo V40 manual.” United Kingdom, 2018.

[126] FORD MEDIA CENTER, “Ford Brazil’s Heavy Truck Division Created The ‘Safe Cap’, A Truly Innovative Safety Device For Truck Drivers.” [Online]. Available: https://media.ford.com/content/fordmedia/feu/gb/en/news/2017/11/06/ford-brazil_s-heavy-truck-division-created-the-safe-cap--a-truly.html. [Accessed: 28-Jan-2018].

[127] harken, “Welcome to the HARKEN project web site.” [Online]. Available: http://harken.ibv.org/. [Accessed: 19-Jan-2017].

[128] Wearvigo, “Vigo Smart Headset.” [Online]. Available: https://www.wearvigo.com/. [Accessed: 16-Nov-2017].

[129] CardioId, “What is cardiowheel?” [Online]. Available: https://www.cardio-id.com/cardiowheel. [Accessed: 21-Nov-2017].

[130] A. Anund and K. Kircher, “Advantages and disadvantages of different methods to evaluate sleepiness warning systems,” 2009.

[131] X. Wang and C. Xu, “Driver drowsiness detection based on non-intrusive metrics considering individual specifics,” *Accid. Anal. Prev.*, vol. 95, no. July, pp. 350–357,

2016.

[132] NADS, “The national advanced driving simulator.” [Online]. Available: http://www.nads-sc.uiowa.edu/sim_nads2.php. [Accessed: 16-Jan-2018].

[133] University of Leeds, “University of Leeds Driving Simulator,” 2018. .

[134] A. Anund, “Sleepiness at the wheel,” Department of Public Health Sciences NASP - National Prevention of Suicide and Mental Health Karolinska Institutet, 2009.

[135] J. R. Lewis and W. P. Beach, “Individual differences and in-vehicle distraction while driving: a test track study and psychometric evaluation,” *Proc. Hum. FACTORS Ergon. Soc. 43rd Annu. Meet.*, pp. 457–461, 1999.

[136] Z. Li, L. Chen, J. Peng, and Y. Wu, “Automatic detection of driver fatigue using driving operation information for transportation safety,” *Sensors (Switzerland)*, vol. 17, no. 6, 2017.

[137] V. L. Neale, T. A. Dingus, S. G. Klauer, and M. Goodman, “An overview of the 100-car naturalistic study and findings,” *Traffic Saf.*, no. January, pp. 1–10, 2005.

[138] UNSW Sydney, “Australian Naturalistic Driving Study.” .

[139] K. L. Campbell, “The SHRP 2 Naturalistic Driving Study,” *Transp. Res. News*, pp. 30–35, 2012.

[140] C. Spindler, “Analysis of the robustness of steering pattern based drowsiness detection,” pp. 1–11, 2014.

[141] Ford, “Stay Alert! The Ford Challenge.” [Online]. Available: https://www.kaggle.com/c/stayalert/data. [Accessed: 17-Jan-2018].

[142] 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,” *Hum. Factors*, vol. 56, no. 5, pp. 986–998, 2014.

[143] E. Aidman, C. Chadunow, K. Johnson, and J. Reece, “Real-time driver drowsiness feedback improves driver alertness and self-reported driving performance,” *Accid. Anal. Prev.*, vol. 81, pp. 8–13, 2015.

[144] J. Schmidt, C. Braunagel, W. Stolzmann, and K. Karrer-Gauss, “Driver drowsiness and behavior detection in prolonged conditionally automated drives,” *IEEE Intell. Veh. Symp. Proc.*, vol. 2016-Augus, no. Iv, pp. 400–405, 2016.

[145] I. Garcia, S. Bronte, L. M. Bergasa, N. Hernandez, B. Delgado, and M. Sevillano, “Vision-based drowsiness detector for a realistic driving simulator,” *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC*, no. October, pp. 887–894, 2010.

[146] K. Dwivedi, K. Biswaranjan, and A. Sethi, “Drowsy driver detection using representation learning,” *Souvenir 2014 IEEE Int. Adv. Comput. Conf. IACC 2014*, pp. 995–999, 2014.

[147] B. Reddy, Y.-H. Kim, S. Yun, C. Seo, and J. Jang, “Real-Time Driver Drowsiness Detection for Embedded System Using Model Compression of Deep Neural Networks,” *2017 IEEE Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 438–445, 2017.

[148] D. Liu, P. Sun, Y. Q. Xiao, and Y. Yin, “Drowsiness detection based on eyelid movement,” *2nd Int. Work. Educ. Technol. Comput. Sci. ETCS 2010*, vol. 2, pp. 49–52, 2010.

[149] Z. Li, S. E. Li, R. Li, B. Cheng, and J. Shi, “Online detection of driver fatigue using steering wheel angles for real driving conditions,” *Sensors (Switzerland)*, vol. 17, no. 3, pp. 1–12, 2017.

[150] J. Liu, C. Zhang, and C. Zheng, “EEG-based estimation of mental fatigue by using KPCA-HMM and complexity parameters,” *Biomed. Signal Process. Control*, vol. 5, no. 2, pp. 124–130, 2010.

[151] G. L. Masala and E. Grosso, “Real time detection of driver attention: Emerging solutions

based on robust iconic classifiers and dictionary of poses," *Transp. Res. Part C Emerg. Technol.*, vol. 49, pp. 32–42, 2014.

[152] "SmartEye." [Online]. Available: http://smarteye.se/. [Accessed: 19-Jan-2018].

[153] P. S. Hamilton, "Open Source ECG Analysis Software Documentation," *Comput. Cardiol.*, vol. 2002, IEEE, pp. 101–104, 2002.

[154] F. A. Carreiras C, Alves AP, Lourenço A, Canento F, Silva H, "BioSPPy - Biosignal Processing in Python," 2015. [Online]. Available: https://github.com/PIA-Group/BioSPPy/. [Accessed: 08-Jan-2018].

[155] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," *Proc. Symp. Eye Track. Res. Appl. - ETRA '00*, pp. 71–78, 2000.

[156] S. Bersch, D. Azzi, R. Khusainov, I. Achumba, and J. Ries, "Sensor Data Acquisition and Processing Parameters for Human Activity Classification," *Sensors*, vol. 14, no. 3, pp. 4239–4270, 2014.

[157] H. Candra *et al.*, "Investigation of window size in classification of EEG-emotion signal with wavelet entropy and support vector machine," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2015–Novem, pp. 7250–7253, 2015.

[158] S. B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques," *Informatica*, vol. 31, pp. 249–268, 2007.

[159] 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 Syst. Appl.*, vol. 41, no. 4 PART 1, pp. 1139–1152, 2014.

[160] SCIKIT-LEARN, "Support Vector Machines (SVM)." [Online]. Available: http://www.bogotobogo.com/python/scikit-learn/scikit_machine_learning_Support_Vector_Machines_SVM.php. [Accessed: 13-May-2018].

[161] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[162] J. Kim, E. Yim, C. Jeon, C. Jung, and B. Han, "Drowsy behavior detection based on driving information," *Int. J. …*, vol. 13, no. 2, pp. 293–300, 2012.

[163] C. Nguyen, Y. Wang, and H. N. Nguyen, "Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic," *J. Biomed. Sci. Eng.*, vol. 06, no. 05, pp. 551–560, 2013.

[164] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 42, no. 4, pp. 463–484, 2012.

[165] Y. Shin, "Application of boosting regression trees to preliminary cost estimation in building construction projects," *Comput. Intell. Neurosci.*, vol. 2015, 2015.

[166] SCIKIT-LEARN, "Nearest Neighbors Classification." [Online]. Available: http://scikit-learn.org/stable/modules/neighbors.html. [Accessed: 15-May-2018].

[167] S. Alaliyat, "Video - based Fall Detection in Elderly ' s Houses," 2008.

[168] S. Chitroub, "Classifier combination and score level fusion: Concepts and practical aspects," *Int. J. Image Data Fusion*, vol. 1, no. 2, pp. 113–135, 2010.

[169] C. Sanderson and K. K. Paliwal, "Identity verification using speech and face information," *Digit. Signal Process.*, vol. 14, no. 5, pp. 449–480, 2004.

[170] P. Bonnin, E. Pissaloux, and A. Cedex, "A new way of image data fusion : the Multi-Spectral Cooperative Segmentation," pp. 572–575, 1995.

[171] L. Xu, A. Krzyżak, and C. Y. Suen, "Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition," *IEEE Trans. Syst. Man Cybern.*, vol. 22, no.

3, pp. 418–435, 1992.

[172] S.-B. Cho and J. H. Kim, “Combining multiple neural networks by fuzzy integral for robust classification,” *IEEE Trans. Syst. Man. Cybern.*, vol. 25, no. 2, 1995.

[173] H. J. Kang, K. Kim, and J. H. Kim, “Optimal approximation of discrete probability distribution with kth-order dependency and its application to combining multiple classifiers,” *Pattern Recognit. Lett.*, vol. 18, no. 6, pp. 515–523, 1997.

[174] H. Tahani and J. M. Keller, “Information fusion in computer vision using the fuzzy integral ,” *Syst. Man Cybern.*, vol. 20, no. 3, pp. 733–741, 1990.

[175] R. Battiti and A. M. Colla, “Democracy in neural nets: Voting schemes for classification,” *Neural Networks*, vol. 7, no. 4, pp. 691–707, 1994.

[176] H. He and E. A. Garcia, “Learning from imbalanced data,” *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, 2009.

[177] R. B. Rao, S. Krishnan, and R. S. Niculescu, “Data mining for improved cardiac care,” *ACM SIGKDD Explor. Newsl.*, vol. 8, no. 1, pp. 3–10, 2006.

[178] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, “Driver inattention monitoring system for intelligent vehicles: A review,” *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 596–614, 2011.

[179] L. R. Rabiner, “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition,” *Proceedings of the IEEE*, vol. 77, no. 2. pp. 257–286, 1989.

[180] I. H. Choi, C. H. Jeong, and Y. G. Kim, “Tracking a driver’s face against extreme head poses and inference of drowsiness using a hidden Markov model,” *Appl. Sci.*, vol. 6, no. 5, 2016.

[181] E. Tadesse, W. Sheng, and M. Liu, “Driver drowsiness detection through HMM based dynamic modeling,” *2014 IEEE Int. Conf. Robot. Autom.*, pp. 4003–4008, 2014.

[182] hmmlearn, “hmmlearn-Tutorial.” [Online]. Available: http://hmmlearn.readthedocs.io/en/latest/tutorial.html. [Accessed: 16-May-2018].

[183] K. Zhang, Z. Zhang, Z. Li, S. Member, Y. Qiao, and S. Member, “Joint Face Detection and Alignment using Multi - task Cascaded Convolutional Networks,” *IEEE Signal Process. Lett.*, no. 1, pp. 1–5, 2016.

[184] H. Li, Z. Lin, X. Shen, and J. Brandt, “A Convolutional Neural Network Cascade for Face Detection,” *IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 5325–5334, 2015.

[185] H. Salman *et al.*, “Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review,” *Neural Comput.*, vol. 2449, pp. 2352–2449, 2018.

[186] M. Peng, C. Wang, T. Chen, and G. Liu, “NIRFaceNet : A Convolutional Neural Network for Near-Infrared Face Identification,” *Information*, pp. 1–14, 2016.

[187] D. Sandberg, “FaceNet,” *GitHub Repos.*, 2017.

[188] X. Cao, Y. Wei, F. Wen, and J. Sun, “Face Alignment by Explicit Shape Regression,” *IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 2887–2894, 2013.

[189] V. Kazemi and J. Sullivan, “One Millisecond Face Alignment with an Ensemble of Regression Trees,” *IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1867–1874, 2014.

[190] Intelligent Behaviour Understand Group (I-BUG), “Facial point annotations.” [Online]. Available: https://ibug.doc.ic.ac.uk/resources/facial-point-annotations/. [Accessed: 17-May-2018].

[191] D. E. King, “Dlib-ml: A Machine Learning Toolkit,” *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758.

[192] E. Murphy-Chutorian and M. M. Trivedi, “Head pose estimation in computer vision: A survey,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 4, pp. 607–626, 2009.

[193] E. R. Pi, “Implementation of a 3D pose estimation algorithm,” no. June, 2015.

[194] Y. Guobing, “Head pose estimation,” *GitHub Repos.*, 2017.

[195] M. Friess and D. Ph, “Head-and-Face Anthropometric Survey of U . S . Respirator Users,” *Natl. Inst. Occup. Saf. Heal.*, vol. 2, no. 11, 2004.

[196] G. Bradski, “The OpenCV Library,” *Dr. Dobb’s J. Softw. Tools*, 2000.

[197] D. W. Hansen and Q. Ji, “In the Eye of the Beholder: A Survey of Models for Eyes and Gaze,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 478–500, 2010.

[198] E. Wood and A. Bulling, “Eyetab: Model-based gaze estimation on unmodified tablet computers,” *Etra*, pp. 3–6, 2014.

[199] G. Balakrishnan, F. Durand, and J. Guttag, “Detecting Pulse from Head Motions in Video,” pp. 3430–3437, 2013.

[200] M.-Z. Poh, D. J. McDuff, and R. W. Picard, “Non-contact, automated cardiac pulse measurements using video imaging and blind source separation,” *Opt. Express*, vol. 18, no. 10, p. 10762, 2010.

[201] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, “Remote plethysmographic imaging using ambient light .,” *Opt. Express*, vol. 16, no. 26, pp. 21434–21445, 2008.

[202] T. Hearn, “Webcam pulse detector,” *GitHub Repos.*, 2017.

[203] Kurtjmac, “Driving Vlog - Pleasantly Pleased (March 2018),” *YouTube*, 2018. [Online]. Available: https://www.youtube.com/watch?v=F2EIceStLkc. [Accessed: 07-Apr-2018].

[204] V. Virgins, “How To Drive On The Highway - The Secrets!,” *YouTube*, 2016. [Online]. Available: https://www.youtube.com/watch?v=hBER4EIT8L0. [Accessed: 10-May-2018].

[205] T. J. VLOGS, “My Trucking Life OTR - Pennsylvania - Trip 11 Day 5,” *YouTube*, 2013. [Online]. Available: https://www.youtube.com/watch?v=xPiROinXSqQ. [Accessed: 10-May-2018].

[206] J. Shen, S. Zafeiriou, G. G. Chrysos, J. Kossaifi, G. Tzimiropoulos, and M. Pantic, “The First Facial Landmark Tracking in-The-Wild Challenge: Benchmark and Results,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2016–Febru, pp. 1003–1011, 2016.

[207] R. Ranjan, V. M. Patel, and R. Chellappa, “HyperFace: A Deep Multi-task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. XX, no. Xx, pp. 1–16, 2017.