Driver Drowsiness Detection with Peripheral Cardiac Signals

Lourenço Abrunhosa Rodrigues
lourenco.abrunhosa.rodrigues@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

December 2020

Abstract

Annually around 1.35 million people die worldwide as a result of road accidents. Of these, 90% occur because of human fault. Such faults have been continuously reduced by the development of safer road architectures and legislation that intends to guarantee the ideal conditions for driving.

However, errors made by human drivers when driving while feeling drowsy result in a constancy of people involved in road accidents, raising the need for a drowsiness detection system. A physiological signal capable of early identifying such state is the heart rate variability, which can be obtained by analysis of the consecutive time intervals between heart beats.

Using peripheral cardiac signals, signals containing cardiac rhythm information and obtained through non-intrusive ways, it is possible to integrate such detection on a vehicle without affecting the driving task.

This work builds the pipeline to use any of three wearable devices: wrist worn PPG band, ECG chest strap and off-the-person ECG collection through a steering wheel, to collect the inter beat intervals, calculate HRV features and detect the drowsiness state of a driver.

A filter was developed to compensate ambient light sensitivity of PPG based devices and the intervals detected from all signals were corrected by an algorithm created to possible wearable contact losses. SVM models with linear kernel and C=0.3 and a selected group of HRV features had good performances , reaching an average 0.62 Matthews correlation coefficient across 12 individuals. Simulator experiments showed good indication that peripheral cardiac signals can be used for drowsiness detection.

Keywords: Heart Rate Variability; Wearable; Drowsiness; Peripheral Cardiac Signals; Machine Learning

1. Introduction

Annually around 1.35 million people die worldwide as a result of road accidents[1]. Of these, 90% occur because of human fault[2]. Such faults have been continuously reduced by the development of safer road architectures and legislation that intends to guarantee the ideal conditions for driving.

However, despite all efforts, errors and distractions caused by the insistence on driving even when feeling drowsy result in a constancy of people involved in road accidents.

For this reason, it has become of the uttermost importance to develop systems capable of identifying driver drowsiness, to act with them to prevent in a more personalized and effective way this dangerous behaviour. Several proposed systems are already available in the market, but are usually based on extrinsic factors, as the simple measurement of time driving, or the monitoring of driving behaviour. Even though their implementation on the vehicle is as non-invasive as one could desire, the fact that they monitor only variables external to the driver leads to performances that fall short of what such vital system should.

On the other side, it is known that the monitoring of physiological data allows insight on the internal mechanisms that produce drowsy states, providing an excellent source of information to assess the drowsiness state of any driver. However more powerful information exists in these signals, the technology to read them normally implicates an higher level of intrusion on the drivers environment, which is why they have been kept away from this field of application.

One of the physiological signals that has revealed an interesting capability to identify an individual's drowsiness state is Heart Rate Variability (HRV), which is obtained through the analysis of the series of time intervals that separate heart beats, usually identified through QRS complexes in an Electrocardiogram (ECG). Again, the need to place chest electrodes to collect the ECG renders this approach impractical, but, fortunately, less invasive alternatives have been proposed to collect the needed information, measuring cardiac rhythm information in a more peripheral way. These non-invasive technologies combine the feasibility of being installed on a vehicle without disturbing the drivers environment, with the ability to infer their drowsiness state from an intrinsic signal, instead of possible manifestations of such state.

This way this work defines peripheral cardiac signals as the set of physiological signals that measure the cardiac rhythm dynamics in a non-invasive form, that is, which collection doesn't demand any change in drivers routine, or that in any way forces him to have his activity affected by the connection with the measuring devices.

Two pieces of equipment already available seem to meet such criteria, the *CardioWheel* by CardioID Technologies, a steering wheel cover that measures a bipolar derivation of ECG through the drivers hands, and wristbands and smart watches with an integrated Photoplethismography (PPG) sensor.

However, while both of them exceed expectations when it comes to practicability, fitting perfectly into anyone's lifestyle and driving, the distancing from the cardiac signals' primary source demands a more careful processing of these peripheral signals in order to extract information as trustworthy as that collected with thoracic electrodes.

Having this, this thesis proposes to answer two main questions, that ultimately combine to produce a drowsiness detection system for driving environments based on peripheral cardiac signals: How to deal with the filtering of such signals, and ensure confidence on the HRV features obtained from them, and if the HRV information obtained from these sources allows such drowsiness state classification as it does with thoracic ECG.

1.1. How to process peripheral signals?

Processing strategies differ depending on which equipment is used, while the *CardioWheel*® has filters built in, and so, returns a signal where the QRS complexes are immediately identifiable, most wrist band PPG sensors return a raw signal, filled with noise, movement and light change artifacts. For this reason an online filtering strategy is implemented to extract only the pulsated component of PPG.

Having the peaks on these signals corresponding to ventricular systole (ECG) and systolic pulse (PPG) peak detection algorithms are implemented to store the timestamps at which heart beats occurred. This stream of timestamps is then used to calculate the series of inter-beat intervals (IBIs) (fig. 1) that are the base of HRV calculation.

As the continuity of the peripheral cardiac signals depends on the constant contact of hands on wheel, or the absence of too strong artifacts on the PPG, it cannot be guaranteed at all moments. Thus it is possible to have missed peaks or false detections in the stream of heartbeat timestamps. To solve this, an IBI correcting algorithm is proposed, and validated, artificially removing or adding peaks and evaluating the error remaining after correction.

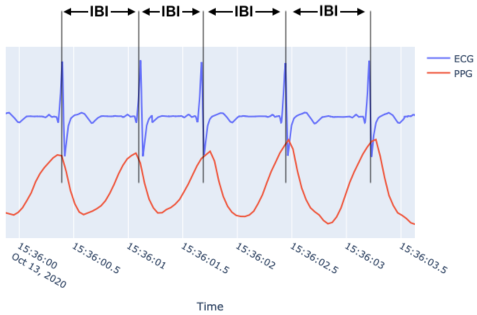


Figure 1: Representation of Inter-Beat Intervals on ECG and PPG signals. IBIs are the interval of time that separates two consecutive heart beats.

It is to note that the wearable nature of these signal's source results in that continuous segments may have duration a lot shorter than what state-of-the-art HRV use, even with correction. Having in mind that short-term HRV (sHRV) need a minimum 5 minutes of uninterrupted HR, and, in wearables, 2-3 minutes would be a much more realistic projection, the analysis conducted must be redirected to the field of ultra-short HRV (usHRV), and, to do so, the validity of features in this ultra short scope will be assessed before using them in classification models.

1.2. How to classify drowsiness from peripheral signals derived HRV?

To start building a classification model on this subject, an already existing dataset containing naturalistic driving data, with both ECG measurement and drowsiness annotations is used to evaluate machine learning algorithms in their capacity to correctly output drowsiness alarms from usHRV features. This database is also used to evaluate the need for class balancing, feature selection and alternative training strategies.

After defining the optimal models and training procedure, data collected with a driving simulator (fig. 2) developed by CardioID/ISEL is used to evaluate the performance of such models when using peripheral signals as data source. This dataset contains drowsiness annotations, hands ECG from *CardioWheel*, wrist PPG measured with the pulseOn wrist band, and chest ECG from a Movesense chest-band. Positive results in this section establish a system using peripheral cardiac signals only to detect drowsiness in drivers, combining the non-invasive advantages of wearables and built-in vehicle systems, with the deep insight physiological signals provide.

Figure 2: Driving simulator setup: A computer simulated environment is presented in the screen, while the driver controls a vehicle using the pedals and Cardiowheel. The simulator not only integrates the inputs to run the environment, but also aggregates inputs from the wheel movements, CardioWheel sensor, and intel realssense camera to a database.

2. Background

Automobile field is one where the need for insight on a person's internal state is becoming already a main topic among researchers. Through the need to reduce road accidents and the unavoidable evidence that most incidents are caused by drivers. According to a report from the Department for Transport[3] (UK) in 2008 3% of fatal accidents and 2% of those that result in serious injury had fatigue as a contributory factor. However, previous research pointed out larger number, namely 10% of collisions [4] and 17% of road crashes[5] that result in injuries and death being sleep related. Its is plausible that fatigue as an accident cause is underestimated in official reports, given the lack of specific formation of police agents to assess its contribution when reporting accidents and the fact that drowsy drivers involved in crashes tend to be wide awake when interviewed because of the induced stress[6]. This underestimation predicts that a much more realistic statistic would be that around 20% of all road accidents are caused by fatigue, either by actually falling asleep on the wheel and or by the decreased performance that it implies.

According to an European Road Safety Observatory report from 2018[7], driving while sleepy or fatigued has a prevalence much higher than what would be expected or even minimally safe, surveys demonstrate that more than a half of the population drives while being drowsy at least once a year, with a range of 10%-40% of them having actually fallen asleep on the wheel. Also, studies from the united states corroborate these results, as about one third of the population feels impaired to perform their daily tasks at least once on a monthly basis[8], which included severe reduction in driving performance. The same report states that fatigue related accidents result in high level injuries, and reaffirms the 20% prevalence of fatigue as a crash contributor. Finally, different studies focused on measuring the increased risk resulting from driving while drowsy, finding the risk to be involved in a car crash to be 4 to 14 times higher than for rested individuals[9, 10, 11].

Other studies[12, 13], directly related stress with poorer driving performance, with the observation that, if on one side, aggressive-coping stressed drivers tend to overtake other vehicles more often and in a more error prone manner, while non-aggressive, but driving disliking drivers tend to be more cautious even though they presented less control. Moreover, following a similar procedure, but coupling it with measurement of reaction times, Różanowski[14] established a positive correlation between perceived stress and poorer task performance in driving environments, which was even more evident in aggressive-coping drivers.

Seeing this, it becomes clear that a need for measures that reduce these human factors preponderance in road accidents is increasingly important. The most obvious course of action would be to remove the human from the driving process, which is already a market direction in the form of autonomous vehicles. However, full implementation of this technology will not occur in the next 30 years, which means that other strategies are needed[15]. Those are the creation of systems that monitor and act upon the stress/fatigue state of the driver. This is advantageous because the only question to be solved is how to measure the internal state of the individual. Mainly three different branches on this subject appear:

- **Driving behaviour -** Analysing driving patterns, as angle of steering wheel corrections, lane positioning, etc.
- **Tracking face position and eye direction -** The position and movements of the head and eyes present characteristic patterns depending on fatigue/stress state.
- **Physiological signals -** Physiological manifestations are the direct result of internal state variations in an individual.

While the first two have very high performances in detecting drowsy individuals, it does so when their state is so clear, that dangerous actions might already be happening. The physiological signals, even though more difficult to process, can detect drowsiness early enough to prevent any unsafe action by the driver.

2.1. HRV

HRV is the study of variation of consecutive heart beat intervals over time. This type of observation provides a window to perceive the balance between systems responsible for the modulation of cardiac rhythm, namely the Autonomous Nervous System (ANS) sympathetic and parasympathetic systems and the identification of anomalous intervals that can be correlated with cardiac disease. Because these clues may not happen at all times, but only in specific periods of the day, the need for continuous monitoring of heart rate appeared, carrying with it the time consuming task of analysing the accumulated data. This resulted in the appearance of computational methods to form indexes that would condensate all the observed data and point out if some worrying information is present.[16]

2.2. HRV and ANS

For this work it is specially important to establish a relation between HRV and the balance between sympathetic and parasympathetic systems, and to understand how different psychological states influence that balance.

As stated in [17], the Autonomic Nervous System (ANS) is constituted by two antagonist systems: the sympathetic, that in a general form prepares the organism for energy expenditure and stress response, and the parasympathetic, that returns the body to its basal, relaxed state.

Sympathetic system, also referred to as the ”fight or flight” system, produces a series of alterations in the body,such as vasodilatation of coronary arteries and vasoconstriction of other vessels, as well as increase in heart rate. This optimizes cardiac output and oxygenation of muscles, that must be optimally active to respond to the stress source.

Contrarily, parasympathetic system will promote a restful state, dilating peripheral circulation and slowing down the heart rate, so that other systemic functions, such as digestion and lachrymal, saliva, urine and fecal secretion take place. For that reason it is called as the ”rest and digest” system.

In healthy subjects, both branches of ANS balance each other, with sympathetic predominance representing an active acceleration of Heart Rate (HR) and the parasympathetic dominance a passive return to the basal state, with consequent deceleration of (HR). These effects are observable in HRV study, specially through frequency domain parameters, as their continuous balancing process produces oscillations at defined frequency ranges.

It is widely accepted in the scientific community that two different bands of spectral analysis of HRV correlate with distinct activity levels of ANS branches[16, 18]. Specifically, high frequencies (0.15 to 0.40 Hz) are commonly related to parasympathetic activity, while low frequencies (0.04 - 0.15 Hz) can be related to a mixture of both activities, or, as some researchers propose[18], sympathetic activity alone if frequency band powers are in normalized units. This allows the direct evaluation of ANS state through HRV as a non invasive form of accessing a person’s internal state.

Other HRV indexes provide additional information, and can be divided in three main groups, depending on the type of analysis needed for its calculation: Time domain, frequency domain and nonlinear domain.

3. Methodology

3.1. Signal Processing

To answer the first question this work proposed, it is necessary to extract the IBIs from each devices’ signal. While the ECG based signals allowed direct application of methods as the Pan-Tompkins[19] to extract R peak locations, and the CardioWheel already provided calculated intervals, the PPG signal suffered from high sensitivity to ambient light conditions, creating step like artifacts that needed to be eliminated. For that, a filter based on moving average removal was designed, with its coefficients calculated accordingly to equation 1.

$$\begin{aligned} y[n] =& x[n] - \frac{3}{W}\sum_{k=-\frac{W}{2}}^{W/2} x[n+k] + \\ &+ \frac{3}{W^2}\sum_{k=-\frac{W}{2}}^{W/2}\sum_{v=-\frac{W}{2}}^{W/2} x[x+k+v] - \\ &- \frac{1}{W^3}\sum_{k=-\frac{W}{2}}^{W/2}\sum_{v=-\frac{W}{2}}^{W/2}\sum_{u=-\frac{W}{2}}^{W/2} x[n+k+v+u] \end{aligned} \tag{1}$$

Where W is window length, and should equal the sampling rate of the signal.

After filtering, an adaptive threshold peak detection algorithm was implemented to locate the PPG peaks. The algorithm parameters, threshold slope (eq.3) and refractory period (eq.4) allowed the detection of peaks even when the amplitude of successive pulses varied, and the discard of false peaks that occured too soon to be a physiologically valid PPG peak.

$$Thrs[n] = Last_y - slope \cdot (n - Last_x) \tag{2}$$

$$slope_{n+1} = -\frac{0.5 * last_y}{F_s * IBI_n} \tag{3}$$

$$RP_{n+1} = 0.6 \cdot IBI_n \tag{4}$$

3.2. IBI corrector

It is a common practice to filter IBI values before performing HRV analysis, as outliers can deviate the variability indexes from their true value, hindering any further conclusions about the recorded signals. In most cases, IBI revision is made by simply eliminating non-physiological intervals, such as the ones shorter than 300ms (above 200 BPM) and those longer than 1500ms (below 40 BPM). Other approaches even define boundaries to how much consecutive IBI can differ, discarding those that cross so defined thresholds. However, while this outlier elimination strategy improves results in conventional HRV time windows, where several hundreds and thousands of IBI are available, windows as short as two minutes may not be able to afford the information loss by discarding outliers.

For this reason, this work proposes a system capable of not only identify outliers, but also of reconstructing the real IBI values from signal corrupted with outliers, combining the reliability of HRV measures based on only physiological values, but also maintaining all the available information, so that the analysis is not compromised by scarcity of data.

This system is based on the ratio between consecutive IBI. Outliers are defined as points which ratio crosses a defined threshold. This forms a basis to evaluate streams of IBI using the same sets of criteria, regardless of the absolute values present in any record.

To define the thresholds and test the performance of the corrector, ECG records from the naturalistic driving experiment SleepEye[20] were used. Signal from all available records was visually inspected, and all segments with clean and correctly identified R peaks were converted into series of IBIs. This way, a dataset of validated IBI streams was ready to evaluate the corrector, consisting of 138 intact series, containing 231470 consecutive pairs of IBIs in total.

This data was firstly used to justify a selection of a lower threshold of 0.8, and a upper threshold of 1.5. These limits were designed to be in accordance with previous research indicating that variations larger than 20% between IBIs indicated outliers[21]. The use of ratios also allows identification of how many heart-beats were missed in cases of longer IBI. Rounding the ratio to the nearest integer would return the number of heart beats encapsulated in the same outlier IBI.

These defined thresholds identified only 0.02% (53) IBI as outliers.

There are three main functions the corrector has to implement: to track the level of reliability of each new IBI, and, having identified an outlier, decide whether to fill a gap, our join two smaller intervals into a physiological value.

Four thresholds are defined, physiological bounds for IBI values, and ratio limits for normal inter IBI variation:

- I_{inf} - Inferior limit of physiological IBI, set to 300ms.
- I_{sup} - Superior limit of physiological IBI, set to 1500ms.
- r_{inf} - Lower bound of accepted ratio, set to 0.8.
- r_{sup} - Upper bound of accepted ratio, set to 1.5.

The algorithm rounds float ratios to the nearest integer. This behaviour is used so that ratios larger than r_{sup}, there is a estimation of how many real IBI were skipped to produce the larger outlier. As an example, with $r_{sup} = 1.5$ and a quotient of 2.8, the system would be capable of realizing that most likely 3 IBIs, instead of only 2, were concatenated into a single value.

The main process consists of consuming a value of a waiting list of IBI, and, by deeming its corresponding ratio to the previously accepted value, decide whether to directly add it to the validated results, to fill detected gaps or to sum it to an adjacent interval.

Before starting this process, and any time the corrector needs to be reset, the corrector must initialize the `pending` list and the `last` value. To do so, the corrector extracts the first IBI from the input and checks if it is inside the physiological range. If so, that value is defined as `last`, and the rest of the available IBIs form the `pending` list.

To fill detected gaps, a series of estimates for the missing IBIs are computed, using the mean value between the partition of the longer interval and the last accepted value (eq. 5).

$$new = \left(\frac{Value}{ratio} + last\right)/2 \tag{5}$$

This is done to simulate a smooth evolution from the last accepted IBI and the partition length needed to have a detected heart beat at the timestamp corresponding to the longer outlier. It allows smooth shortening or widening of estimated intervals to accommodate outliers that are not integer multiples of the last valid value, instead of having a sudden jump to a series of identical partitions of that outlier.

After defining a filling value, the outlier gets this estimated value subtracted from it, and is replaced at the beginning of the `pending` list to proceed the evaluation, if the remaining value continues to be large enough to be an outlier, the filling process is repeated.

Finally, if the system detects a shorter interval, it tries to join it with an adjacent value. This serves to correct instances where a false peak was detected, leading to a real IBI divided into two parts. The corrector chooses the smallest value between the previous (`last`) and following intervals, and adds the outlier to it. Finally, it checks that such addition does not result in a ratio above r_{sup}, if it does, the shorter interval is added without any processing, as it would mean that it did not correspond to a partitioned IBI.

To make sure that the system always produces realistic estimates, in regards to physiology, any time a proposed value reports a normal ratio, but a non-physiological value, the corrector is reset by running the initialize process on the current `pending` list.

3.3. Detecting Drowsiness with HRV features

To establish a set of models capable of correctly identify dangerous states of drowsiness, machine learning strategies were implemented, using a dataset from a previous naturalist driving study, SleepEYE[20]. This study consisted of 20 individuals who had their ECG measured and Karolinska Sleepiness Scale (KSS) self-report annotated during 90 minute drives in public roads in Sweden. Using these measurements, ECG can be transformed in HRV features and KSS annotations can justify a binary classification of alert/drowsy. Each individual drove twice, first in a day period, after a normally slept night, without influence of alcohol or caffeine, and the second in the night, after spending the day awake in normal activity. This measurement design intended to force alert and drowsy data from all individuals, even though it is not guaranteed that each individual record does not have a wide range of KSS scores associated with both alert and drowsy states.

Four models were tested, Support Vector Machines (SVM), one class SVM (ocSVM), Gradient Boosting Trees (GBT), and Artificial Neural Network.

General and Individual models were tested, and, having set that individual models were the only viable option for this classification task, the SVM model showed better performance than the other three, being that one used to further select features and test class balancing strategies. Model performances were measured using Matthews Correlation Coefficient (MCC).

3.4. Implementation of the system

To test the entire system, a simulator experiment was designed, making 13 volunteers driving in two sessions of 30 minutes, while collecting data with the three devices and stating every five minutes their drowsiness level with the Karolinska Sleepiness Scale.

From each session signals', HRV features were calculated in windows of two minutes with 50% overlap, and the KSS annotations were interpolated to match this time windowing. The features from each device to train and test a model for each one of them, and the chest band based model was also used to classify the data from the other devices. To have a direct measure of how well can these devices be used to detect drowsiness, but also to understand if a single model can be used for data from any source.

4. Results & discussion

This thesis studied the feasibility and technical requirements of producing a peripheral cardiac signal based drowsiness detection system. Three main dimensions of this problem were approached and answered: how to collect these peripheral signals and convert them into streams of IBI values, how to use those values in drowsiness detection and whether such system could be agnostic to the original signal measured.

The first question was answered by introducing three different types of wearable devices capable of collecting cardiac rhythm information: the chest strap Movesense, the capacitive steering wheel CardioWheel, and the wrist PPG sensor PulseOn. While the ECG based devices provided built-in filtering that allowed direct detection of R peaks and subsequently IBI, the PPG sensor suffered from sensitivity to external conditions, such as perceived ambient illumination from both light and hand position changes. This created sudden offset changes in the signal that needed a special filter to eliminate, that could not depend on frequency filtering due to the step nature of those artifacts. Instead, an online filter that mimics recursive moving average removal was created. By applying such filter to mimic a window of one second, resulting signals would maintain only the oscillatory component, where cardiac rhythm is encoded. Additionally, an adaptive threshold peak detection algorithm was implemented to locate the peaks of PPG signal. the algorithm used also a refractory period of 0.6 times the length of the last detected interbeat interval to avoid false peak detection, and reset its threshold after 1.5 times the last detected interval passes without a new peak detection. This created a detection system robust against changes in pulse amplitude, false peaks created by sudden movement and interruptions in the signal pulses. For the other devices, while CardioWheel directly provided the IBI itself calculated, Pan-Tompkins algorithm was used to detect R peaks in Movesense signals, correcting the peak locations by selecting the maximum value in a 0.4 seconds window centered in the initial estimates.

While the peak detection methods used in the different signals proved capable of identifying the peaks present in them, moments of poorer contact between the individual and the devices lead to missing peaks and added artifacts that corrupted some of the intervals collected. Even though normal procedures to treat such outliers consist of simply eliminating them, the ultra short nature of the analysis time windows used, 2 minutes, required a more conservative approach. Hence a IBI corrector system was created, evaluating the ratio between consecutive IBI to detect both missing peaks and false detections to accordingly estimate the location of the non-detected peak and divide the longer IBI or two join two shorter intervals into the true IBI. By testing this system on artificially corrupted segments of visually validated ECG, it was shown that the system is capable of reconstructing the sequence of IBI from signals corrupted with 10% missed detections and additional 10% false peaks with less than 7.5 milliseconds of mean absolute deviation from the true signal (table 1). The system was tested to the limit of having 40% of the signal values corrupted, and still managed to retrieve a stream of values with an MAD of 38.03ms, which, while very unlikely that such a large portion of the signal produces faulty IBI values, its still a smaller temporal deviation than the uncertainty in IBI determination on a 25Hz signal as the PPG is. This system is relevant to ensure that all collected information is used to calculate the HRV has confidently as possible, but the author leaves also the suggestion of its usage on analysis of longer term HRV, as it maintains the true succession pairs used in non-linear analysis as the Poincaré plot and the traditional sample elimination does not.

The second question, how to use the IBI values to detect drowsiness, was answered by searching the best subset of HRV features and the best model architecture to do so. An initial set of time and frequency domain features was used to compare four decision models, SVM, ocSVM, GBT and a ANN. In the process of testing which model performed best, the author realized that a general models, this is, a model trained to classify drowsiness in any arbitrary individual was performing poorly, independently of model architecture. This lead to the investigation of personalized models, which showed great improvements for part of the population. The individuals that continued to perform badly showed upon further analysis of their data that the limitations of the experiment and sleepiness scale used for the database, SleepEye, brought:

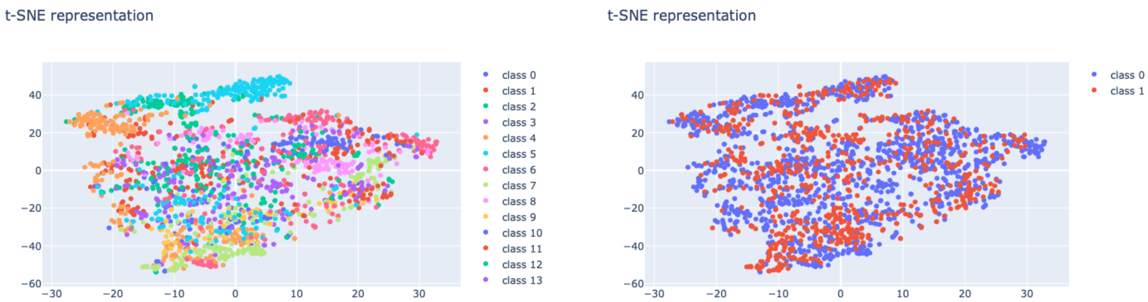
- Unbalance in classes, while the experiment was designed to have both alert and sleep deprived driving sessions, not all participants managed to provide enough KSS ratings associated with being sleepy for the model to properly learn the separation boundary between the two classes, even with class balancing methods applied.
- Imprecise self rating of their own state, being KSS a subjective drowsiness scale, the confidence in the annotations is proportional to the capability each individual has to self assess its state and correctly understand the levels of the scale. By looking into some of the ratings provided by subjects in this dataset, consecutive values with high ranges of variation raised the suspicion that some individuals were not accurately reporting their KSS level.
- The fundamentally continuous nature of drowsiness, as it is not a biological switch, where people would be either fully alert or fully drowsy, it is a process that sets in continuously, which makes the definition of a dangerous drowsiness level a rather arbitrary process, blurring the class separation in this problem. It was observed that the best results were obtained by individuals that reported both very low KSS values ($<$4) and high ones ($>$7), while those that concentrated ratings in values between 5 and 7 had the poorest performances.

By evaluating only the population whose annotations showed a good understanding of the scale, trustworthy self report and balanced experience of both alert and drowsy states, the models trained and tested for each of the 12 selected individuals attained a mean performance across them of 0.64 ± 0.04 and 0.49 ± 0.05 MCC for SVM and GBT respectively, while the other two models continued to perform poorly, thus being discarded. At last, to confirm that the poorly performing individuals were not the cause why the general model failed, a new general model was trained and tested with data from the 12 selected ones. This new model failed, and it was shown, using t-Distributed Stochastic Neighbor Embedding (t-SNE) representations of feature space, that the real reason for that was that each individual forms its own cluster in feature space, and while a frontier can be defined between the alert and drowsy data of one individual, the displacement of the various subjective clusters makes it impossible to determine a single common boundary (figure 3).

From here, the SVM model was selected as the best fitted to classify personalized state of drowsiness. Features used in the classification were revised, eliminating VLF because it did not hold significance when calculated in a short time window as 2 minutes, and two non-linear features were

Table 1: MAD of IBI reconstruction with different levels of contamination.

Missed detection density	False peak density 0.00	0.02	0.04	0.06	0.08	0.10
0.00	0.13	0.20	0.30	0.58	1.56	1.99
0.05	2.66	2.14	2.43	2.79	3.75	3.98
0.10	5.23	5.63	5.97	6.62	6.20	7.42
0.15	10.40	10.08	11.56	11.84	12.70	12.39
0.20	15.96	17.27	17.64	17.19	18.35	19.17
0.25	24.39	25.19	24.00	27.20	26.62	25.95
0.30	35.93	36.19	36.84	35.79	37.37	38.03



(a) **(b)**

Figure 3: t-SNE representation of the dataset formed by all the well separated individuals data. While both plots distribute the same data, (a) colors each point according to the subject the point comes from, and (b) colors the points according to the class (0=alert, 1=drowsy) they belong to.

added: first α component of Detrended Fluctuation Analysis (DFA) and Pointcaré SD_2. Unsupervised feature selection using MAD as the relevance metric was applied, which resulted in the elimination of LF feature. And Finally, hyper-parameters were fine tuned, defining an SVM with linear kernel and C parameter 0.3 as the best architecture for drowsiness detection, which attained a mean performance of 0.62 ± 0.03 MCC, which indicates a strong correlation.

This answers the question on whether IBI values can ultimately be used to detect drowsiness, but all these models were tested and trained with data collected through a chest ECG, and a final question must be analysed: can the same model detect drowsiness, but from IBI measured from a peripheral signal?

The experiment conducted in this thesis aimed to answer that. By applying the tools developed in the rest of the work, the simultaneously collected signals (chest ECG, hands ECG and wrist PPG) were converted into IBIs, and ultimately, HRV features were calculated for every two minute window in each of the signals. Unfortunately, from the 13 volunteers recruited, only two managed to survive the selection criteria applied to the SleepEye dataset. Three individuals had only one session, two individuals had missing data in one of their sessions, and four had reported the same state (either alert or drowsy) throughout the two sessions, not providing the two classes needed for training and testing the models. Additionally, two individuals had very unbalanced class distributions, and suffered from the same limitations found in the SleepEye data annotations, KSS values all bellow 8, and very few minutes reporting a 7 drowsy state.

The two individuals that survived the selection criteria produced models with very good MCC scores when trained and tested with data from the same device, with all devices. Those scores ranged from 0.62 to 0.81. And, to answer the final question, the model trained with data from the Movesense device, remained well performing when applied to data from the peripheral cardiac signals, ranging scores from 0.34 to 0.61, and with the high note of the performance of classifying PulseOn data with the model trained with its own data or the Movesense one is the same (table 2). Additionally, McNemar's test was used to compare the classifications of the entire dataset of each device with the Movesense trained model, for each individual, and it showed that all crossed classifications were statistically equivalent to the base classification, Movesense train on Movesense data (table 3).

This results indicates that the system this thesis proposes is very possibly feasible, and well performing.

5. Conclusions

The outcomes of this thesis confidently support the feasibility of the signal processing tools designed for PPG signals, in terms of filtering and peak detection. Bringing this signal closer to be equally considered as a solution for cardiac rhythm monitoring in the driving context. Moreover, the algorithm responsible for recovering the true sequence of IBIs showed to be robust against elevated levels of signal corruption, proving itself to be a valuable instrument to filter outliers but simultaneously maintain the original relations of adjacency, and use the most of the collected information for a more trustworthy calculus of HRV features.

The finding that personalized models outperformed a general one was already an idea proposed by previous research, but the way such fact could be visualized through t-SNE plots solidified it, maybe redirecting the development of HRV based models to take this as a starting point. One other interesting point to develop in the future is the fact that while individualized models proved to be the possible way to detect drowsiness, training each model for every new user of this system is not doable in a market perspective. However, it is hypothesised that a limited set of individual models can be representative of the possible ranges of HRV for a general population. By finding such set and combine them in a voting scheme or other ensemble classification framework a general and ready to apply drowsy detection system based on HRV can be created.

This work proposed also a model architecture that had the better and most consistent performance across 12 different individuals, as well as a selected set of HRV features specified for drowsiness detection in two minute time windows. The feature selection combined critical evaluation according to their validity in a ultra-short analysis framework, and machine learning feature selection strategies. The definition of a model template as this can serve as the basis for more complex ensemble models for general classification, as stated previously, or as a model waiting only for a training batch of data to be deployed in a personalized strategy.

Finally, the experiment results indicated that a peripheral cardiac signal based system for drowsiness detection is feasible, and also that the same installed system is agnostic to the signal source, being possible the use of different wearables measuring the cardiac rhythm through either ECG or PPG.

It is expected that such findings accelerate the incorporation of such a system in vehicles, helping reduce the burden of road casualties caused by drowsy drivers.

However, future work has to be developed to affirm this with certainty, firstly to compensate the limited size of the analysed population. A new study has to be conducted to evaluate if the findings of this work hold. The reduced number of individuals here analysed is pointed as the main limitation of this work, however the recruitment of volunteers during this pandemic time was mostly nonexistent, and the recruits consisted of the CardioID team and two professors involved in the development of the Simulator. Unfortunately the time window that aligned the readiness of the simulator with the ease in lockdown measures, as well as the calendars of each of the participants, didn't allow the retake of some of the sessions to be able to add more individuals to the final analysis.

References

[1] "Road traffic injuries," https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries, accessed: 2020-09-24.

[2] N. H. T. S. Administration, "2016 fatal motor vehicle crashes: Overview," NHTSA's National Center for Statistics and Analysis, 1200 New Jersey Avenue, SE Washington, DC United States 20590, Tech. Rep., 10 2017.

[3] P. Jackson, C. Hilditch, A. Holmes, N. Reed, N. Merat, and L. Smith, *Fatigue and road safety: a critical analysis of recent evidence*. Department for Transport, 01 2011.

[4] G. Maycock, *Driver sleepiness as a factor in car and HGV accidents*. Transport Research Laboratory, January 1995.

[5] D. Flatley, L. Reyner, and J. Horne, "Sleep-related crashes on sections of different road types in the uk (1995– 2001)," in *Road Safety Research Report No. 52*. Department for Transport, 2004.

[6] R. Robertson, E. Holmes, and W. Van Laar, *The Facts about Fatigued Driving in Ontario, A Guidebook for Police*. Traffic Injury Research Foundation, 2009.

[7] European Road Safety Observatory, *Fatigue*. European comission, 2018.

[8] National Sleep Foundation, *"Sleep America" Pool 2002*. National Sleep Foundation, 2002.

Table 2: Results for signal source performance comparison, for each subject, columns correspond to the origin of test data, and rows to training data.

Subject 152			
	Movesense	CardioWheel	PulseOn
Movesense	0.81±0.10	0.46±0.26	0.61±0.20
CardioWheel	–	0.77±0.11	–
PulseOn	–	–	0.61±0.19
Subject 128			
	Movesense	CardioWheel	PulseOn
Movesense	0.70±0.13	0.34±0.20	0.54±0.21
CardioWheel	–	0.62±0.13	–
PulseOn	–	–	0.70±0.14

Table 3: McNemar's test results regarding similarity of classifications from different sources.

Subject	Pair	χ^2	p-value	result
152	Movesense+CardioWheel	1.00	0.508	Equivalent
	Movesense+PulseOn	3.57	0.125	Equivalent
128	Movesense+CardioWheel	0.50	0.727	Equivalent
	Movesense+PulseOn	1.29	0.453	Equivalent

[9] J. Connor, G. Whitlock, R. Norton, and R. Jackson, "The role of driver sleepiness in car crashes: a systematic review of epidemiological studies," *Accident; analysis and prevention*, vol. 33, no. 1, pp. 31–41, 01 2001. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/11189120

[10] T. a. Dingus, S. G. Klauer, V. L. Neale, A. Petersen, S. E. Lee, J. Sudweeks, M. a. Perez, J. Hankey, D. Ramsey, S. Gupta, C. Bucher, Z. R. Doerzaph, J. Jermeland, and R. Knipling, "The 100-car naturalistic driving study phase ii – results of the 100-car field experiment," *Dot Hs 810 593*, 2006.

[11] J. Herman, B. Kafoa, I. Wainiqolo, E. Robinson, E. McCaig, J. Connor, R. Jackson, and S. Ameratunga, "Driver sleepiness and risk of motor vehicle crash injuries: a population-based case control study in fiji (trip 12)," *Injury*, vol. 45, no. 3, pp. 586–591, 03 2014. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/23830198

[12] D. Hennessy and D. Wiesenthal, "The relationship between traffic congestion, driver stress and direct versus indirect coping behaviours," *Ergonomics*, vol. 40, pp. 348–361, 03 1997.

[13] G. Matthews, L. Dorn, T. Hoyes, D. Davies, I. Glendon, and R. Taylor, "Driver stress and performance on a driving simulator," *Human factors*, vol. 40, pp. 136–49, 04 1998.

[14] K. Różanowski, O. Truszczyński, K. Filipczak, and M. Madeyski, *The level of driver personality and stress experienced as factors influencing behavior on the road*. WIT Press, 2015.

[15] T. Litman, *Autonomous vehicle implementation predictions*. Victoria Transport Policy Institute Victoria, Canada, 2017.

[16] U. Rajendra Acharya, K. Paul Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: a review," *Medical and Biological Engineering and Computing*, vol. 44, no. 12, pp. 1031–1051, 2006. [Online]. Available: https://doi.org/10.1007/s11517-006-0119-0

[17] R. Gordan, J. K. Gwathmey, and L. H. Xie, "Autonomic and endocrine control of cardiovascular function," *World J Cardiol*, vol. 7, no. 4, pp. 204–214, Apr 2015.

[18] J. Sztajzel, "Heart rate variability: a noninvasive electrocardiographic method to measure the autonomic nervous system," *Swiss Med Wkly*, vol. 134, no. 35-36, pp. 514–522, Sep 2004.

[19] J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, 1985.

[20] C. Fors, C. Ahlstrom, P. Sorner, J. Kovaceva, E. Hasselberg, M. Krantz, J. F. Gronvall, K. Kircher, and A. Anund, "Camera-based sleepiness detection: final report of the project sleepeye." *Vip Publication*, 2011.

[21] R. E. Kleiger, J. Miller, J. Bigger, and A. J. Moss, "Decreased heart rate variability and its association with increased mortality after acute myocardial infarction," *The American Journal of Cardiology*, vol. 59, no. 4, pp. 256 – 262, 1987. [Online]. Available: http://www.sciencedirect.com/science/article/pii/00029149879