Monitoring Driver's Respiration Using Smart Seatbelt

Implications for Driver Sleepiness

Master's thesis in Biomedical Engineering

Jason Meek
Banitama Supartha

DEPARTMENT OF ELECTRICAL ENGINEERING

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2021
www.chalmers.se

Master's thesis 2021

Monitoring Driver's Respiration Using Smart Seatbelt

Implications for Driver Sleepiness

Jason Meek
Banitama Supartha

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering
Division of Biomedical Engineering
Biomedical Signals and Systems
Chalmers University of Technology
Gothenburg, Sweden 2021

Monitoring Driver's Respiration Using Smart Seatbelt
Implications for Driver Sleepiness
Jason Meek, Banitama Supartha

Supervisors: Ke Lu, Electrical Engineering & Johan Karlsson, Autoliv
Examiner: Bengt-Arne Sjöqvist, Electrical Engineering

Master's Thesis 2021
Department of Electrical Engineering
Division of Biomedical Engineering
Biomedical Signals and Systems
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2021

Monitoring Driver's Respiration Using Smart Seatbelt
Implications for Driver Sleepiness
Jason Meek, Banitama Supartha
Department of Electrical Engineering
Chalmers University of Technology

Abstract

Driver inattention and sleepiness is one of the major factors leading to unsafe driving situations, accidents and in the worst cases, traffic fatalities. Although Sweden currently has the lowest traffic fatality rate in Europe, there is always reason to strive to lower this number, both in Sweden and the world. Autoliv Inc., a leading supplier of vehicle safety systems, is conducting research and development in the improvement of in vehicle driver monitoring systems. One such idea is to use a specially designed seatbelt buckle sensor to analyse the breathing characteristics of a vehicle occupant. The main questions of this study are. Is it possible to accurately measure occupant respiration using this sensor and if so, can useful features be extracted from the respiration signal to help predict driver intention or sleepiness with the aid of classification algorithms? The study begins by looking at previous research done into driver sleepiness classification using respiration derived features. This is followed by a description of the methods needed to take the raw respiration signal to a group of extracted features. Results from each stage in the methods are presented, after which the results are discussed, and overall conclusions made. The results show that it is indeed possible to extract accurate respiration signals using the belt tension sensor system. In total, 20 features have been extracted. Driver respiration rate, inspiration/expiration ratio, and respiration rate variability are to name a few. A number of features have been identified as having a clear relationship to sleepiness, while others show little observable relationship. Using feature selection, the most impactful features are fed into several classifiers, and the accuracy of those features in predicting the sleepiness of the subject is measured. The resulting accuracy ranges from 61% to 79%, depending on the type of the classifier and the features used.

Keywords: sleepiness, respiration, seatbelt, belt tension sensor, classifier, machine learning, signal processing, vehicle safety

Acknowledgements

We would like to thank Johan Karlsson from Autoliv and Ke Lu from Chalmers as our supervisors, whom we received continued support as well as ideas when feeling a little stuck and in need of inspiration. Autoliv as a whole deserves a thanks as they have given us access to world class data and insight into an incredibly interesting research field. We also would like to thank our family and friends for the support throughout our academic years.

Jason Meek, Banitama Supartha, Gothenburg, June 2021

Contents

List of Figures

List of Tables

Nomenclature

ANS Autonomic Nervous System
AUC Area Under ROC Curve
BB Breath-by-breath
BCG Ballistocardiography
BR Breathing Rate
BRV Breathing Rate Variability
BTS Belt Tension Sensor
ECG Electrocardiograph
ED Expiration Duration
EEG Electroencephalograph
EMG Electromyograph
EOG Electrooculograph
HR Heart Rate
HRV Heart Rate Variability
I/E Inspiration/Expiration
ID Inspiration Duration
KNN K-nearest Neighbor
KSS Karolinska Sleepiness Scale
OSA Obstructive Sleep Apnea
PA Peak Amplitude
ROC Receiver Operating Characteristic
RR Respiratory Rate
SD Standard Deviation
SST Synchrosqueezing Transform
SVM Support Vector Machine
$TEDD$ Thoracic Effort Derived Drowsiness
UE Unusual Respiration Event
VBB Breathing Rate Variability
$VBBref$ Reference Breathing Rate Variability

1

Introduction

European member states recorded around 22 800 traffic fatalities in 2019[1]. There are on average, 51 traffic related deaths per million of the population in each EU state. Sweden has the lowest, with 22 deaths per million and Romania being the highest, with 96 deaths per million. Meanwhile, in the United States, National Highway Traffic Safety Administration (NHTSA) published in 2019 that sleepy driving related accidents contribute as much as 697 (1.9%) of traffic fatalities[2]. That number did not include nonfatal traffic accidents with life-changing injuries.

The European Commission (EC) has several actions to push the traffic fatalities number lower. A major one is the updated General Safety Regulation document[3]. This document lays out regulations for type approval of vehicles, trailers, components and other systems regarding safety and protection of passengers and vulnerable road users. A new safety feature mandated by the updated regulation is that of improved driver drowsiness and attention warning systems. Car manufacturers should demonstrate that new vehicles that are going into the market are equipped with these systems. NHTSA has worked on several projects to combat drowsy-driving related traffic accidents. One of the projects is to work closely with manufacturers to design, test, and implement drowsy driving detection systems[4].

Autoliv Inc. is one of the world's largest suppliers for vehicle safety systems, such as airbags, seatbelts, steering wheels, and other systems linked to each of these. One component of interest that Autoliv produce, is a seat belt buckle with a built in tension sensor. This buckle was initially developed for the US market in order to fulfill the regulation that vehicles must be fitted with a child seat detection system and is named belt tension sensor (BTS). To adapt to the new regulations for vehicle passenger monitoring and safety, Autoliv is working on (among other things) developing the next generation of driver monitoring and inattention/drowsiness detection systems. The system is designed to take into account physiological measurements of the driver. For this particular project, the possibility to measure breathing and breathing-related parameters using the BTS to estimate driver sleepiness is explored.

1.1 Aim

The aim of this project is to develop a system that is able to use seatbelt tension and movement to detect driver sleepiness based on the recordings from the Autoliv BTS. In order to fulfill this aim, the following sub-aims will need to be considered.

1. Determine the possibilities of detecting sleepiness using respiration measurements based off of research done by previous works.
2. Implement data cleaning algorithms to remove noise and unwanted measurements from the BTS measurements, and make comparisons between the BTS and reference measurements to determine if the BTS is a vaiable tool for Autoliv to use for respiration measurement.
3. If the previous two steps are feasible, develop and implement a proof of concept drowsiness detection algorithm that takes parameters derived from the BTS data as its primary source of input.

1.2 Research Question

In this project, the main research question is to what degree can the BTS sensor developed by Autoliv be used as an aid to detect driver drowsiness. As with the aim, this question can be broken down into sub questions.

1. The first question that must be answered is if it is possible to use respiration data (chest movement) to differentiate between different drivers' sleepiness levels?
2. If this is possible, can the BTS system act as a viable method for vehicle occupant respiration measurement?
3. Can useful parameters with a proven link to sleepiness level be extracted from the BTS data?
4. Finally, is it possible to develop an algorithm based off of previous research that can be applied to the BTS data?

1.3 Scope and Limitations

Scope
The scope of this project is fairly broad and covers the analysis of existing methods used for respiration measurement, feature extraction and sleepiness detection. It will also cover the processing of the BTS and reference band data so as to enable comparison and extraction of features. Finally, it will include the analysis of features and the development of a classification system to try and infer driver sleepiness from useful features. The data collection is outside the scope as it was performed a year previous to the undertaking of this project, so is the development and design of the BTS used for respiration measurement. Usage of additional measurements for sensor fusion such as vehicle speed, acceleration, location and steering wheel angle are outside the scope.

Limitations

The project has been limited to using only the BTS and reference chest band as respiration measurements. As time is short, it was not possible to perform any testing of the system or development of a real time processing solution. The broad nature of the project and set time budget, limits the ability to delve into depth in each of the main sections, with a slightly more proof-of-concept feel to it. What is able to be disclosed in detail is also limited by company wide non disclosure agreements. This includes the intricacies of the BTS system as well as the full experimental setup in the vehicle. As no more data was collected after the project began, it has been limited to the existing collected data. Similarly, the analysis of features with relation to KSS scores is limited by the accuracy of the KSS system, which is a subjective one. Since the BTS and reference band measurements are of different units, it has been decided that results for these signals will be quoted as unitless.

2

Theory

This chapter will discuss background information on sleepiness, sleepiness assessment, and driving and sleepiness.

2.1 Physiology of Sleepiness

Sleepiness and sleep onset results in a host of measurable changes in the human body. An area of interest is in the physiological changes that may occur. A few such changes are briefly discussed in this section.

There are several documented research studies on physiological changes during sleep and sleep onset. In a sleep study (polysomnography), researchers monitor several physiological parameters when the subject is asleep. The recording includes several EEG channels, EOG (electrooculogram) channels, EMG channels connected to electrodes on several points on the body, ECG, and respiratory-related measurements (airflow, chest and abdomen movement, and O_2 saturation)[5].

The results from sleep studies are primarily measurements during stable sleep. To correctly make a conclusion on what happened during sleep onset, the measurement must focus on the period where the sleep onset begins. One of the early studies that investigated details on sleep onset shows that respiration rate (RR) changes when an individual is transitioning from a state of wakefulness to sleepiness [6]. A transition from wakefulness to sleepiness is defined as transition from alpha state to theta state in the EEG wave. The expressed change in respiratory tempo correlates with falling metabolism rate and increasing respiratory resistance.

To put it into more detail, the fluctuation between alpha state and theta state on sleep onset is thought to affect the regulatory system that affects control of respiration. What happens in the state transition is explained in the following chain reaction. The tidal volume during the transition falls significantly due to falling metabolic rate. This affects the estimated minute ventilation as well. The drop in estimated minute ventilation increases the partial pressure of carbon dioxide in the blood, which then increase the respiratory drive. The increase in the respiratory drive is expressed in increase in the inspiration period and ratio of inspiration period

to respiratory rate [6].

There also exists interesting observed changes in the autonomous nervous system (ANS) with an increase in sleepiness. In this case it must be mentioned that sleepiness and sleep onset are slightly different. The sleep onset mentioned previously is when a person is transitioning from a state of wakefulness to sleep, while sleepiness is occurring all within a state of wakefulness. It has been shown that the heart rate variability (HRV) has a significant relationship to sleepiness. Both [7] and [8] find that there is a change in activity of the ANS with an increased level of sleepiness. They use the alterations in HRV as a method to measure the ANS activity. Another researcher observed changes in ANS activity and its correlation with the increase of sleepiness level [9]. Increased sleepiness was shown to be significantly correlated with increased parasympathetic activity, measured via the response of pupil light reflex[9]. The idea here is that with the ANS activity changing as sleepiness increases, it is possible to observe changes in physiological signals linked to ANS processes.

Although it is accepted that there is some physiological change during sleep onset, the magnitude of the change might vary among individuals. For example, a study found that during sleep onset, people with obstructive sleep apnea (OSA) have less respiratory cycle fluctuations compared to people with no OSA [10]. The less responsive ANS was thought to be the cause behind the lack of irregularity in breathing during sleep onset [11].

2.2 Sleepiness Assessment

It is easier to assess subjective sleepiness rather than measuring the physiological condition of sleepy individuals in a driving environment. To facilitate that, a psychometric scale is needed to quantify drowsiness. Two such scales are the Karolinska Sleepiness Scale and Observer Rating of Drowsiness.

The Karolinska Sleepiness Scale [12], or KSS for short, is the most common way to quantify the drowsiness of the subject. The subject must self-report the score based on their sleepiness. The KSS is a 9-point scale, going from 1 up to 9.

Meanwhile, the Observer Rating of Drowsiness is a method of evaluating the sleepiness of a subject from the footage of their face and upper torso [13]. Two or more trained observers rate the subject's drowsiness from that footage, hence the name. The sleepiness rating uses a Likert scale-based assessment, starting with 0 for not sleepy, 25 for slightly sleepy, 50 for moderately sleepy, 75 for very sleepy, and ends up at 100 for extremely sleepy. These are but two examples of methods used to monitor participant sleepiness levels. In this study, the KSS system is the system of choice.

Table 2.1: Subject sleepiness and its corresponding KSS score

Score	Sleepiness
1	extremely alert
2	very alert
3	alert
4	rather alert
5	neither alert nor sleepy
6	some sign of sleepiness
7	sleepy, but no effort to keep awake
8	sleepy, some effort to keep awake
9	very sleepy, great effort to keep awake, fighting sleep

2.3 Driving and Sleepiness

Sleepiness is a common contributor to vehicle accident numbers, especially on a dull and monotonous roads [14]. This is reflected by a higher percentage of tiredness-related accidents on highway roads compared to rural roads[15]. One such study (using a driving simulator) confirms that when the feeling of sleepiness increases, more adverse safety events (centerlane crossing and crashes) will occur[16].

Although drivers usually know that they are sleepy, they often have no recollection about sleeping on the wheel while they did. The relatively long period of sleep onset compared to the actual period where the driver is asleep might be the cause of this [15]. The stronger the feeling of sleepiness is, the larger the possibility for potential accident. To make things more dangerous, the feeling of sleepiness could subside and drivers might overestimate their alertness level [14]. Therefore, it is important to warn the driver about their actual sleepiness and encourage the driver to take some rest.

One way to assess sleepiness in a driving environment is to monitor the behaviour of the driver. Monitored behaviours can include, among others, lane-keeping performance and steering wheel control performance[17]. Other visible signs for sleepiness are head nodding, frequent blinking, heavy eyelids, and yawning.

Another way to assess sleepiness is by monitoring physiological parameters. In a driving environment, it is preferable to employ measurement methods that do not interfere with driving activity. Clearly, devices such as EEG caps, EMG electrodes, and EOG electrodes are ruled out here. Less obtrusive sensors that can be used for physiological monitoring include steering wheels with electrodes to measure heart rate (HR), eye tracking equipment for gaze signal acquisition (pupil size, percentage of eye closure), seat pressure sensors, thermal imaging cameras and seatbelt tension for respiration signal measurement (which is the main topic in this report) [18].

2.3.1 Respiration and Sleepiness

A major interest in this project is linking respiration measurement and driver sleepiness. A normal adult breathes at a rate of 12-16 breathes per minute [19], with inspiration/expiration ratio being 1:2 on average [20]. As discussed in 2.1, sleepiness has the ability to affect measurable physiological signals. The following research has been done into observing the effects on respiration during increased sleepiness.

Brandy et al. [21] measure the chest movement (respirations) directly using a wearable chest band. Said chest band also records HR data. They then extract HRV, BR variance, and power spectral density of those data to determine fatigue through a neural network. Although they have not done the neural network classifier part yet, the study concludes that the HR and the HRV decrease if the subject becomes sleepy. They also found that RR drops when a subject falls asleep.

Igasaki et al. have published two papers relating to sleepiness estimation using respiration derived features [22][23]. In the first, they use the HRV and breathing rate variability (BRV) to infer sleepiness onset while driving, via logistic regression. The motivation for HRV is made simply that it is commonly used for similar tasks and shown to be effective. For BRV, they mention that since the ANS is able to affect the respiratory system, and that there is an observed activity in the ANS when sleepiness is higher, there should be a possibility to observe a changes in BRV as sleepiness increases. The observed values in presence and absence of sleepiness showed significant differences for all participants, but there did not seem to be a clear trend throughout. For the second study they decide to forgo BR and use other respiration derived parameters to infer sleepiness by training a SVM. The idea is to explore other respiration parameters that can be linked to sleepiness. Breathing parameters derived include mean and standard deviation (SD) of inspiration duration (ID), expiration duration (ED), peak amplitude (PA), driving, and timing (which will be explained later). They also measure the frequency and duration of unusual respiration events (UEs), which is a condition where the breath amplitude exceeds a certain threshold. The motive behind the selected features is that the researchers believed that the selected parameters are good indicators of thoracic respiration. They discovered that mean ED, SD ED, SD ID, SD PA, and UE duration are positively correlated with sleepiness, while mean timing are negatively correlated with drowsiness.

Rodriguez-Ibanez et al. [24] look into the relative variability of breath-by-breath (BB) interval. Their idea is that as a drive progresses and the driver becomes more sleepy, their BRV will increase compared to the start of the drive, allowing for sleepiness to be detected by the increase of relative BRV. They do this by finding a 40-second reference window in the first 5 minutes of the measurement where the BB time is the most stable. The magnitude of this variability is used to scale the BB variability in the rest of the recording. The resulting parameter is termed Thoracic Effort Derived Drowsiness (TEDD) index. The result shows that the TEDD index increases significantly as the level of sleepiness increases, which translates to positive correlation between breath interval relative variability and sleepiness. A

more recent paper from the same researcher looks into instantaneous respiratory frequency (IRF) [25]. The result shows that a sleepy driver has a lower mean IRF and higher standard deviation of IRF compared to an awake driver. Lee et al.[26] use 15 parameters extracted from 8-channel EEG and a respiration sensor. The study goes into little detail regarding respiration, but it claimed that the decrease in regularity of respiration is a useful marker for increasing fatigue level.

Kiashari et al. [27] investigated using a thermal imaging system to measure driver respiration and drowsiness with the knowledge that respiration can vary significantly with an increased level of sleepiness [6][28]. The RR and inspiration-expiration (I/E) time was extracted from the respiration data. Analysis of the RR and I/E showed that the mean RR decreased during periods of sleepiness and the mean I/E showed a tentative increase.

2.3.2 Respiration Measurements and Parameter Extraction

The methods used for gathering respiration data can vary significantly from study to study. This subsection will discuss the ways different studies obtained respiration data and the parameters that were extracted from this data. Although none of these studies use a belt tension setup like in this report, the idea will be to find comparable methods that can be applied to this problem. Plethysmography measure change of volume in an organ or whole body. Therefore, pulmonary plethysmography is measurement of change in lung volume [29]. Study in sleepiness commonly use pulmonary plethysmography to measure respiratory function. In this report, when referring to the respiration signal, it is in the form shown in Figure 2.1, a time varying representation of the change in tidal volume.

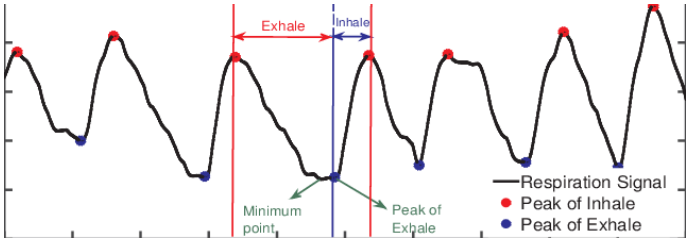


Figure 2.1: Annotation of a respiration signal in the form most commonly referred to in this report [30].

Principles of Measurement

Physiological measurement in a vehicle setting needs to be as unobtrusive as possible. A direct measure of respiration, for example, spirometry, is highly accurate but is obtrusive and therefore not well suited for people actively driving a vehicle. Another, obtrusive method is to use probes placed at the base of the nostrils that measure the temparature change of the air flowing past [31]. The inhaled and exhaled air will be of different temperatures, which is measured by the probe. The system converts the change in temperature to an electrical signal that represents the respiration. Again, such a system wont be well accepted by car drivers. For the measurement of

respiration unobtrusively, there exist two main physical effects that can be measured [18]. These are, mechanical and thermal effects.

A popular way to measure respiration mechanically is by the expansion and contraction of the thorax or abdomen in accordance with the change in lung volume and diaphragm movement. One such method is Ballistocardiography (BCG). BCG is the remote measurement of cardiac activity, commonly performed by measuring the displacement of the body caused by heart and lung movements with pressure sensors or strain gauges [18][32]. Apart from using pressure and strain gauge sensors embedded in seats or seatbelt webbing, Respiratory Inductance Plethysmography (RIP) bands are a popular measurement tool. This device consists of a zigzagging conductor embedded in an elastic chest band, that when changed in length (by chest expansion/contraction), alters the magnetic field generated by current flowing in the conductor. When placed around the thorax or abdomen, respiration can be measured by converting the change in magnetic field to a signal whose amplitude follows the change [31][33]. Optical methods can fall within the mechanical sensing realm too. BCG respiration measurement can be performed indirectly by an imaging system in which the change in distance between a camera lens and participants chest is measured [34][31][18]. This change in distance can be transformed into a displacement-based respiration signal. Thermal effects of respiration come in the form of changes in body surface temperature. Warming and cooling of the region around the nostrils as a person is breathing in and out may be measured by thermal imaging cameras, after which respiration can be derived based on a conversion of temperature change to flow [27][18]. In this project, measurement of respiration by the BTS falls within the mechanical realm. It can be seen as a form of BCG, with the change in belt tension measured at the buckle due to chest displacement, being related to respiration. The observed signals will be very similar to that shown in Figure 2.1.

Past Studies Respiration Data Collection
A wearable chest band with the BioHarness 3.0 sensor attached, is utilised in [21]. The band is used to measure the HR and BR, with subsequent data streamed to a mobile device for analysis. In studies, [24], [22], [23], [26] [25] and [35], a RIP band or similar such device is used to measure respiratory signals. In each case, after the respiration signal is obtained, various features can be extracted from this data. The placement of the chest band appears to have an effect on the measured signal. Subsequently, [25] and [35] experimented with placing three bands on each subject at the thoracic, diaphragm and abdominal positions. It was concluded that the optimal location to place such a band is the abdominal position [25].

As mentioned, there exist other forms that can be used to measure respiration. One of these is via imaging. In [27], a thermal imaging system is used to measure respiration according to the heat around the targets nostrils. An inspiration causes the area to cool while an expiration the opposite. Their motivation is that this sort of driver assessment will be truly non-invasive. Another imaging-based solution is employed by both [34] and [31], which use the Kinect camera system developed

by Microsoft to detect depth changes in the thoracic and abdominal regions to measure respiration. Again with the motivation to enable a less invasive monitoring method. Although not usually intended as a respiratory measurement tool, ECG signals can be processed to provide accurate respiratory measurement [36] [37]. The method proposed by [36] uses a combination of wavelet decomposition, filtering and averaging to obtain a respiratory signal, while [37] uses a more Fourier Transform based approach. One downside of many of these approaches is that they can become costly and bulky. [38] attempts to remedy this by using a small accelerometer sensor attached to the subjects chest. This is used to measure the movement of the rib cage, which is then translated into a respiratory signal. A compiled table showing studies that have been conducted and what respiration parameters were extracted can be seen in Appendix A.1.

Data Pre-processing

It is worth noting that prior to extracting features, signal preprocessing can be a useful step depending on what features are to be extracted. Downsampling, bandpass filtering, or a combination of the two are commonly used. A two-step approach is used by [25], where the respiration signal is first downsampled and the passed through a bandpass filter. The downsampling takes the signal from 100Hz to 20Hz, while the filter is comprised of a Butterworth bandpass of second order, with a passband of 0.05-0.5Hz. The rationale behind this range is that general respiration rates will fall well within this spectrum. On the other hand, [35] chooses to forgo the downsampling and instead apply a fourth order Butterworth bandpass filter with a passband of 0.05 to 0.5Hz, after which a nonlinear enhancement function is applied. This function effectively normalizes the filtered signal to help reduce the effect of large artifacts. While looking at a different end application (respiration and stress), [34] filter the respiration signal using a bandpass filter of 0.1 - 1Hz, under the assumption that a person will usually have a RR between 0.1 and 0.35 Hz. Additional smoothing of the respiration signal was performed on 0.5 second segments post filtering. Yoon et al.[39] take a slightly different filtering approach to the common Butterworth method. A second order Savitzky–Golay smoothing filter is used to get rid of movement artifacts in the respiration waveform, enabling smoothing while keeping the general shape of the waveform.

Useful Parameters

After processing the signal, the breathing parameters are ready to be extracted from the respiration data. Respiratory rate (RR) is the most commonly extracted parameter from breathing data[21][27][22][25], while some other authors derived parameters related to inspiratory/expiratory period and magnitude[23][27]. Rodriguez-Ibanez et al.[24] use RR variability and use that parameter to derive a parameter called Thoracic Effort Derived Drowsiness Index (TEDD). Kiashari et al.[27] investigated the mean and SD of RR and I/E ratio with a rolling window of 2 minutes. Igasaki et al. [22][23] use inspiration and expiration duration and magnitude, as well as two other derived parameters called **driving** and **timing**. The timing parameter is the ratio between inspiratory period and the full respiratory period. The driving parameter is the change of amplitude during the inspiratory period. The same authors

extract UEs, for which they find a correlation between occurrence and duration to sleepiness increase. In this project, an interest will be placed on extracting breathing rate, I/E based features such as the I/E ratio, driving, timing and breath period as well as the UEs and variability of breathing rate in the form of TEDD. Theis decision is based on the fact that past studies have noticed interesting changes in these features as the level of sleepiness varies.

2.3.3 Drowsiness Classification

A drowsiness classifier works by taking breathing parameters and determining the subject's drowsiness from those input. Initially, several classifiers need to be trained from a dataset that contains the parameters and the target class. After training, the classifiers' performances are measured against a different dataset. The performance of the classifier might vary depending on the type of the classification scheme.

Rodriguez-Ibanez et al.[25] use simple thresholding from the TEDD index to determine the subject's drowsiness. A score of 0-3 is classified as being alert, 3-6 as fatigued, and above 6 as sleepy. This TEDD classification scheme is taken further by [35], in which a host of extra processing and signal quality checks are performed before making a decision from the TEDD value. This is claimed to improve the accuracy and specificity of the classification, which is still a sleepy or not-sleepy decision based on hitting a certain threshold.

Machine learning-based methods provide a way to classify drowsiness from several parameters. There are a large number of different algorithms that can be used depending on the type of parameters and what sort of classification is to be made. Commonly used algorithms are support vector machine (SVM), random forest, k-nearest neighbours (KNN), logistical regression, and various neural network setups.

Support vector machine (SVM) based models work well when there is a binary (or more than two class) classification to be made. SVM works by mapping input parameters into a higher dimension, after which a hyperplane is created that separates and categorizes the mapped inputs [40]. The hyperplane that separates the class are selected such that it provides optimal margin between two classes. An illustration of this classifier is shown in Figure 2.2.

K-nearest neighbors (KNN) works by classifying a new data point to the majority class of nearest data points with known class around the new data point. To tune this classifier, a learning algorithm search for the number K that returns the best performance. To better illustrate the principle of KNN, refer to Figure 2.3.

Random forest, or random decision forests, are classifier that consists of an ensemble of decision trees. Each decision tree has randomly selected set features from all available feature. The classification of an input is the total tally of all decision trees in the forest. Random forest is useful in overcoming the problem of poor generalization[43]. Figure 2.4 illustrate how random forest classifier works.

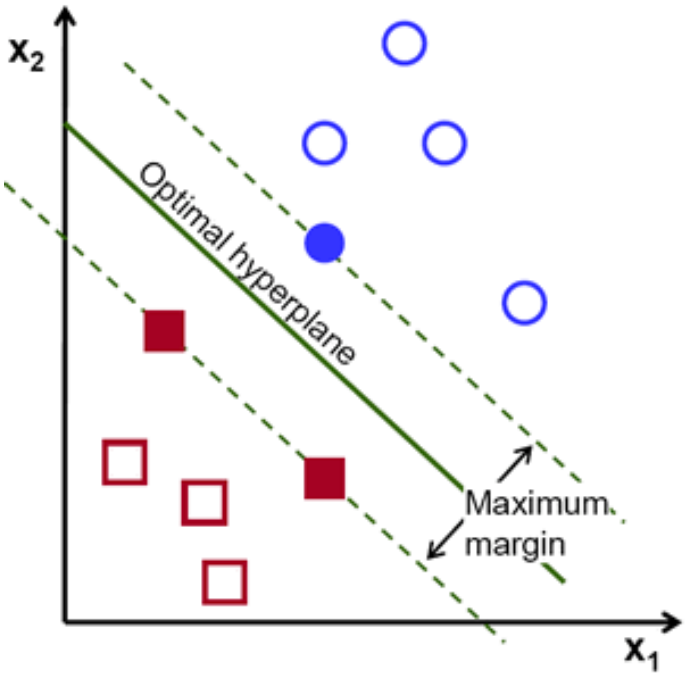


Figure 2.2: Overview of SVM classifier. The inputs with solid colors are the *support vector*. Image reference: [41]

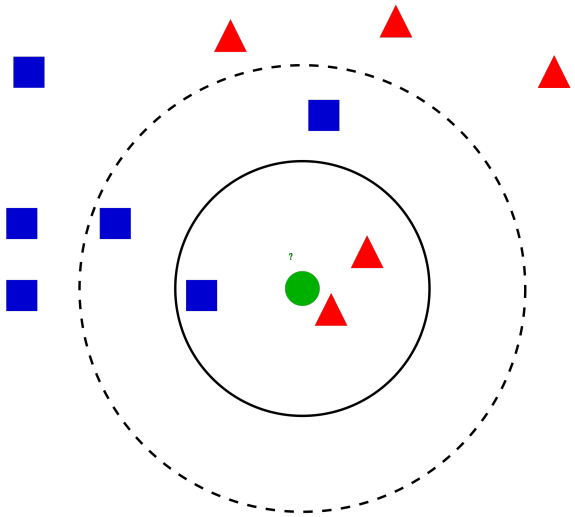


Figure 2.3: Overview of KNN classifier. The new data point (green circle) are going to be classified. The class the new data point falls into depend on the majority of the K-nearest neighbor. For K=3, it will be classified into red class. For K=5, it will be classified into blue class. Image reference: [42]

In the first paper by Igasaki et al., logistic regression was used to distinguish between the presence and absence of sleepiness using HRV and BRV based parametersm calculated over 5 minute intervals [22]. KSS scores equal or greater than 7 are considered sleepy, while below 7 is considered not-sleepy. The final results showed a mean accuracy of 74% for BRV or HRV parameters used on their own, and 84% when all parameters were used. In their second paper, a combination of mean and

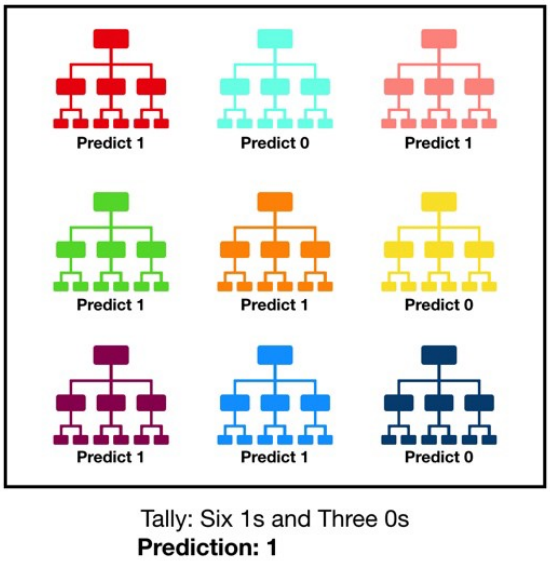


Figure 2.4: Overview of random forest classifier. This forest has nine decision trees with different class output. The classification outcome is determined with the majority class output (in this case, 1). Image reference: [42]

standard deviation of respiration based parameters calculated over 5 minute intervals are used to train a SVM [23]. The target values are a KSS score of 7 or above for sleepy, below 5 for alert and between 5 and 7 for neither sleepy, nor alert. The result was the classification learner accurately distinguish 89.4% of the sample into correct class from just the breathing parameters. Kiashari et al.[27] have used the mean and standard deviation of RR and I/E ratio for two minute segments to train a SVM as well as KNN classifier with a fusion of the derived parameters. Models were trained using leave one out, with the SVM performing best. A claimed accuracy of 90%, sensitivity of 92%, specificity of 85% and precision of 91%. Lee et al. use EEG and respiration features to train a SVM that classifies the driver state as one of 6 levels [26]. Different numbers of features were experimented with for training, with the best accuracy result from using 12 out of 16 features. Although not for drowsiness classification, [44] use a particle swarm optimization (PSO) based learning technique for classifying driver inattention. This approach uses random forest classifiers and neural networks to act on images and 7-point pressure measurements, which are then combined using PSO to achieve the best classification. A table sumarizing the different classification approaches to sleepiness detection, as well as what respiration based features are used is attached in appendix A.1.

3

Methods

The following chapter describes the methods and processes that were taken to achieve the project aims. Figure 3.1 shows a high level overview of the developed system. Each of the subsequent sections describe an area of this flow diagram.

3.1 Data Collection

The data collection was done as a part of a research study by Autoliv Inc., VTI, Smarteye AB and Linköping University, funded with a grant as part of the EU project, MEDIATOR. The experiment was approved by the Swedish government and the Swedish Ethical Review Authority. The data was collected from January 2020 to March 2020 [45]. The experiment consisted of 36 women and 53 men for a total of 89 participants. They drove on a dual-lane motorway, with varying degrees of automation and day condition in each drive. Of the 89 participants, 43 were randomly assigned to drive a Volvo XC90 with the remainder assigned to a Volvo V60. Each drive consisted of driving in one direction for 90km, turning around and returning the same 90km. The researchers used a 2x2 combination of automation and condition: with level 2 assistance turned on/off, and with day/night driving condition. This resulted in each participant undertaking 4 separate drives, two during the day and two at night.

The vehicle of interest for this project is the XC90 since it is fitted with the BTS. Apart from the BTS, installed instrumentation includes dual control equipment, a Raspberry Pi-based OBD logger, and a GPS recorder. In addition, the researchers gathered recordings from three cameras pointed to the forward roadway, driver's face, and driver's torso. A test leader, always ready to take control of the car when signs of dangerous driving appeared, accompanied the participant in the driving session.

The mean age for the participants was 38 years. Data regarding the weight and stature of each participant is also available. For three days prior to driving session, the participants were instructed to sleep at least 7 hours, go to bed no later than 00.00 and wake up no later than 09.00.

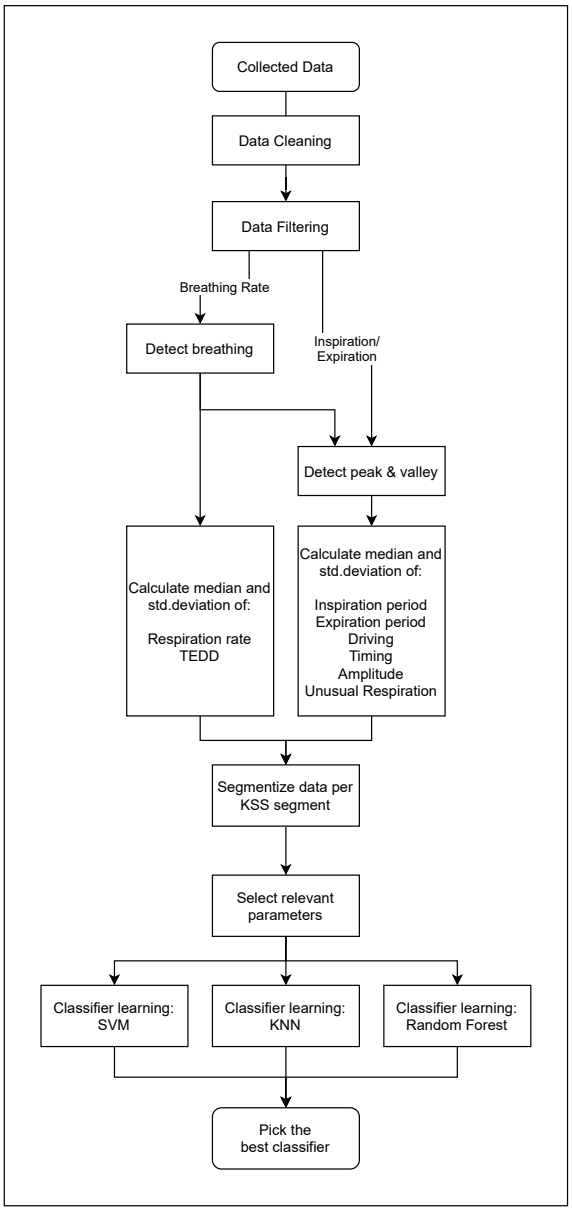


Figure 3.1: Overview of the data processing for classifier learning

The participant was fitted with several physiological sensors, such as 64-channel EEG, lead-II ECG, a single chest band for respiration measurement, and vertical electrooculograph. The chest band used in the data collection is connected to NeXus-10 biofeedback measurement system. In the driving session, the participants self-reported their sleepiness every five minutes using the Karolinska Sleepiness Scale. The physiological data of interest then are collected from two source. The first source is the chest band with a sampling rate of 256Hz, and the second is the BTS reading with a sampling rate of 200Hz. As mentioned in the introduction, the measurement unit of the BTS and the chest band is not directly comparable. Therefore, the obtained result will be unitless. When analyzing the video footage of participants, all work was done at the Autoliv office with strict rules not to save or distribute any of the videos. Additionally, all participants are anonymous. This is to ensure

GDPR and ethical regulations are followed.

3.2 Data Pre-processing

MATLAB R2020b on Windows 10 is used to process the data. The details of what happens in every step of data processing is described below.

3.2.1 Data Cleaning

The first step that must be taken in processing the data is to get rid of poor quality data. A sequential process for assessing the data quality was developed specifically for the data collected in this project and is further described below.

1. Set all timestamps to use the same timezone and convert the timestamps into unix time
2. Exclude any BTS or reference data that has large discontinuities in its timestamp vector
3. Clip remaining BTS and reference data to the same start and end time
4. Convert the BTS and reference data into MATLAB time series
5. Create a universal time vector and resample both the BTS and reference to this time vector, thus synchronizing reference and BTS
6. Filter and normalize the BTS and reference data
7. Perform correlation of the BTS and reference data
8. Investigate cases where there is a very low correlation and exclude recordings that have obvious problems

The first step deals with forcing a universal timezone. This is done because the timestamp vectors for the BTS and reference data are using differing time zones, which causes synchronising issues when comparing the two. Converting both to unix time allows for easier processing of the data. The second step is an initial check for poor quality data. This is done by calculating the difference between adjacent time values in the BTS timestamp vector and excluding any recordings where there is a discontinuity larger than 1 second as such data will cause problems later when synchronising segments of each recording with the KSS periods. The corresponding reference recording is removed in order to keep the number of reference and BTS recordings the same. Ensuring that the BTS and reference data starts and ends at the same point is an important factor since it will allow for a meaning full comparison when correlating the two sets. A *start drive* time was selected as the time corresponding to the first BTS timestamp, while the *end drive* was set as the last KSS recording time.

The next issue to overcome is that the reference and BTS data have different sampling frequencies, resulting in a different number of samples even when starting and ending at the same time. Steps five and six solve this issue by creating a universal time vector with a sample frequency of 8Hz. The BTS and reference time series data is resampled to this time vector, ensuring both have the same sample rate and use a

shared time vector. Ensuring the BTS and reference data for each recording/drive is the same length allows, among other things, for easier comparison between the two. A resample rate of 8Hz is chosen since it is an integer factor of both 200Hz and 256Hz, is sufficiently high to avoid any aliasing and is low enough to greatly reduce the amount of data that must be processed.

Before the data can be compared, it is passed through a bandpass filter and normalised. The applied filtering is the same as the first branch of filtering described in the *Data Filtering* part. Pearson's correlation coefficient is obtained for each recording pair of BTS and reference data, after which, recordings with a low correlation are investigated further. The correlation coefficient is calculated for each drive, for which there are originally 166. Due to the nature of the data, cases with low correlation are manually inspected before being discarded as there are cases where the two data match almost perfectly but have a consistent time offset. In cases where there is a clear difference in the BTS and reference data, it is hard to tell which is "better" and thus, for such cases, both recordings will be discarded.

3.2.2 Data Filtering

To get the desired breathing features, such as breathing rate, chest amplitude and inspiration/expiration timing, the data needs to be filtered. Since the breathing rate of normal people is between 12-25 bpm (0.2 - 0.45 Hz), it will be important to focus on this component while attenuating components above and below this frequency band. For this, the signal will go through two branches of processing. The first branch consists of a bandpass filter followed by nonlinear normalization, and is termed the *breathing rate* filter. This branch is intended to filter the data in a way to ease breath detection and breathing rate calculation. The nonlinear normalisation follows what was done by [35]. It effectively normalises the data while also reducing the amplitude difference between usual respiratory data and large spikes. The second branch, a highpass filter followed by a Savitzky-Golay filter [9], is geared to inspiration/expiration observations. This branch of filtering is intended to smooth the respiration signal, while retaining the original shape of the waveform, allowing for a more accurate calculation of inspiration and expiration. It is termed the *Inspiration-Expiration* filter or *IE* filter for short.

The filters used in the first branch are comprised of two fourth order Butterworth filters with cutoff frequencies of 0.15 and 0.5 Hz respectively. These values are slightly narrower than in previous works, and are chosen due to the increased amounts of high and low frequency noise introduced by the real world driving setup. The highpass filter in the second branch is a fourth order Butterworth filter with a cutoff of 0.05 Hz, while the Savitzky-Golay filter is of fourth order with a frame length of 2.5 times the sample frequency. This frame length was chosen via an iterative process and enables the retention of the original waveform shape while smoothing out larger high frequency jitter. The Matlab *filtfilt* function is used for filtering to avoid any phase shifting when filtering. The overall flow can be seen in Figure 3.2.

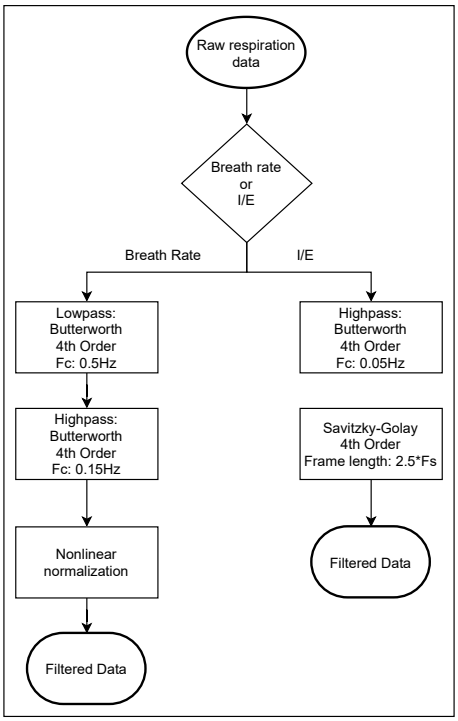
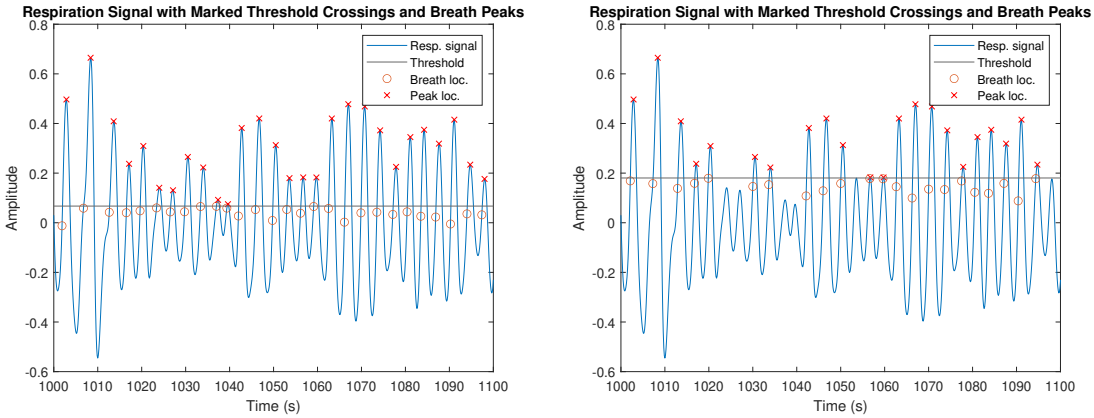


Figure 3.2: Data filtering setup

3.2.3 Breath Detection

In order to progress further with extracting RR and inspiration/expiration parameters, a viable breath detection algorithm must first be implemented. The described method takes inspiration from a threshold crossing application described in [24]. In essence, a window of minimal breathing variability is selected from the first n minutes of recording (which [24] set to 40 seconds). From this window, a threshold is set based on a percentile value of the breathing data in the window. This threshold is applied to the entire recording, where the location of a positive slope that breaches the threshold is considered a breath.

In practice, there are a number of parameters that need to be tuned. To begin with, a window length of 40 seconds was selected, with the first 10 minutes of the recording being used for optimal window location. A revised window length of 33 seconds was selected after iterating through window lengths ranging from 20 to 60 seconds. The 60th percentile of the breathing data in the identified window is used to calculate the breath threshold, down from the original design of 70th percentile used in the literature. This was done as over time it became evident that using the 70th percentile resulted in many viable breaths going undetected. This threshold is unique for each recording and is recalculated each time a new recording is processed. A representation of these two threshold levels is shown in Figure 3.3. When detecting the threshold crossing points, the algorithm will reject a breath occurrence if it falls within 2 seconds of the previous breath as we deem this too short a period for a valid breath to occur.



(a) **(b)**

Figure 3.3: Breath detections and the corresponding respiration peak using the breath detection algorithm with a threshold set to 60th percentile vs 70th.

3.2.4 Peak and Valley Detection

To correctly derive the inspiration and expiration phase from a breath, an accurate detection of a respiration peak and valley is needed. The algorithm for peak and valley detection works as follows:

Utilising the crossing index calculated in the breath detection stage, the occurrence of each breath can be found in the respiration data filtered by the second filtering branch. The manner in which a breath is detected means that after a threshold crossing, there should occur a reparation peak, followed by a valley. Thus, between two threshold crossings a peak will be seen followed by a valley. The algorithm observes the respiration data between two threshold crossing indexes and looks for the peak index using the *findpeaks()* function. The data is then inverted and the same process is applied to find the valley index. In the event that the *findpeaks* function finds more than one peak or valley, the maximum value is taken. The algorithm then performs a check to see if the detected peak occurs before the detected valley. In cases where a valley occurs before a peak, both are ignored and the algorithm moves on to the next pair of threshold crossing indexes. Similarly, if a peak and no valley is detected or a valley and no peak, the algorithm will ignore that segment of data and move on to the next.

3.3 Feature Extraction

3.3.1 Respiration Rate

The calculation of RR is built upon the breath detection algorithm seen previously. For each threshold crossing, the respiratory peak occurring right after the crossing is detected. The RR can be calculated by finding the difference between adjacent peaks (in samples), with a simple conversion to breaths per minute at the end. The result is a time varying RR. This RR is then smoothed by a moving average filter

over 4 breathing periods.

3.3.2 Inspiration/Expiration Based Parameters

The combination of inspiration and expiration to form useful parameters has been investigated in a number of previous works. Using the extracted inspiration peaks and expiration valleys the inspiration time and expiration time can be easily calculated. The algorithm will calculate inspiration time as the time difference between a peak and the preceding valley. The expiration time is calculated as the time difference between a peak and its subsequent valley. The total breath period is calculated by finding the time difference between adjacent respiration valleys. Additionally, the peak-to-peak magnitude, or inspiration magnitude can be calculated by finding the difference in magnitude between a peak and the preceding valley.

From these four parameters, it is possible to formulate additional parameters. Two such parameters are *driving* and *timing*. The names for these parameters come from [23] and have been reproduced for ease of understanding. 'Driving' is a parameter that determines how fast the subject inhales. It is calculated by dividing a specific peak-to-peak magnitude by its corresponding inspiration period. Meanwhile, 'timing' is a parameter that measures the ratio of inspiration time to total respiration time of each breath. It is calculated by dividing the inspiration period by the total breath period.

$$Driving = \frac{InspP2P}{InspPeriod} \qquad (3.1) \qquad Timing = \frac{InspPeriod}{BreathPeriod} \qquad (3.2)$$

3.3.3 Unusual Respiration

To detect yawning or any other unusual torso movement, it will be useful to flag respiration peaks that appear unusually large. After inspecting video recordings of the participants under test for large spikes in respiration amplitude, it was concluded that many of the large spikes relate to a yawn or unusual body movement. With the known link of increased yawning and sleepiness, an algorithm to detect these unusual respirations is needed. The implemented algorithm uses a moving filter that finds the weighted average of the previous 10 respiration periods as well as the standard deviation. If a respiration peak is 4 or more standard deviations above the weighted average of the last 10 peaks, the peak is classed as an unusual respiration.

3.3.4 Breathing Rate Variability Analysis

The TEDD index has been identified by two separate papers as a valuable indicator to detect sleepiness onset in people. Following the work done in [24], a similar style of TEDD index is derived.

First, the breathing rate variability (VBB) for an entire recording is calculated. This is the absolute difference between adjacent breathing rate values. The VBB is then smoothed using a moving mean filter over 10 breath periods, and is referred to as

sVBB. Based on the 33 second window of maximum breath stability calculated in the breath detection algorithm, it is possible to calculate the mean VBB for this window (VBBref). Finally, the TEDD score can be calculated:

$$TEDD = \frac{sVBB}{VBBref} \tag{3.3}$$

3.4 Data Segmentation

Since the KSS readings are obtained approximately every 5 minutes, it is difficult to infer any real value from what is happening second to second inside these 5 minute blocks. For this reason, it has been decided that each of the extracted parameters will be segmented into 5 minute blocks that correspond to each KSS period. The median and standard deviation of breathing rate, inspiration time, expiration time, respiration peak-to-peak amplitude, driving, timing and breath period is calculated. For UE, the mean number of events and the median duration of events for each segment is calculated, together with the mean of TEDD. Since the start of the data has been forced to the first instance of BTS data, there is at times, a gap of less than 5 minutes between the start of the data and the first KSS reading. If this occurs, the first KSS reading that contains data for a full 5 minutes prior to its occurrence is used as the first segment of that particular recording. Partial segments that occurred before this are discarded. Doing so ensures that each segment contains data from exactly 5 minutes of recording, allowing for a better basis of comparison. On average there are around 18 KSS segments per drive, thus, occasionally discarding up to the first two segments does not have any major impact on the amount of data available.

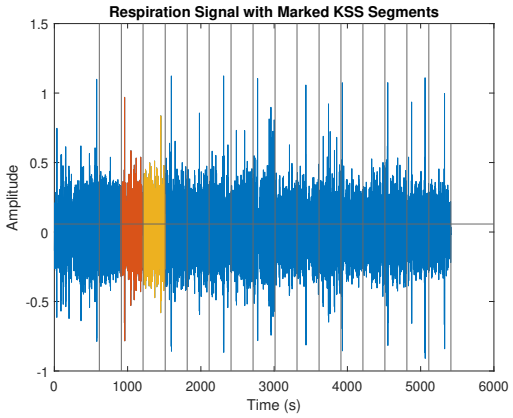


Figure 3.4: Filtered respiration data with KSS segments represented by vertical lines and two 5 minute segments highlighted.

An example of segmented respiration data can be seen in Figure 3.4. Each of the vertical lines denotes an instant at which a KSS reading is taken. The preceding 5 minutes of data before each of these KSS instants is taken as that KSS data

segment. The orange and yellow overlays highlight two 5 minute segments of data. A similar process is repeated for all the extracted feature parameters, after which the mean, median and standard deviation of the feature data in each segment is calculated. The median is used for most features as opposed to the mean due to the skew nature of the distributions. A data table is compiled where each row represents a 5 minute segment of data. The columns are comprised of the calculated feature parameters, the KSS score of each segment, participant number, car automation, day/night condition and the segment number, used as a drive progression indicator.

3.5 Sleepiness Detection

The main methods used for sleepiness detection are three different machine learning algorithms. Before these models can be trained and tuned, the extracted features must be analysed in relation to sleepiness states so as to enable useful features to be selected for classifier training.

3.5.1 Statistical Analysis and Parameter Selection

To have any sort of success in classification, it is important to ensure that the extracted features are of good quality and have a correlation with the driver sleepiness. Before exploring different classifiers for sleepiness detection, statistical analysis of the extracted feature parameters is performed. This serves as an insight into parameter-sleepiness correlation as well as an initial platform to perform parameter selection.

Using the Mann-Whitneys U-test (Wilcoxon rank sum), one is able to tell whether the two sample medians are significantly different or not [46]. The null hypothesis would be that the two samples are from continuous distributions with equal medians, while the alternative is that they are not. A key assumption is made that the two samples are independent (unpaired). In this case, the two samples will be made up of data flagged as sleepy and not sleepy. For each type of extracted parameter, a test is done to check for significance between sleepy and alert states. This can enable a preliminary understanding if there is a link between the observed feature and sleepiness level. However, more will need to be done to select a feature subset to best train the classifier.

Good features are identified by applying forward and backward sequential feature selection algorithms to find a more optimal feature subset. The forward case starts with 1 feature and will sequentially add feature by feature until the best classification result is found. The reverse case starts by using all features and removes features until the best subset is found. The *sequentialfs()* function in MATLAB is used for this task. Additionally, a more basic approach is looked at in which a feature subset is randomly generated. Training and validation is run, performance metrics are saved, and a new feature subset is generated. This is repeated 500 times and the best subset is selected and compared to the sequential selection.

3.5.2 Classification Learning

Previous studies have identified SVM classifiers as good methods for classifying driver sleepiness. Random forest, an ensemble-based learning technique that uses multiple decision trees to perform a majority vote for binary classification, is another interesting and powerful method. There is also KNN which is easy to train and in certain instances can provide a good performance to cost ratio. In this project, SVM, KNN and Random Forests are investigated further for sleepiness classification.

Although each method has its own intricacies and hyperparameters that must be adjusted, the same general training and testing procedure is followed for each. The model training will be in the form of a classical train and validate cycle in which participant level leave-one-out cross validation is used and can be summarised as follows.

1. Select all data from participant i as validation data
2. Train the model using all remaining data
3. Use the trained model to predict sleepiness from the validation data
4. Save any performance indicators from that iteration
5. Use the next participant as validation and train with all remaining data
6. Iterate until all participants have had the chance to act as a validation set

The resulting vector contains the predictions for every drive of all participants. Such a setup is used as when training on physiological data that has high intraparticipant correlation, it can be easy for the model to overfit. Additionally, this approach allows for model hyperparameters to be tuned each iteration in in order to find an optimal subset using a simple grid search method. The test and validation data are the labels for sleepy or not sleepy, with sleepy being a KSS of 7 or above and not sleepy, a KSS of 6 or below. Table 3.1 shows the selected hyperparameters that provided the best results for each of the learning methods.

The predicted sleepiness state is recorded for each validation cycle, with the median accuracy, F1 score, and area under ROC curve (AUC) serving as performance indicators. The accuracy is defined as the rate of correctly predicted class. The ROC (Receiver Operating Characteristic) curve is a representation of sensitivity for a given specificity, and the area of the curve measures how good the classifier is. The F1 score is a measure of classifier accuracy that takes into account the ratio of precision and recall. It can be defined as the following equation:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3.4}$$

KNN	
Distance	Euclidean
Number of neighbours	43
Distance weight	Inverse
Standardization	Yes
SVM	
Kernel function	Radial basis function
Outlier fraction	2%
Kernel scale	Automatic
Standardize	Yes
Random Forest	
Method	Ada Boost
Learning rate	0.2
Learning cycles	300
Max splits per node	..

Table 3.1: Hyperparameters selected that provide the best results for each of the machine learning models.

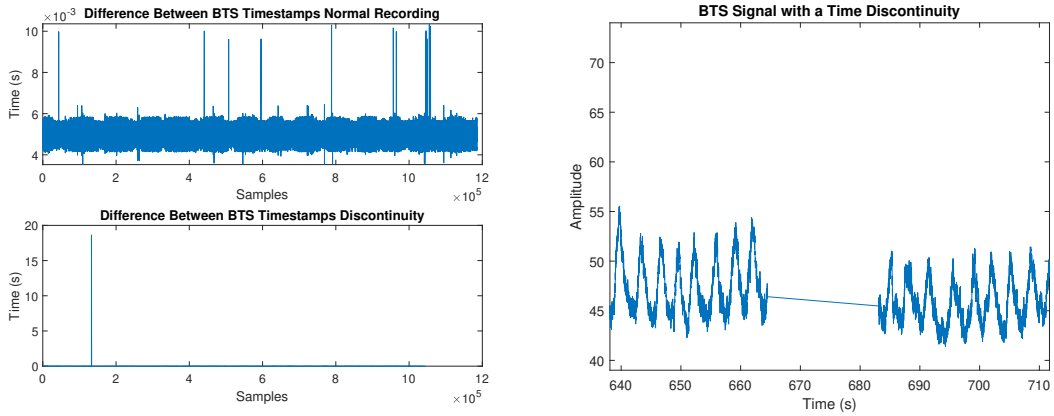
4

Results

Undertaking the processes described in the methods, the following results are observed. The results are roughly broken down into sections according to the main research questions.

4.1 Reference and BTS Comparison

The first step in the project process was to clean the reference and BTS data and perform comparisons between the two to see how effective the BTS data can be. Before comparison occurred, the BTS and reference data was checked for large time discontinuities in recording as this will interfere wit any later processing and comparison. An example of a BTS recording with large time discontinuities is highlighted below. Small discontinuities are accepted, with anything breaching 0.5 seconds becoming problematic.



(a) Plots of two BTS recording's time stamp difference with the second figure showing a case of large discontinuity.

(b) BTS signal for the case of the large discontinuity in (a).

Figure 4.1: Difference between BTS timestamps for a normal recording and one with a large discontinuity shown in (a). An example of a large timestamp discontinuity shown in the BTS signal (b).

For the case of the large time discontinuity of almost 20 seconds in Figure 4.1, the

BTS recording will be discarded, along with the accompanying reference recording. In all, four cases of time stamp discontinuities have been identified.

The next step is to clip the BTS and reference measurements to the same start and end time as well as resample them down to the same time vector. This vector has a sample rate of 8Hz.

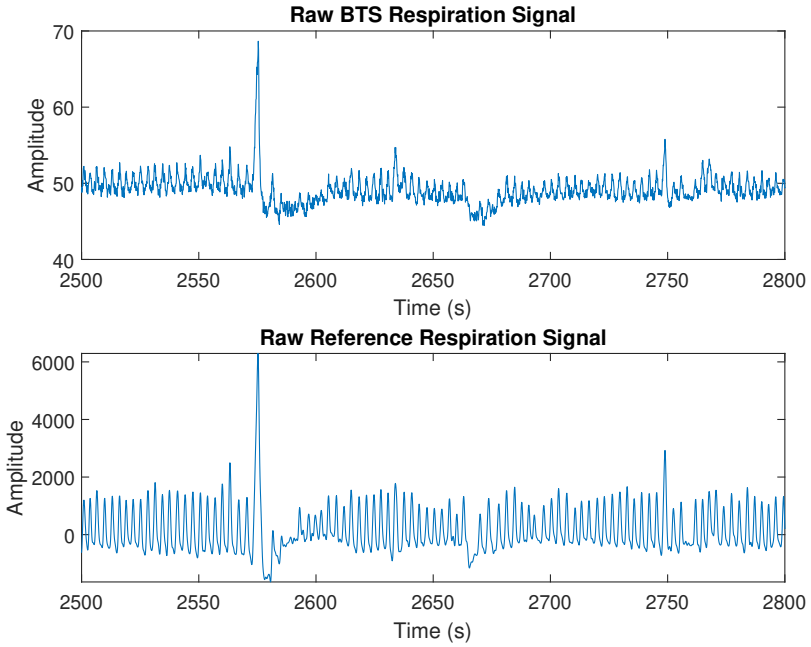
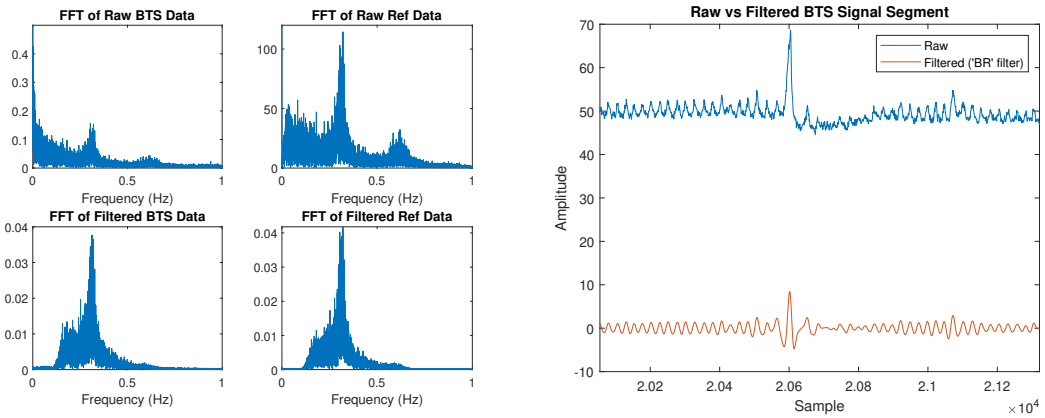


Figure 4.2: Segment of Raw BTS and reference data from a certain recording after time synchronisation and resampling.

As can be seen from Figure 4.2, there are both similarities and differences between the raw BTS and reference data. In this particular segment, the timing and occurrence of respiration synchronises well. Each peak in the data represents the full inspiration, while a valley shows the expiration point. It is however easy to see that the BTS data is far more noisy than the reference, has a large constant DC offset and more short term DC offsets. This is what the filtering stage will seek to remedy.

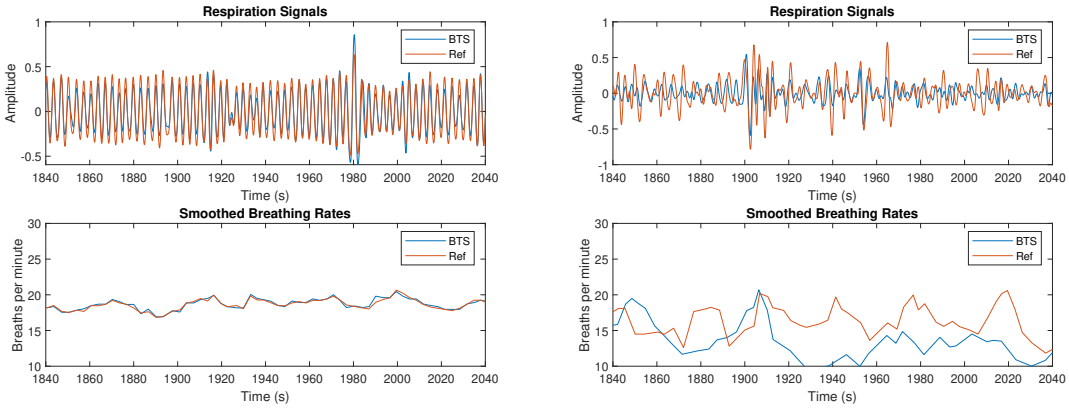
Applying the respiration filtering branch yields the results highlighted by Figure 4.3. In 4.3a, both the BTS and reference data show a large amount of low frequency noise, with the BTS possessing a large DC spike. The magnitude of the BTS spectrum is far smaller than that of the reference, which is to be expected since there it's signal power is smaller than that of the reference. What is easy to see is the spike at around 0.3Hz, present in BTS and reference. This frequency component represents the respiration signal. Applying the designed filter followed by normalization yields the second set of FFTs. The filter successfully attenuates the low and high frequency components while keeping the frequencies in the respiration zone. The input vs output signal for a small segment of BTS data in Figure 4.3b shows the filter successfully removing the DC offset and high frequency noise.



(a) **(b)**

Figure 4.3: Frequency domain representation of the BTS and reference signals before and after applying the BR filter (a). Corresponding filtered and unfiltered time domain signals in (b).

The signal clipping and respiration filter is applied to all BTS and reference recordings, resulting in data that is of equal length, sample rate, and order of magnitude. The two sets of signals are now in a form where they can be compared.

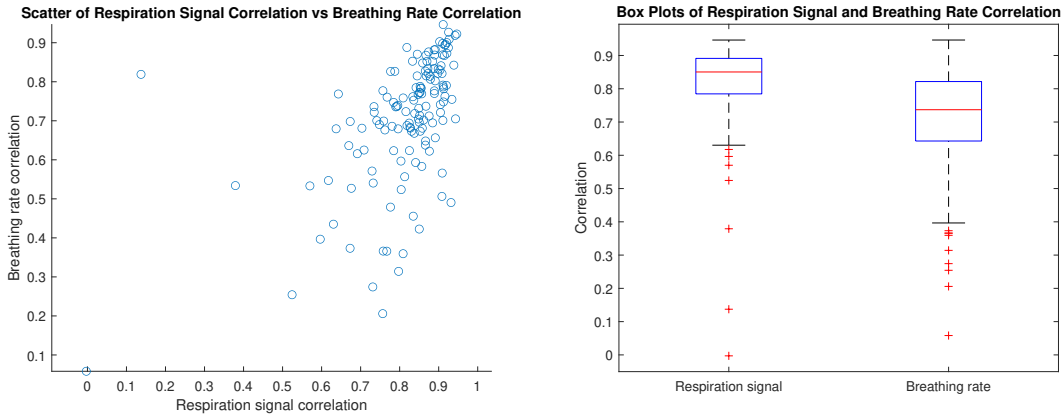


(a) **(b)**

Figure 4.4: Good quality respiration signals and BRs overlaid in (a), with (b) showing a case where the BTS and reference signals did not match up (BTS signal of small amplitude and noisy).

For each of the drives, the Pearson's correlation coefficient for BTS and reference data is calculated. This is done for both the respiration data and BR data. The BR results will be further explored in section 4.2.

Based on the results seen in Figure 4.5, the median correlation for the respiration signals is 0.85, while the median correlation for BR is 0.74. These plots do not show the points with overly poor correlation between the BTS and reference. The pairs with poor correlation were removed from the dataset after manual inspection.



(a) **(b)**

Figure 4.5: Scatter plot showing the correlation of BTS and reference respiration signals vs the correlation of BTS and reference BRs (a). Box plots of the distribution of correlation values between BTS and reference for respiration signals and BRs (b).

Out of the 166 drive recordings, 20 have been removed following correlation analysis and manual inspection. The manual inspection is necessary as not all points of poor correlation result in bad data. For example, the point in the scatter plot with almost zero correlation happens to be due to a constant time offset of 60 seconds between the reference and BTS. Upon inspection, it was decided that this measurement was good enough to keep, even with the poor correlation. This phenomenon is highlighted in Figure 4.6, and can be solved by simply delaying the BTS signal by a minute. Two such cases of constant time offset between the reference and BTS were found, with neither being discarded.

Although not related to the comparison between BTS and reference data, the idea that very tall and short drivers will have an observable impact on the tensions recorded by the BTS system is a valid concern. Figure 4.7 shows the mean value of BTS peak amplitudes for tall, short and in-between height drivers as well as the average base belt tension for the same three classes. There are only 4 participants in the tall and short groups, with all other participants falling into the *in-between* group. The median peak values are 1.5, 3.3 and 3.0 for the short, in-between and tall groups. Similarly, the median values for the base tension are 47.7, 47.8 and 47.9.

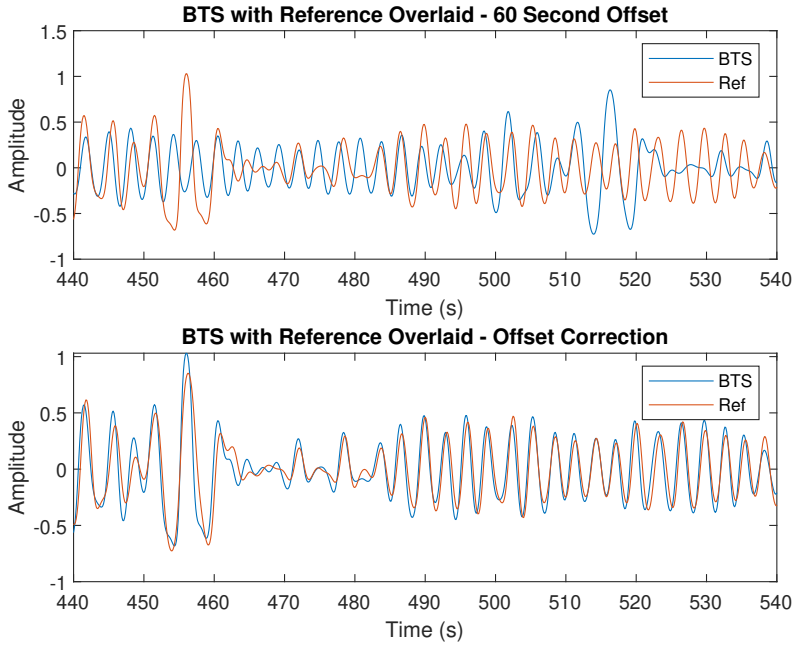
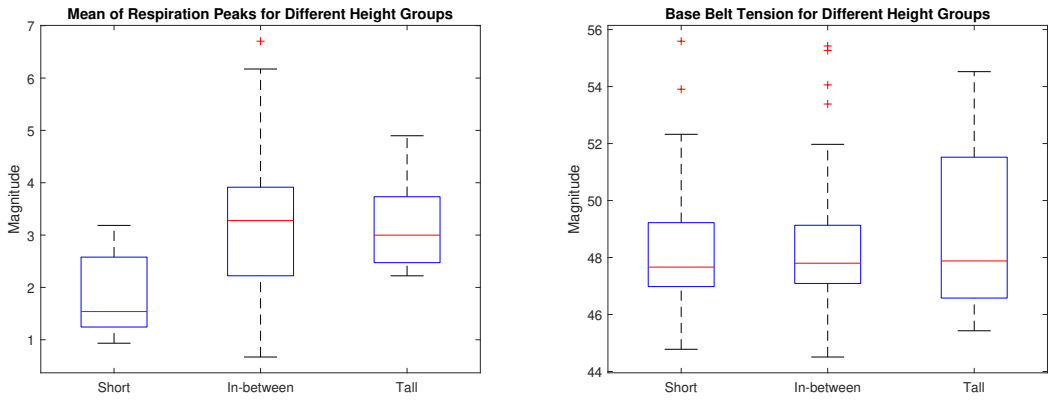


Figure 4.6: Respiration data that had an initially poor correlation due to the reference data lagging the BTS by a constant 60 seconds.



(a) **(b)**

Figure 4.7: Mean respiration peak values for different height ranges in (a) and the average base belt tension for different height ranges in (b). Short: 162cm or less, in-between: 162-180 cm, Tall: over 180cm.

4.2 Parameter Extraction

Based on the methods described for extracting the respiration based parameters, the following results can be observed.

4.2.1 Respiration Rate

Before calculating respiration rate, the occurrence of a breath must be detected. The threshold crossing method is used, with the following figure showing the detected breath and respiration peaks based on using a 60th percentile vs 70th percentile.

Using the threshold crossing method on the respiration filtered signal, the rate of respiration is calculated. Figure 4.8 shows the result of the breath detection algorithm on a segment of respiration data. The BR is calculated by finding the difference in time between adjacent breath peaks (marked by 'x'). The dip in BR around the 2070 second mark is due to the respiration signal failing to cross the threshold, resulting in a bigger time difference between the two adjacent breath peaks.

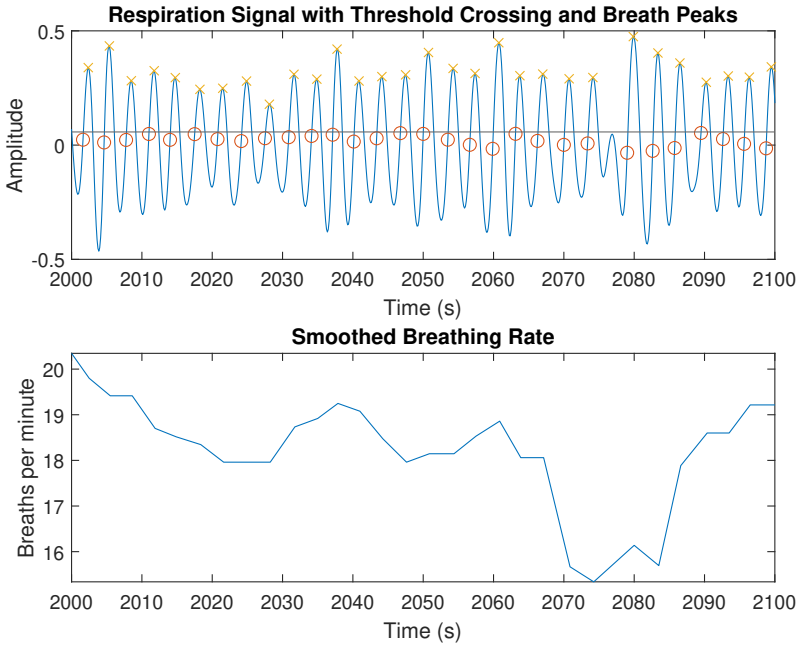


Figure 4.8: Threshold crossings and the associated breath peaks above with the calculated instantaneous and smoothed BR below.

The distribution of median BRs for each of the drives is shown in Figure 4.9. The median is 17.4bpm, with an interquartile range of 3.88bpm, an upper adjacent of 22.6bpm and a lower adjacent of 12.6bpm.

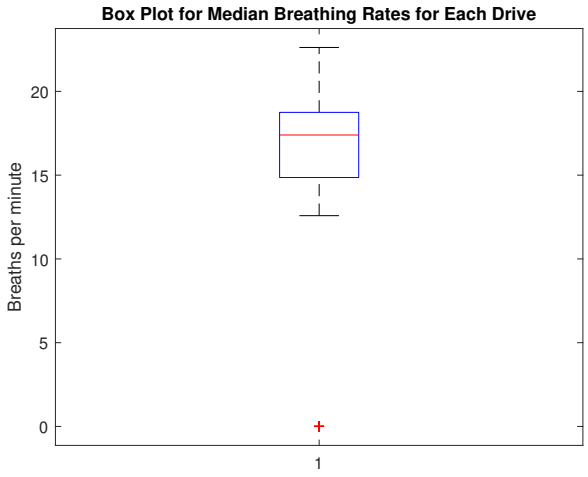
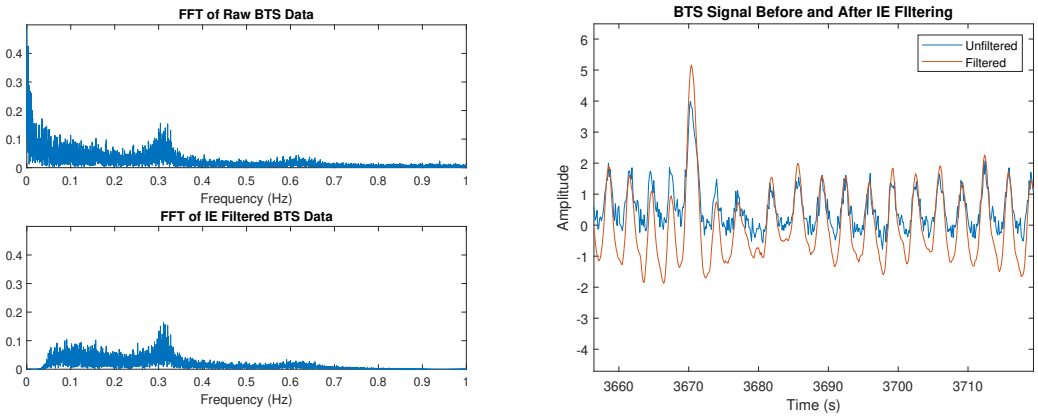


Figure 4.9: Box plot showing the distribution of median BRs for each of the drives.

4.2.2 Inspiration/Expiration Parameters

The calculation of inspiration and expiration time is more sensitive to the shape of the waveform, and thus it is advantageous to keep the filtered waveform as close in shape to the original one. For this, the "IE" filter is used (second filtering branch in 3.2.2). An example of the effect of the IE filter in the frequency and time domain is seen in Figure 4.10.



(a) **(b)**

Figure 4.10: Frequency domain representation of the BTS and reference signals before and after applying the IE filter (a). Corresponding filtered and unfiltered time domain signals in (b).

Using the peak and valley detection algorithm, the respiration peaks and valleys can be detected. In Figure 4.11, the detected respiration peaks and valleys are highlighted. Although the method has been tuned to reject breaths that occur too close together or where a valley occurs before a peak between two threshold crossings, some strange behaviour is still possible. The breath events between 5070 and 5080 appear to be a single breath, while the algorithm finds two separate breaths. The

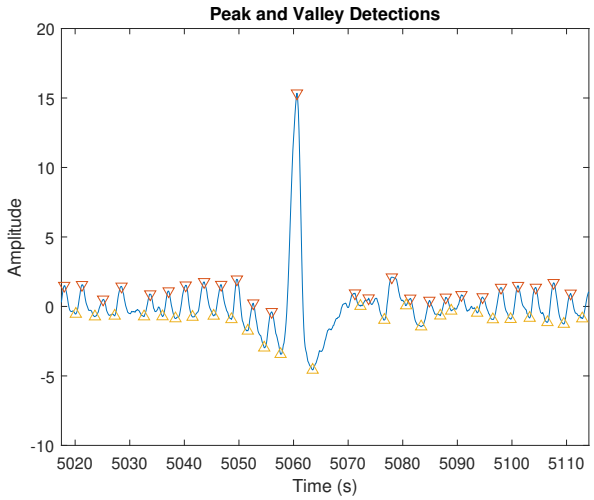


Figure 4.11: The respiration peaks and valleys found by the I/E peak and valley detection algorithm when applied to an IE filtered BTS signal.

inspiration, expiration, breath period, driving, and timing can all be calculated from the detected peaks and valleys. For each of the drives, the median of these four parameters is calculated. The indicators for each set are showin in Table 4.1. The reason for using the median as opposed to the mean is that the distributions of the parameters are fairly skewed and contain a number of outliers. This is due to the case where the IE algorithm rejects a peak and valley if the valley comes before the peak. The resulting time difference between the adjacent peak or valley is thus large and will result in a spike in inspiration or expiration time. For each drive, the median inspiration-expiration ratio is shown in Figure 4.12.

Table 4.1: Indicators derived from the median inspiration expiration parameters.

Indicator	Median	IQR	Upper and Lower Adjacent
Inspiration period	1.5s	0.25s	2.0s & 1.25s
Expiration Period	1.69	0.38s	2.75s & 1.13s
Timing	0.48	0.042	0.53 & 0.4
Driving	3.18	1.45	5.77 & 0.84

4.2.3 Unusual Respiration Events

Following the UE algorithm, if a peak is n standard deviations above the weighted average of the previous 10 respirations, it is considered an unusual event. Using a deviation of 4 standard deviations, the detected UE for a full drive is seen in Figure 4.13. Figure 4.14 shows two detected UEs in more detail. The first is of a yawn while the second is a torso twist accompanied by a large inhalation. In all cases, the events have been verified by video comparison.

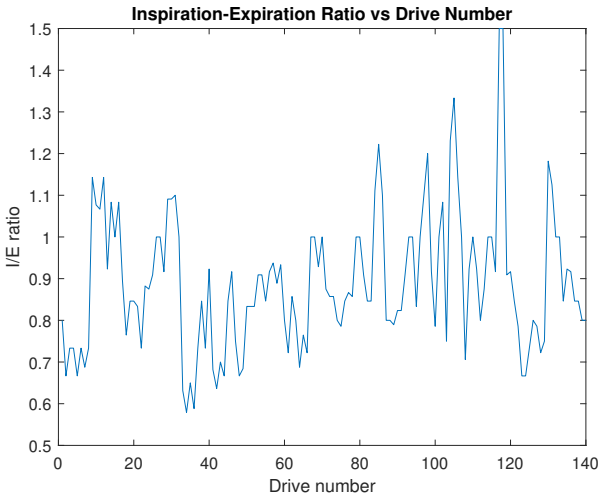


Figure 4.12: Median I/E ratio values for each of the drives.

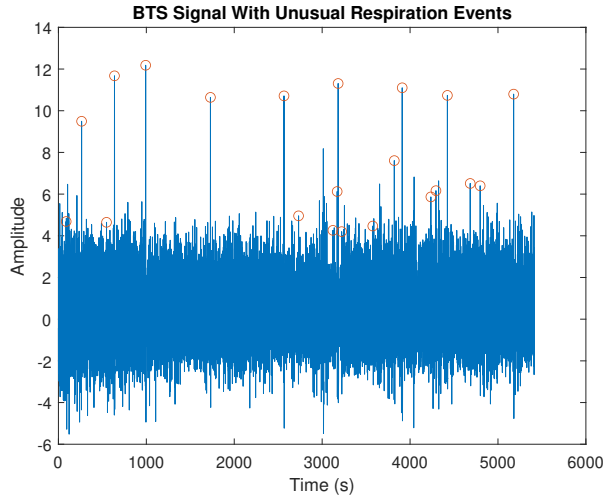
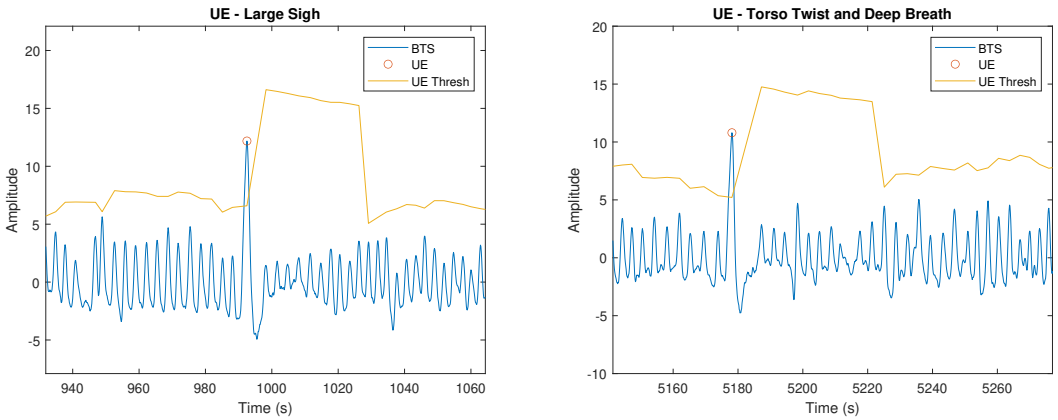


Figure 4.13: Detected UEs for the BTS respiration signal over a full drive.



(a) **(b)**

Figure 4.14: Two individual unusual respirations with the threshold overlaid. (a) represents a yawn, while (b) is that of a torso twist and large inhalation. The occurrence of the events was verified by examining the video recording.

4.3 Classification Learner

The following results show how each feature extracted from the respiration data links to sleepy and not sleepy states as well as how well the developed classifiers are at distinguishing between a sleepy or not sleepy labeled segment of data.

4.3.1 Parameter Analysis

Table 4.2 shows the results when a Wilcoxon rank-sum test is applied to each of the parameters using a 5% significance value. The highlighted red results are for cases that do not fail the null hypothesis, and can be classed as having a statistically insignificant difference in data distribution for sleepy and not sleepy.

Table 4.2: Rank-sum test results for feature parameters, with a null hypothesis that the medians for each feature are the same in sleepy and not sleepy states.

Parameter	**p-value**	Median not sleepy	Median sleepy
Median BR	0.0127	17.91	17.71
Std BR	2.00e-09	1.96	2.14
Delta Mean BR	**0.202**	-	-
Median Insp	4.52e-06	1.5	1.5
Std Insp	3.08e-15	0.87	1.05
Median Exp	**0.108**	1.75	1.625
Std Exp	2.37e-08	0.65	0.71
Median P2P	8.73e-10	4.81	5.3
Std P2P	4.73e-40	1.98	2.55
Median Driving	5.89e-05	3.03	3.21
Std Driving	7.33e-26	1.28	1.48
Median Timing	5.51e-06	0.48	0.48
Std Timing	8.87e-12	0.11	0.12
Median BP	0.0311	3.25	3.38
Std BP	1.13e-14	1.19	1.33
UE Count	3.08e-05	2	2
UE Period	1.14e-04	9	10.1
Median TEDD	3.74e-13	1.41	1.66

A phenomenon that has been observed, is that the drive condition (day drive vs night drive), has a large and consistent correlation with sleepiness level. Drives that occur at night show an increased average level of recorded KSS scores. This relationship is highlighted by Figure 4.15. The last KSS segment to observe is chosen as 18 since there are a small number of drives containing more than 18 segments. Including these can lead to underrepresented results.

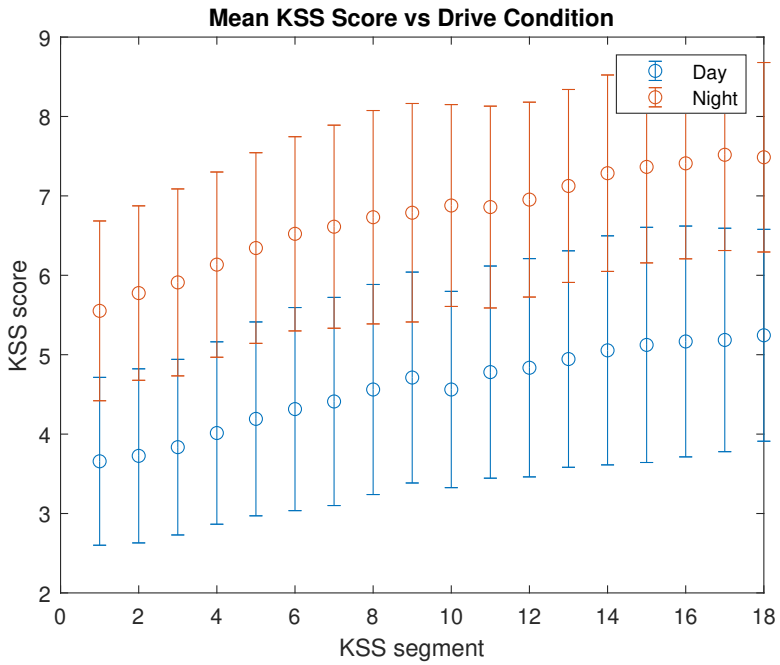


Figure 4.15: Error plot showing how the mean KSS score for each segment of the drive progresses.

Figures 4.16, 4.17, 4.18 and 4.19 show how the median of different parameters change during the drive as well as how the data is distributed between sleepy/not-sleepy and day/night conditions. The error plots (subfigure (b)) are made up by calculating the median or mean of all values for each 5 minute KSS segment number. This enables the data to be visualised in a drive progression format. Similar plots for all features in Table 4.2 have been included in Appendix B. In all cases, a sleepiness value of 0 (negative) represents not sleepy and a value of 1 (positive) is sleepy.

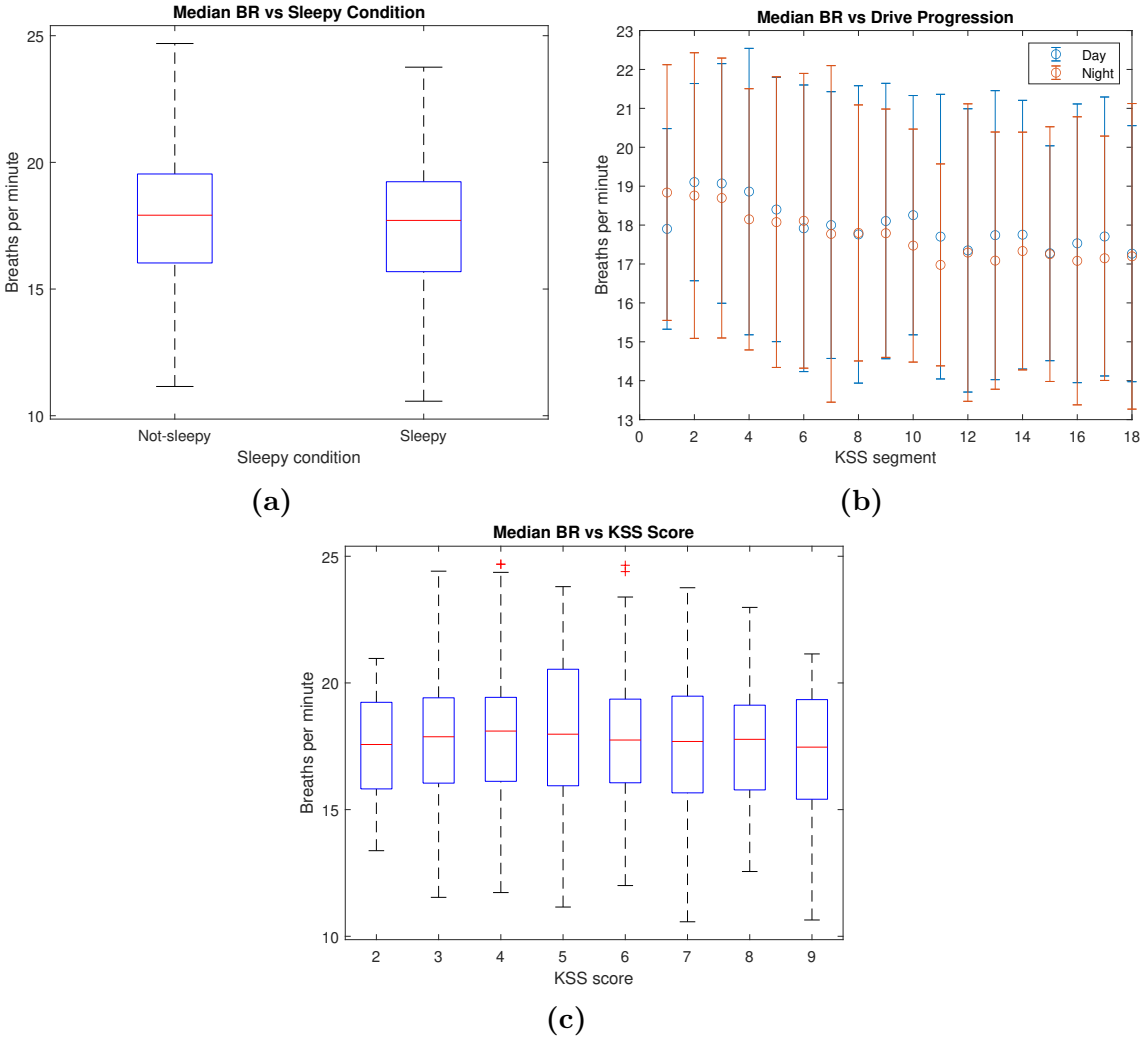


Figure 4.16: The changes seen in median BR for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

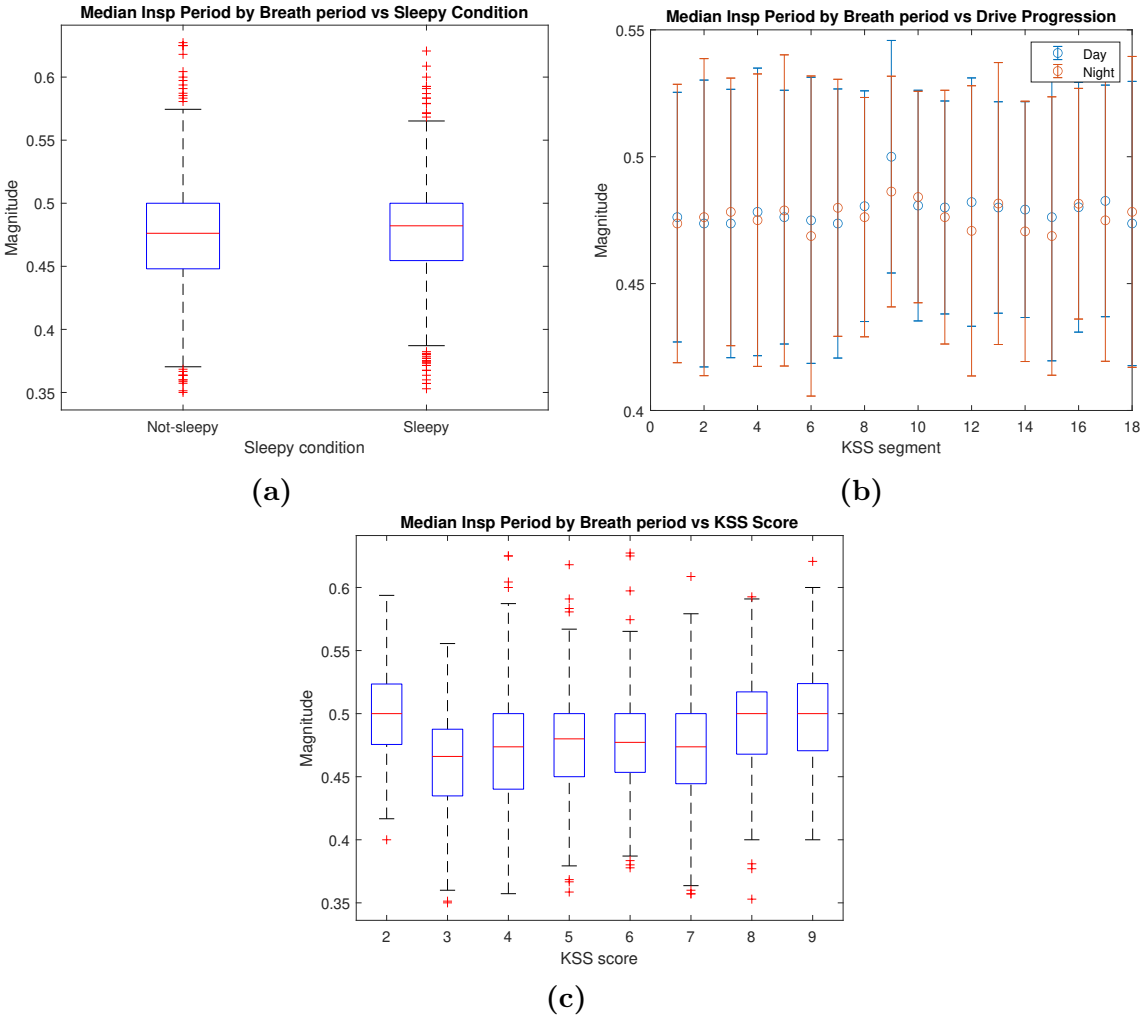


Figure 4.17: The changes seen in median inspiration period to breath period ratio (timing) for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

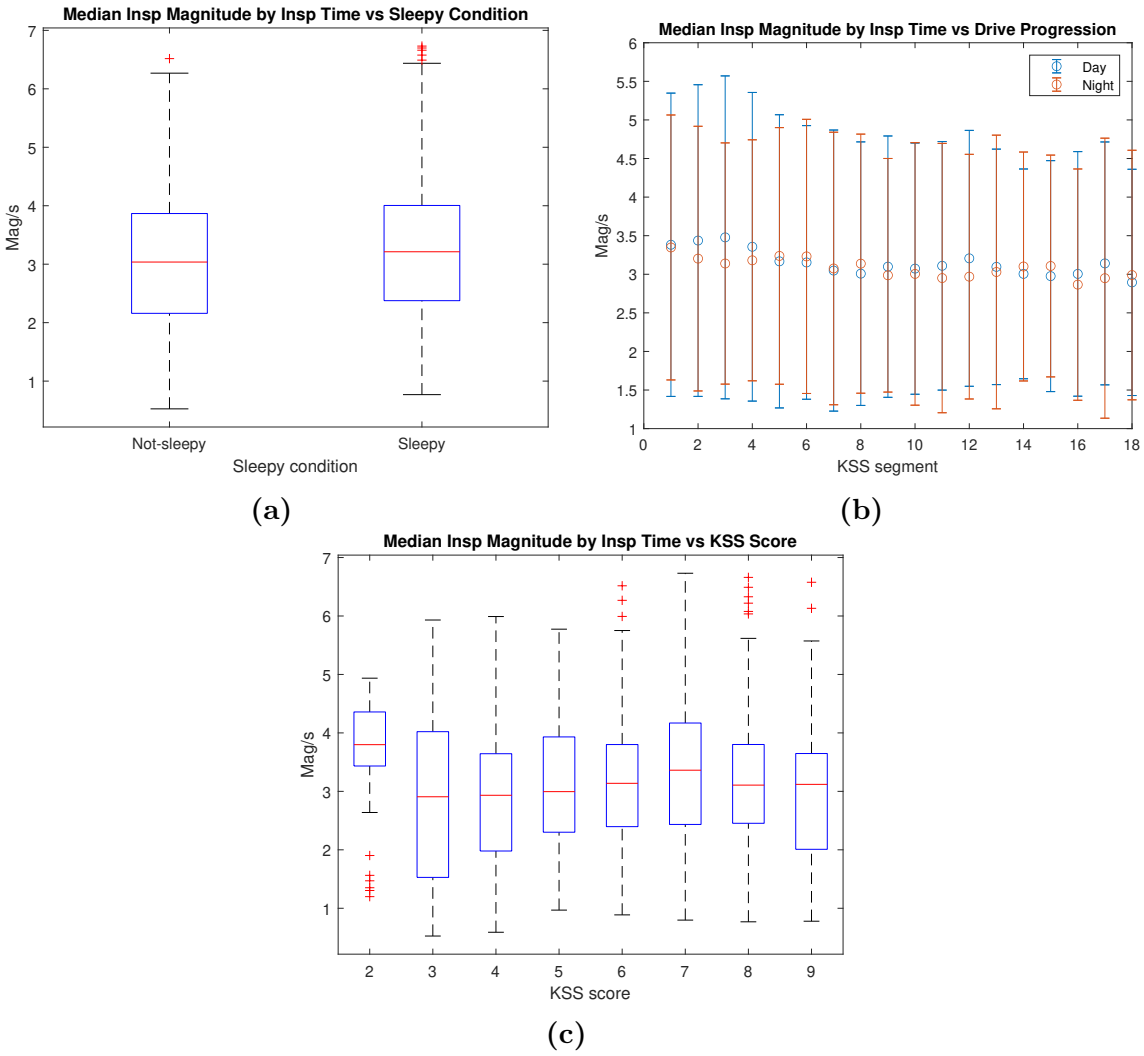


Figure 4.18: The changes seen in median inspiration magnitude to inspiration period ratio (driving) for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

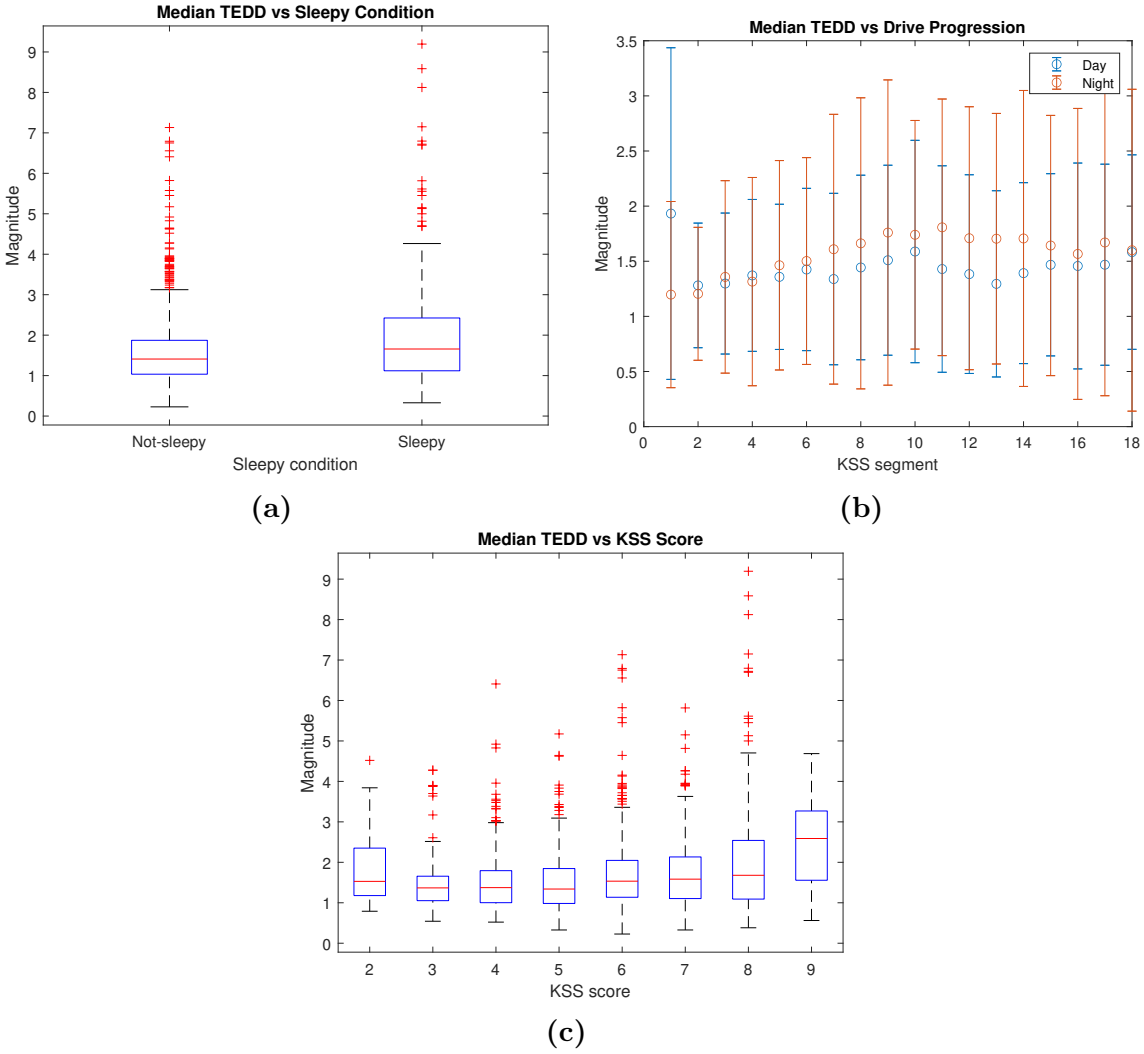


Figure 4.19: The changes seen in median TEDD values for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

An example of data that has a highly significant difference in sleepy vs not sleepy distribution is shown in Figure 4.20. On the other hand, 4.21 highlights an example of distributions that the Rank Sum test considers insignificantly different.

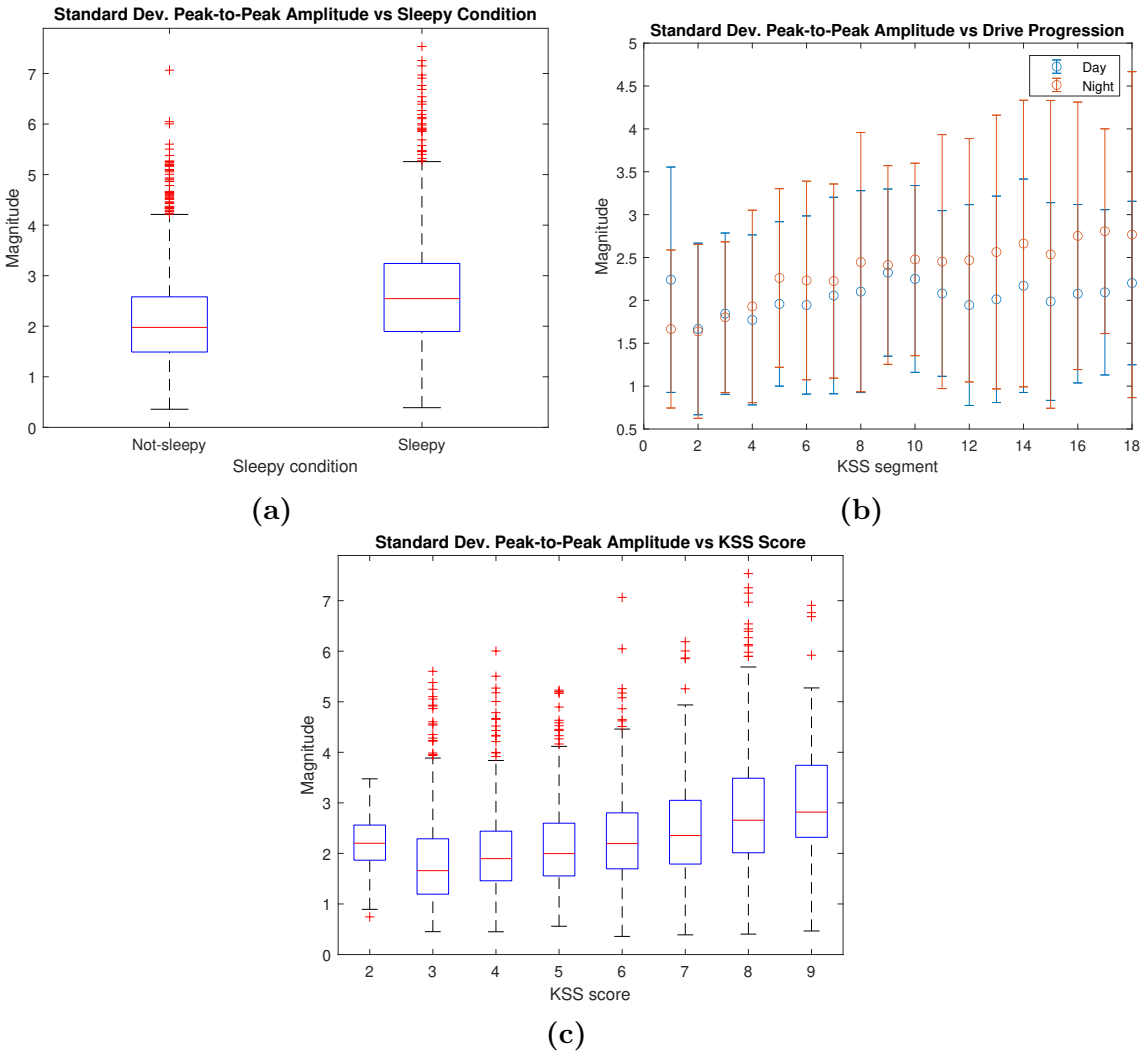


Figure 4.20: The changes seen in the standard deviation of breath peak-to-peak amplitudes for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

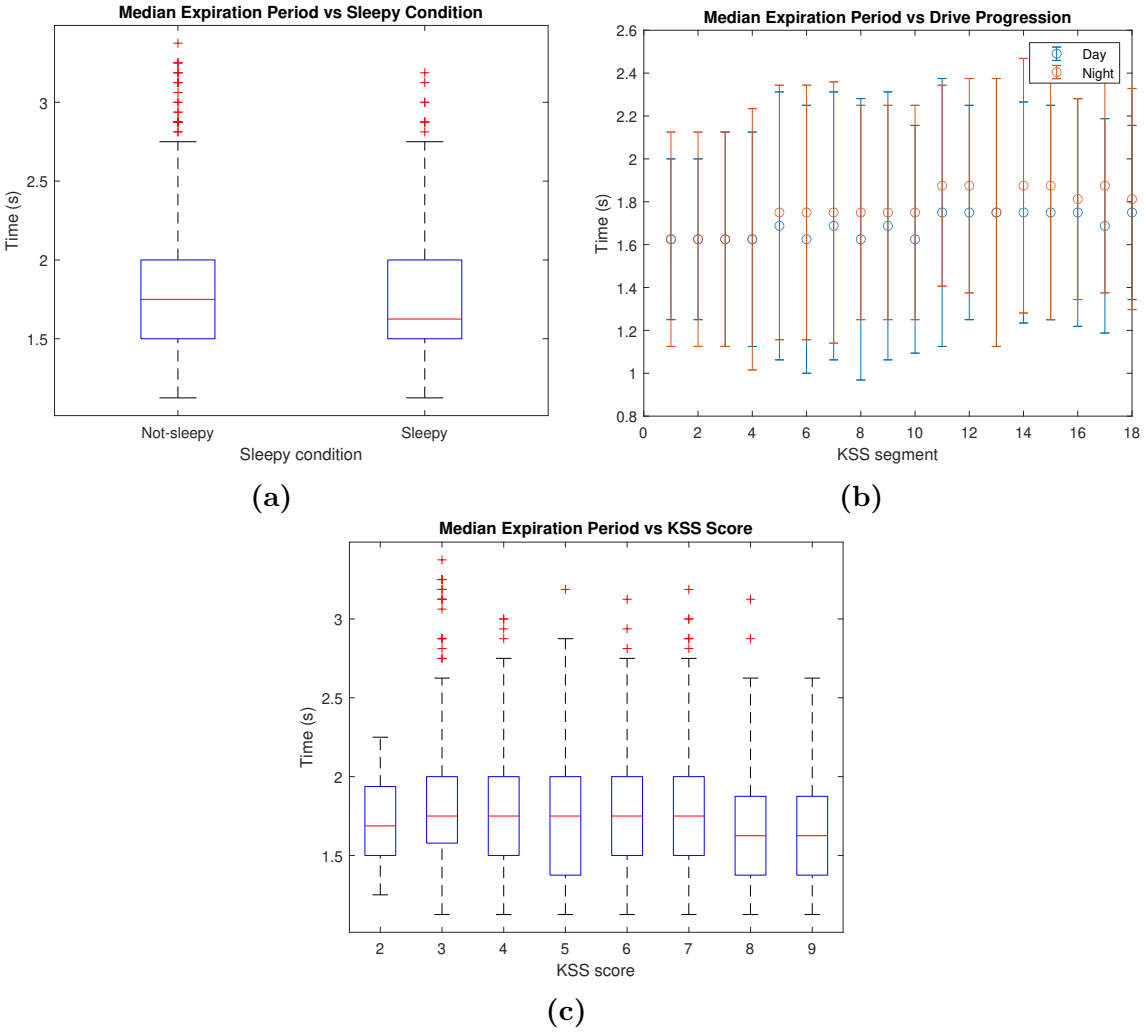


Figure 4.21: The changes seen in median expiration times for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

4.3.2 Classification

To train the classifier efficiently, a set of features is selected from all available breathing features. The selected features are shown in Table 4.3. Two additional features, day/night condition and segment number, are used in separate training sessions. Classification accuracy, F1 score and area under ROC curve (AUC) are shown in Table 4.4. The associated confusion matrices for KNN, SVM and random forest can seen in Tables 4.5, 4.6, and 4.7. The results in which day/night condition and segment number are added to the extracted features for training are labeled **Extra** while the results that use only selected features are labeled **Normal**. For the sake of comparison, the results from classifiers trained only on day/night condition and segment number are also shown with the **None** label.

Table 4.3: Selected features for classifier training

Features	Description
Mean BR	Mean of BR
Median BR	Median of BR
Var BR	Variance of BR
Median Insp	Median of inspiration period
Std Insp	Standard deviation of inspiration period
Median Exp	Median of expiration period
Std Driving	Standard deviation of driving
Median Timing	Median of timing
Std BP	Standard deviation of breathe period
UE Count	Unusual respiration event count

On average, the KNN classifier has the best accuracy, F1 score and AUC among the three classifiers, while the SVM and random forest are slightly above/below each other depending on the training features used and observed score. The amount of datapoints for sleepy and not sleepy segments in the experiment is:

- Sleepy: 885
- Not sleepy: 1623

Table 4.4: Performance comparison between three classifier algorithm.

Classifier	**Accuracy**	**F1**	**AUC**
Normal			
SVM	0.62281	0.42824	0.57755
KNN	0.64872	0.50543	0.55902
Random Forest	0.62919	0.45747	0.56143
Extra			
SVM	0.76396	0.68521	0.79849
KNN	0.77990	0.69704	0.83767
Random Forest	0.75319	0.66220	0.81123
None			
SVM	0.80622	0.71097	0.79316
KNN	0.80742	0.71158	0.84365
Random Forest	0.80742	0.71158	0.84404

Table 4.5: Confusion matrix for KNN classifier.

Class	Predicted: Awake			Predicted: Sleepy		
	Normal	Extra	None	Normal	Extra	None
Actual: Awake	1441	1367	1349	182	256	274
Actual: Sleepy	699	296	209	186	589	676

Table 4.6: Confusion matrix for SVM classifier.

Class	Predicted: Awake			Predicted: Sleepy		
	Normal	Extra	None	Normal	Extra	None
Actual: Awake	1388	1373	1349	235	250	274
Actual: Sleepy	705	344	211	183	541	674

Table 4.7: Confusion matrix for random forest classifier.

Class	Predicted: Awake			Predicted: Sleepy		
	Normal	Extra	None	Normal	Extra	None
Actual: Awake	1336	1346	1349	287	277	274
Actual: Sleepy	643	342	209	242	543	676

5

Discussion

5.1 Respiration Measurements for Sleepiness Detection

The question whether it is possible to use respiration measurements to determine driver sleepiness is an important one as it sets the basis for the entire project. The results from previous papers indicate that it is possible to differentiate between sleepy and nonsleepy drivers using respiration measurements. There are a range of parameters that have been used, including, but not limited to, the mean and variance of breathing rate, inspiration/expiration ratio, and other similarly derived parameters. In general is it found that as drowsiness levels increase, the rate of respiration will decrease, while the variability will increase. The UE duration, ID and ED are expected to increase as sleepiness levels progress. The I/E ratio is considered useful but can have slightly conflicting results at times.

Although these previous results look promising, it is important to note that the experimental setup for the vast majority is in a highly controlled environment with relatively few participants. This includes driving simulators and real-world driving under controlled observation such as closed roads or tracks. Such setups are more ideal than a natural driving situation, and thus, the resulting performance may be skewed towards an unrealistic level of accuracy.

Additionally, the respiration features emerge from a mix of task-related fatigue and sleepiness. This phenomenon is visible in Figure 4.16 where breathing rate is inversely correlated with drive progression in both sleepy and non-sleepy participants. An improvement in data processing might help isolate sleepiness from the respiration feature.

5.2 BTS Ability to Record Respiration Data

The second research question is whether the BTS can accurately detect driver respiration. Based on the observed results, the BTS sensor is able to detect a time-varying signal of driver respiration. After removing unwanted frequency components, for-

matting the reference and BTS data to be comparable, and removing recordings with evident problems, there is a high similarity between the signals recorded by the reference band and the BTS. This is highlighted by a median correlation value of 0.85 obtained after correlating the reference and BTS signals from each drive. Such a metric cant tell us exactly how good the quality of the system is, only that it agrees with the reference band results, which is assumed to be of acceptable accuracy.

Although a correlation of 0.85 may seem slightly low, in the context of comparing two physiological signals recorded using completely different systems, this result is promising. No matter how perfect the quality and filtering of the two signals, there will always be slight differences that the correlation method picks up on. The reference band is specially designed to detect respiration signals, and therefore (in theory) a trustworthy benchmark. Past works have shown that the placement of the chest band can alter the quality of the recorded signal. There are cases of the reference band obtaining poor recordings, however, these are few and far between and not frequent enough to discredit this as our benchmark. If the signal recorded by the BTS agrees with the reference, one should thus be confident that the BTS is indeed an accurate tool for respiration measurement. This is reinforced by the fact that when the frequency spectrum of each is examined, there is a noticeable peak around the region one would expect the breathing rate to fall.

In a handful of BTS recordings, large timestamp discontinuities are present. This has been shown to be highly problematic in severe cases. A likely reason for this is in the data collection and logging system and has been addressed by those setting up the hardware. Additionally, the raw BTS signals have a large DC offset which the reference does not. This is due to the base tension each participant excerts on the seatbelt when they are sitting in the drivers seat. Thankfully, this as well as any slow amplitude change in the signal caused by slow movement is possible to be filtered out.

For the recordings in this experiment, the overall quality of the BTS signal looks good. However, the reliability may be impacted by the base tension of the seatbelt. It remains to be seen how the BTS performs in a case with reduced static belt tension. With such a setup, the tension seen by the BTS will likely be reduced, thus lowering the overall sensitivity of the system. This will impact on the performance of extracted features that rely on measuring a change in amplitude of the respiration signal. One such feature would be the driving, in which the inspiration change in amplitude is divided by the inspiration duration.

The stature of the subjects also affect the BTS signal quality. Ideally, the diagonal part of the seatbelt should contact the torso in as much area as possible. Several cases are considered where the contact area is reduced. First, a tall subject with long legs or long reach might sit farther back to the point that part of their torso is positioned behind the B-pillar. Another case where the contact area is minimized is where the subject is short and the seatbelt's vertical adjustment located on B pillar is not adjusted to the subject's body height. Suboptimal contact between the

torso and the seatbelt will in theory result in lower belt tension. Figure 4.7 gives a rough overview of what one can expect. The results highlight the lower observed peak magnitudes and thus belt tension in the shortest participants. The inbetween and tallest groups show more similar values. Interestingly, the base belt tension for all three groups shares an almost identical median value. Combining these two observations, it appears that the height of the participant is not overly affecting the average tension. Shorter people are likely to have smaller chest movements and thus lowered inspiration peak tension. Unfortunately, the tall and short groups contain only four participants each and thus the findings are not entirely trustworthy. More research will need to be conducted to find out exactly how the stature of a participant is affecting the BTS data.

5.3 Feature Extraction

The question here is given that the BTS is evidently a viable method to record respiration data, can useful features be extracted from this data? The results show that extraction of breathing rate, inspiration/expiration times, number of unusual breathing events and the variability of the breathing rate are possible.

Calculating the breathing rate median for each drive yields an interquartile range of between 14.86 and 18.74 breaths per minute. This is exactly in the range of normal breathing rates. The correlation between the BTS calculated breathing rates and those from the reference data has a median value of 0.74. This is lower than the correlation between the respiration signals, and can likely be attributed to the fact that if a breath is detected in BTS, but not in the reference, then the breathing rate at that breath location will be substantially higher for the BTS. Averaging over each 5 minute KSS segment will subdue these short time fluctuations.

Using the IE filter, the results show that it is possible to remove low frequency DC and high frequency jitter components from the BTS signal while maintaining the original shape of the waveform. The peak and valley detection algorithm is shown to work well, and thus the resulting inspiration and expiration times will only be limited by the quality of the recorded signal. Observing the inspiration and expiration times shows that there is no consistent ratio between the two. It is expected that for a normal person, the ratio will be around 1:2 or 0.5. On average, the ratio seen from the results is at around 0.88 (Table 4.1 and Figure 4.12). One reason for this discrepancy is the way the inspiration and expiration period is defined. It is possible that the developed method is finding a premature end for the expiration period as it currently looks for the minimum point in the expiration valley. In reality the end of the expiration period may be slightly in front of this valley peak and thus the calculated expiration period is shorter than what it really should be. Figure 4.10 highlights how the raw BTS has infact a region of flatness in each expiration valley, while the filtered signal does not. If the filtering could be imporved to keep this flat region while also removing high frequency noise, the IE ratio could potentially be more accurately calculated.

Detection of unusual respiration events is also possible. Comparing the events detected by the algorithm with those at the same time instant in the video recording showed that the detected events almost always corresponded to a yawn or large body movement. The threshold definition for the detection of UE peaks in the respiration signal is the main parameter that has an effect on the algorithm's results. As mentioned, an exponentially weighted moving average of the previous 10 respiration peaks is used, with a UE occurring if a peak is 4 or more standard deviations above the average. This value was selected based on a visual analysis of the algorithm's performance for different threshold levels. With such a method it is difficult to tell what sort of UE is being observed as the waveform shape for a yawn or large breath before brief speech is very similar.

5.4 Parameter Analysis

The final question of this project is whether it is possible to develop an algorithm to detect sleepiness and use it on the BTS data. Answering this question involves two steps. The first is to assess the correlation between extracted features and sleepiness levels, and form a method for feature selection. The second is to use these features and train various models to perform classification.

The first area of analysis, the rank-sum test, shows that all features except *delta mean BR* and *median exp* have a statistically significant difference in average value. Upon examining the median values for each feature in the sleepy and not sleepy states, there are cases where the statistical test shows there is significant difference in the distributions, while the value of the two medians is almost exactly the same. An example is for the *median timing* feature, where the median value for both sleepy and not sleepy states is 0.48. Upon examining the boxplots for this feature (Figure 4.17), the distributions look almost the same over the interquartile range. This highlights the fact that the result of the rank-sum test must not be taken as law and should always be further investigated.

An interesting observation encountered is the significant impact the day/night drive condition has on KSS scores and thus sleepiness. This is highlighted by Figure 4.15, where the KSS scores differ significantly as the drive progresses during the day vs night. Such a trend follows what was found in [45]. This trend may be exaggerated by the way experiment was conducted, however, one would expect people driving at night to be under a larger sleep pressure induced by the circadian rhythm. For this reason, a plot of the drive condition vs feature values is included for each feature. Although not a perfect representation of sleepiness, there is a strong correlation, with an added benefit that the sample size for day/night is much more even than sleepy/not sleepy.

Previous works have found that the rate of respiration decreases the more sleepy a driver is. The results shown in Figure 4.16 show that there is a very slight decrease in the respiration rate in sleepy cases vs not sleepy cases. The difference is small, and not as pronounced as in previous studies. Moreover, the night vs day does not

show much difference, as do the values for each KSS bin. This can be down to a number of factors, a major one being that this study has used real world driving experiments with a large group of participants. Far from the controlled lab setups with small sample sizes often seen in other studies. One observable trend though is that the respiration rate tends to decrease in both day and night cases as the drives progress. This is interesting as there is the trend for increasing KSS scores as the drives progress (Figure 4.15). This means that there is a decrease in respiration rate as KSS values increase.

For the two inspiration/expiration based feature, *timing*, the results seem fairly inconclusive. This feature shows the ratio between inspiration time and total breath period. In Figure 4.17b, there is no real trend for day vs night drive condition. Looking at the distributions in (a), there is a slightly higher median and lower quartile value for the sleepy state. The progression as KSS increases shows a tentative increase in value. For *driving*, an indicator of respiration size over time, the general trend seems to be that the value decreases marginally as the drives progress. Day/night conditions and progression as KSS increases in value yield little insight. It must be noted that observed results will be directly affected by the inspiration and expiration time, for which the calculation method has been show to have room for improvement.

It is expected that the value of TEDD (breathing rate variability) will increase when a person is becoming more sleepy. From the results, this is indeed the case, with the value for sleepy cases being consistently higher than for non-sleepy cases, this is promising and is showing similar trends to what was seen by previous works. Similarly, the values for a night drive are higher than those of a daytime drive. It seems that the higher the KSS value, the higher the TEDD score, which is inline with the results from sleepy/not sleepy. Again, the difference is not as large as what is found in previous studies. A factor here is that the magnitude of TEDD relies heavily on the scalar derived from the window of minimum variability at the recording's start. With each of the recordings starting with the participant already driving on the open road, this period of minimum variability is going to be quite different every time compared to if the experiment is done in the same controlled simulator.

The standard deviation of peak-to-peak breath amplitude (Figure 4.20) showed the lowest p-value in the Rank Sum test. Observing the difference between sleepy and not sleepy for this features shows that the values are most definitely higher for sleepy participants. The same trend is observed for day vs night drives, with consistently higher values seen during night time driving. On the other hand, the median expiration period showed a rather insignificant p-value, which would lead one to expect little difference between not-sleepy and sleepy values. When looking at figure 4.21a, there is in fact a reduction in the median value when sleepy, although the rest of the distributions are almost identical. This could explain the insignificant result in the test. Observing the results in (b) and (c), it is strange to see that there is a slight increase in expiration period as the drive progresses, while a marginal decrease as

the KSS values increase. At this time, we are not sure why this is being observed, but can gather that this feature (in its current form) will not work very well on its own as a link to sleepiness.

There are several features that have visible differences between sleepy and not sleepy states, as well as day/night drive conditions. Unfortunately, there is an ever persistent overlap of values in different classes. Such an overlap is likely caused by a high variance in measurements. This variance can be attributed to the nature of performing a real world experiment with a large group of different individuals. Additionally, the self scoring system of the KSS is subjective and although in theory one of the best ways to measure sleepiness, will always result in inaccuracies being carried through.

5.5 Classification

Based off of these results, one would expect that the features that show the biggest difference between sleepy and not sleepy states will produce the best results when used in a classifier. Indeed, if the classifier is examining a purely linear relationship, this should be largely true. However, in the case of the classifiers used, this is not so simple, as there are far more complex patterns evolving as more than one feature is used. Initially, the built in MATLAB sequential feature selection functionality was used to select features. This proved problematic when the validation was switched to a participant-level leave-one-out approach as this is group-based and MATLAB's cross validation is unable to function in the way needed. For this reason, a more basic approach was used in which the full training process was done 500 times, each with a random selection of features. The feature subset resulting in the best performing classifier is selected as the final feature subset to use. It was decided to use the same feature subset for all three classifiers so as to keep the comparison as equal as possible.

Using the best subset of features and with the accompanying classifier hyperparameters, the three different classifiers were trained. The overall validation accuracy, F1 score and AUC are shown in Table 4.4. If only the extracted features are used to train the classifiers (and not the day/night condition or drive segment number), the results are all fairly poor. When those two extra features are included, all three performance scores increase by at least 10%. Most interestingly is that when only the day/night condition and drive segment number are used as features, the classifier achieves the best results. This highlights the significance of the day/night condition in the experiment. Although an interesting observation, it would be best practice to exclude these two features from the classification models as it is heavily biased towards the collected dataset. Additionally, it would become dangerous to have such weight placed on these two features in a real driving situation. There is always the possibility that drivers start a drive already sleepy, even during the day. Therefore, it will be important that they are assessed as similarly as possible to someone getting in the car at night, who may in fact not even be sleepy. Plots for the resulting ROC curves for each classifier can be seen in Appendix B.19, B.20,

B.18.

On average the classifiers are better at predicting not sleepy segments (true negative rate) as opposed to true positive rate (sleepy segments). Class imbalance was not an issue, since the support points between awake datapoints and sleepy datapoints sits at 65% to 35%. Then again, the best accuracy among all classifier when using only breathing features is at 64.87%, which sits around the number of awake datapoints. Based on the fact that a number of extracted features showed a noticeable change between sleepy and not sleepy states, one would expect the classifier to perform better. A potential problem could be due to the high variance and overlap of extracted feature values for sleepy and not sleepy states. This overlap can lead to difficulties in establishing defined decision boundaries or trees, and thus lower performance.

6

Conclusions

With reference to the research questions, obtained results and discussion, the following conclusions and future recommendations can be made.

6.1 Conclusions

Is it possible to use respiration data to differentiate between driver's sleepiness levels?
It is clear to see from the previous research that it is indeed possible to differentiate whether a driver is becoming sleepy based off of respiration data. The conclusion is made that this is possible, albeit with an ideal experimental setup used by the majority of studies.

Is the BTS system able to function as a viable respiration measurement platform?
Based on the comparison results between the BTS and reference chestband, it is safe to conclude that there is good similarity between the BTS and reference data in the majority of recordings. Therefore, the BTS can be considered a viable system to record respiration (for this experimental setup).

Can useful parameters with a proven link to sleepiness level be extracted from the BTS data?
It has been shown that breathing rate, inspiration/expiration based features, unusual respiration events and the variability of breathing rate are possible to extract from the BTS data. Of the extracted features, several have a marginal to significant difference between sleepy and non-sleepy states, while others have very little correlation. It is concluded that although not perfect, a link between respiration features and sleepiness level is observable in a number of extracted features.

Is it possible to develop an algorithm based off of previous research that can be applied to the BTS data?
The developed classifiers have been shown to process some ability to classify between sleepy and non-sleepy feature segments. If no day/night condition or drive progression number is included in the features, then the performance is quite poor.

Contrarily, if these features are included, the classifier performance increases dramatically. The conclusion is that it is possible to use the extracted features to perform a classification, but the performance is lower than expected.

6.2 Recommendations and Future Work

An obvious recommendation is to perform more data collection to further improve the dataset. This will allow for a more diverse group of data to be analysed and processed, giving an even better real world view of what is happening.

There will be interesting work to be done in investigating seatbelt systems that have lower tension retractor systems as this will likely reduce the tension seen at the BTS buckle. In this regard, features that utilize signal magnitude will need to be explored more deply as they could potentially become unreliable. Similarly, the inspiration peak tension is lower in shorter participants than taller ones. It will be useful to perform a more in depth analysis of how the BTS recordings are affected by people of different stature as well as the amount of clothing being worn.

Mentioned in previous works, but not put into practice in this study is using multiple chest bands instead of one to measure the reference signal. Having one band placed around the chest and another at the mid to lower abdomen could increase the overall reliability of the reference measurement, allowing for a more concrete analysis of the performance of the BTS. If only one band is to be used, the optimal position for a single band is around the upper abdomen just below the sternum.

Not investigated much in this project, but what could be useful, is the fusion of data from other sensors to aid the interpretation of events that can potentially mislead the classification. Such sensors can include gps, accelerometers, steering wheel angle etc.

As has been seen with the inspiration and expiration time calculation, there exists room to better refine the process and concretely define where a breath starts and stops, as opposed to just looking for a minimum or maximium point. Additionally, there is room to improve the IE filtering to enable to smoothing of the respiration signal, while keeping the flatter portion of the expiration cycle.

More time and work is needed to improve the classification system. There is room to use more robust and in depth feature selection methods together with hyperparameter tuning. One advantage could be to use Python for the machine learning stage as there is more flexibility to customize the whole procedure. Participant level cross validation will be easier to integrate into feature selection and hyperparameter optimization in Python. Using reference measurement derived features to train the classifier is one way that could be used to check the consistency of the results. Currently, only features extracted from BTS data are used for training the classifiers.

Bibliography

[1] E. Commission, *2019 road safety statistics: What is behind the figures?* [Online]. Available: https://ec.europa.eu/commission/presscorner/detail/en/qanda_20_1004.

[2] National Highway Traffic Safety Administration, "Overview of motor vehicle crashes in 2019," no. December, 2020. [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813060.

[3] E. Parliament, "Regulation (EU) 2019/2144," vol. 2019, 2019.

[4] National Highway Traffic Safety Administration, "NHTSA Drowsy Driving Research and Program Plan," pp. 1–8, 2016. [Online]. Available: http://www.nhtsa.gov/staticfiles/nti/pdf/DrowsyDriving_StrategicPlan_030316.pdf.

[5] L. Novelli, R. Ferri, and O. Bruni, "Sleep classification according to AASM and Rechtschaffen and Kales: Effects on sleep scoring parameters of children and adolescents," *Journal of Sleep Research*, vol. 19, no. 1 PART. 2, pp. 238–247, 2010, ISSN: 09621105. DOI: 10.1111/j.1365-2869.2009.00785.x.

[6] A. Kay and P. Wilkin, "Respiratory Instability During Sleep Onset," *Sleep (Rochester)*, no. model 1400, 1992.

[7] S. Rios-Aguilar, J. L. M. Merino, A. Millán Sánchez, and Á. Sánchez Valdivieso, "Variation of the Heartbeat and Activity as an Indicator of Drowsiness at the Wheel Using a Smartwatch," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 3, no. 3, p. 96, 2015, ISSN: 1989-1660. DOI: 10.9781/ijimai.2015.3313.

[8] J. Vicente, P. Laguna, A. Bartra, and R. Bailón, "Detection of driver's drowsiness by means of HRV analysis," *Computing in Cardiology*, vol. 38, pp. 89–92, 2011, ISSN: 23258861.

[9] W. H. Press and S. A. Teukolsky, "Savitzky-Golay Smoothing Filters," *Computers in Physics*, vol. 4, no. 6, p. 669, 1990, ISSN: 08941866. DOI: 10.1063/1.4822961.

[10] D. W. Jung, S. H. Hwang, Y. J. Lee, D. U. Jeong, and K. S. Park, "Apnea-Hypopnea Index Prediction Using Electrocardiogram Acquired during the Sleep-Onset Period," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 2, pp. 295–301, 2017, ISSN: 15582531. DOI: 10.1109/TBME.2016.2554138.

[11] U. Wiklund, B. O. Olofsson, K. Franklin, H. Blom, P. Bjerle, and U. Niklasson, "Autonomic cardiovascular regulation in patients with obstructive sleep apnoea: A study based on spectral analysis of heart rate variability," *Clini-*

cal Physiology, vol. 20, no. 3, pp. 234–241, May 2000, ISSN: 01445979. DOI: 10.1046/j.1365-2281.2000.00251.x.

[12] T. Akerstedt and M. Gillberg, "Subjective and objective sleepiness in the active individual.," eng, *The International journal of neuroscience*, vol. 52, no. 1-2, pp. 29–37, May 1990, ISSN: 0020-7454 (Print). DOI: 10.3109/00207459008994241.

[13] W. W. Wierwille and L. A. Ellsworth, "Evaluation of driver drowsiness by trained raters," *Accident Analysis & Prevention*, vol. 26, no. 5, pp. 571–581, 1994, ISSN: 0001-4575. DOI: https://doi.org/10.1016/0001-4575(94)90019-1.

[14] L. A. Reyner and J. A. Horne, "Falling asleep whilst driving: Are drivers aware of prior sleepiness?" *International Journal of Legal Medicine*, vol. 111, no. 3, pp. 120–123, 1998, ISSN: 09379827. DOI: 10.1007/s004140050131.

[15] G. MAYCOCK, "Sleepiness and driving: The experience of uk car drivers," *Journal of Sleep Research*, vol. 5, pp. 229–231, 4 Dec. 1996, ISSN: 0962-1105. DOI: 10.1111/j.1365-2869.1996.00229.x.

[16] A. Williamson, R. Friswell, J. Olivier, and R. Grzebieta, "Are drivers aware of sleepiness and increasing crash risk while driving?" *Accident Analysis and Prevention*, vol. 70, pp. 225–234, Sep. 2014, ISSN: 00014575. DOI: 10.1016/j.aap.2014.04.007.

[17] R. J. Van Loon, R. F. Brouwer, and M. H. Martens, "Drowsy drivers' underperformance in lateral control: How much is too much? Using an integrated measure of lateral control to quantify safe lateral driving," *Accident Analysis and Prevention*, vol. 84, pp. 134–143, Nov. 2015, ISSN: 00014575. DOI: 10.1016/j.aap.2015.08.012.

[18] S. Leonhardt, L. Leicht, and D. Teichmann, "Unobtrusive vital sign monitoring in automotive environments—A review," *Sensors (Switzerland)*, vol. 18, no. 9, pp. 1–38, 2018, ISSN: 14248220. DOI: 10.3390/s18093080.

[19] K. E. Barrett, S. M. Barman, S. Boitano, and H. Brooks, *Ganong's Review of Medical Physiology*, 24th. 2012, p. 619, ISBN: 978-0071780032.

[20] E. Sembroski, D. Sanghavi, and A. Bhardwaj, *Inverse Ratio Ventilation*. Treasure Island: StatPearls, 2020.

[21] B. Warwick, N. Symons, X. Chen, and K. Xiong, "Detecting driver drowsiness using wireless wearables," *Proceedings - 2015 IEEE 12th International Conference on Mobile Ad Hoc and Sensor Systems, MASS 2015*, pp. 585–588, 2015. DOI: 10.1109/MASS.2015.22.

[22] T. Igasaki, K. Nagasawa, N. Murayama, and Z. Hu, "Drowsiness estimation under driving environment by heart rate variability and/or breathing rate variability with logistic regression analysis," *Proceedings - 2015 8th International Conference on BioMedical Engineering and Informatics, BMEI 2015*, no. Bmei, pp. 189–193, 2016. DOI: 10.1109/BMEI.2015.7401498.

[23] T. Igasaki, K. Nagasawa, I. A. Akbar, and N. Kubo, "Sleepiness classification by thoracic respiration using support vector machine," *BMEiCON 2016 - 9th Biomedical Engineering International Conference*, pp. 6–10, 2017. DOI: 10.1109/BMEiCON.2016.7859630.

[24] 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," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 6055–6058, 2011, ISSN: 1557170X. DOI: 10.1109/IEMBS.2011.6091496.

[25] N. Rodríguez-Ibáñez, M. A. García-González, M. Fernández-Chimeno, H. De Rosario, and J. Ramos-Castro, "Synchrosqueezing Index for Detecting Drowsiness Based on the Respiratory Effort Signal," *IFMBE Proceedings 41 XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013*, IFMBE Proceedings, vol. 41, L. M. Roa Romero, Ed., pp. 965–968, 2014. DOI: 10.1007/978-3-319-00846-2.

[26] B. G. Lee, B. L. Lee, and W. Y. Chung, "Mobile healthcare for automatic driving sleep-onset detection using wavelet-based EEG and respiration signals," *Sensors (Switzerland)*, vol. 14, no. 10, pp. 17 915–17 936, 2014, ISSN: 14248220. DOI: 10.3390/s141017915.

[27] S. E. H. Kiashari, A. Nahvi, H. Bakhoda, A. Homayounfard, and M. Tashakori, "Evaluation of driver drowsiness using respiration analysis by thermal imaging on a driving simulator," *Multimedia Tools and Applications*, vol. 79, no. 25-26, pp. 17 793–17 815, 2020, ISSN: 15737721. DOI: 10.1007/s11042-020-08696-x.

[28] N. J. Douglas, D. P. White, C. K. Pickett, J. V. Weil, and C. W. Zwillich, "Respiration during sleep in normal man," *Thorax*, vol. 37, no. 11, pp. 840–844, 1982, ISSN: 00406376. DOI: 10.1136/thx.37.11.840.

[29] B. M. Koeppen and B. A. Stanton, *Berne & Levy physiology*, 7th. Elsevier, 2017, pp. 447–450.

[30] V. Zakeri, A. Akhbardeh, N. Alamdari, R. Fazel-Rezai, M. Paukkunen, and K. Tavakolian, "Analyzing seismocardiogram cycles to identify the respiratory phases," *IEEE Transactions on Biomedical Engineering*, vol. PP, pp. 1–1, Oct. 2016. DOI: 10.1109/TBME.2016.2621037.

[31] J. Solaz, J. Laparra-Hernández, D. Bande, N. Rodríguez, S. Veleff, J. Gerpe, and E. Medina, "Drowsiness Detection Based on the Analysis of Breathing Rate Obtained from Real-time Image Recognition," *Transportation Research Procedia*, vol. 14, pp. 3867–3876, 2016, ISSN: 23521465. DOI: 10.1016/j.trpro.2016.05.472.

[32] P. Kang, M. Liao, M. R. Wester, J. S. Leeder, and R. E. Pearce, "Ballistocardiography – A Method Worth Revisiting," *Conf Proc IEEE Eng Med Biol Soc.*, pp. 4279–4282, 2011, ISSN: 1946-6242. DOI: 10.1109/IEMBS.2011.6091062.Ballistocardiography. arXiv: NIHMS150003.

[33] D. J. Marlin and C. M. Deaton, "15 - pulmonary function testing," in *Equine Respiratory Medicine and Surgery*, B. C. McGorum, P. M. Dixon, N. E. Robinson, and J. Schumacher, Eds., Edinburgh: W.B. Saunders, 2007, pp. 211–233, ISBN: 978-0-7020-2759-8. DOI: https://doi.org/10.1016/B978-0-7020-2759-8.50020-9.

[34] Y. Shan, S. Li, and T. Chen, "Respiratory signal and human stress: non-contact detection of stress with a low-cost depth sensing camera," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 8, pp. 1825–1837, 2020, ISSN: 1868808X. DOI: 10.1007/s13042-020-01074-x.

[35] F. Guede-Fernández, M. Fernández-Chimeno, J. Ramos-Castro, and M. A. García-González, "Driver Drowsiness Detection Based on Respiratory Signal

Analysis," *IEEE Access*, vol. 7, pp. 81 826–81 838, 2019, ISSN: 21693536. DOI: 10.1109/ACCESS.2019.2924481.

[36] J. Boyle, N. Bidargaddi, A. Sarela, and M. Karunanithi, "Automatic detection of respiration rate from ambulatory single-lead ECG," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 6, pp. 890–896, 2009, ISSN: 10897771. DOI: 10.1109/TITB.2009.2031239.

[37] S. Tateno, X. Guan, R. Cao, and Z. Qu, "Development of Drowsiness Detection System Based on Respiration Changes Using Heart Rate Monitoring," *57th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, pp. 1664–1669, 2018. DOI: 10.23919/SICE.2018.8492599.

[38] A. R. Fekr, M. Janidarmian, K. Radecka, and Z. Zilic, *A medical cloud-based platform for respiration rate measurement and hierarchical classification of breath disorders*, 2014. DOI: 10.3390/s140611204.

[39] J. W. Yoon, Y. S. Noh, Y. S. Kwon, W. K. Kim, and H. R. Yoon, "Improvement of dynamic respiration monitoring through sensor fusion of accelerometer and gyro-sensor," *Journal of Electrical Engineering and Technology*, vol. 9, no. 1, pp. 334–343, 2014, ISSN: 19750102. DOI: 10.5370/JEET.2014.9.1.334.

[40] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995, ISSN: 1573-0565. DOI: 10.1023/A:1022627411411.

[41] R. Gandhi, *Support vector machine - introduction to machine learning algorithms*, Jul. 2018. [Online]. Available: https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47.

[42] A. Ajanki, *K nearest neighbor*, May 2007. [Online]. Available: https://commons.wikimedia.org/wiki/File:KnnClassification.svg.

[43] T. K. Ho, "Random decision forests," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 1, 1995, 278–282 vol.1. DOI: 10.1109/ICDAR.1995.598994.

[44] A. Koesdwiady, R. Abdelmoula, F. Karray, and M. Kamel, "Driver Inattention Detection System: A PSO-Based Multiview Classification Approach," *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, pp. 1624–1629, 2015. DOI: 10.1109/ITSC.2015.264.

[45] C. Ahlström, R. Zemblys, H. Jansson, C. Forsberg, J. Karlsson, and A. Anund, "Effects of partially automated driving on the development of driver sleepiness," *Accident Analysis & Prevention*, vol. 153, p. 106 058, 2021, ISSN: 0001-4575. DOI: https://doi.org/10.1016/j.aap.2021.106058.

[46] W. W. LaMorte, *Mann Whitney U Test (Wilcoxon Rank Sum Test)*, 2017. [Online]. Available: https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704%7B%5C_%7Dnonparametric/bs704%7B%5C_%7Dnonparametric4.html (visited on 04/26/2021).

[47] H. De Rosario, J. Solaz, P. Gameiro, and L. Costa, "Using smart materials to monitor physiological signals of driver's inattention," *3rd International Conference on Driver Distraction and Inattention*, no. 78, pp. 1–7, 2013.

A

Appendix 1

Table A.1: Comparison of studies that use breathing parameters to derive sleepiness

Study Name	Author	Method	Variables	Performance
Driver drowsiness thermal imaging	Kiashari et al. [27]	Respiration signal extraction using thermal imaging and drowsiness classification using SVM	Mean and SD of RR and inspiration-expiration time. (SST used for rate extraction)	Accuracy of 90%
Detecting Driver Drowsiness using Wireless wearables	Warwick et al. [21]	Use a wearable band sensor to measure HR and BR. Freq representation of BR sent into a neural net.	HR, HRV, BR decrease	..
Drowsiness Detection by Thoracic Effort	Rodriguez-Ibanez et al. [24]	Developed the TEDD index that indicates a drowsy driver based on its output being higher than a certain value	Respiratory rate variability (RR calculated with a threshold crossing method)	Sensitivity: 83% Specificity: 95% (These are for the most drowsy state)
Logistic Regression Analysis with BR and HR variability	Igasaki et al. [22]	ECG + chest band for HR + BR detection, using Japanese version of KSS for evaluation (KSS-J)	HRV, BRV, logistic regression of 4HRV + 4BRV + KSS-J data	HRV accuracy: 71% BRV accuracy: 72% Combined accuracy: 81%.

Sleepiness Classification by Thoracic Respiration Using SVM	Igasaki et al. [23]	Fed extracted parameters related to respiration into a SVM. Targets based on KSS-J	ID, ED, Max to min amplitude, ID-to-ED ratio and others (took mean, SD, var.. Of these)	Heavy sleepiness accuracy: 94%. Light or heavy accuracy: 96%. Differentiate between 3 classes of sleepiness accuracy 89%. Overall accuracy: 94.1%.
Mobile Healthcare for Automatic Driving Sleep-Onset Detection	Lee et al. [26]	Two-channel EEG and respiratory sensor (sampled at 1hz), then extraction of various EEG waves using wavelet	15 EEG features + BR regularity	Accuracy: 98.6%. Sensitivity: 99.1%. Specificity 99.5%. All when using at least 12 features.
Driver Drowsiness Detection Based on Respiratory Signal Analysis	Guede-Fernández et al. [35]	Use the TEDD algorithm on respiration data with extra filtering	BR variability	Specificity: 97%. Sensitivity: 90%.
Syn-chrosqueezing Index for Detecting Drowsiness Based on the Respiratory Effort Signal	Rodríguez-Ibáñez et al. [25]	Use syn-chrosqueezing (SST) on respiration data to extract BR.	BR mean and standard deviation.	Mean BR is lower when drowsy. SD of BR is higher when drowsy.

Using smart materials to monitor physiological signals of driver's inattention	De Rosario et al. [47]	Strategically positioned smart material to measure HR and BR.	..	No info on sleepiness classification yet, since paper focused on the biological signal acquisition.

B

Appendix 2

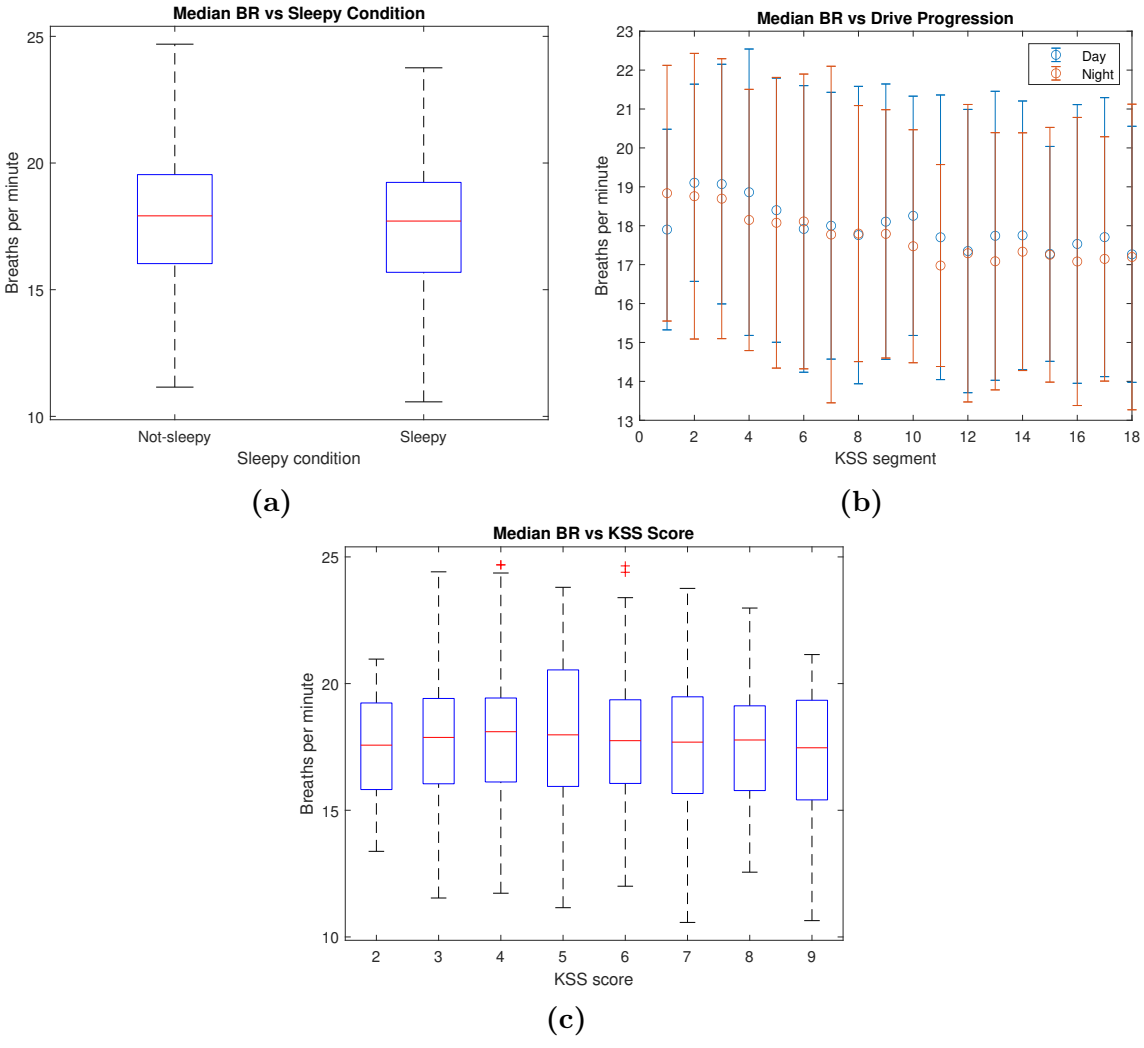


Figure B.1: The changes seen in median BR for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

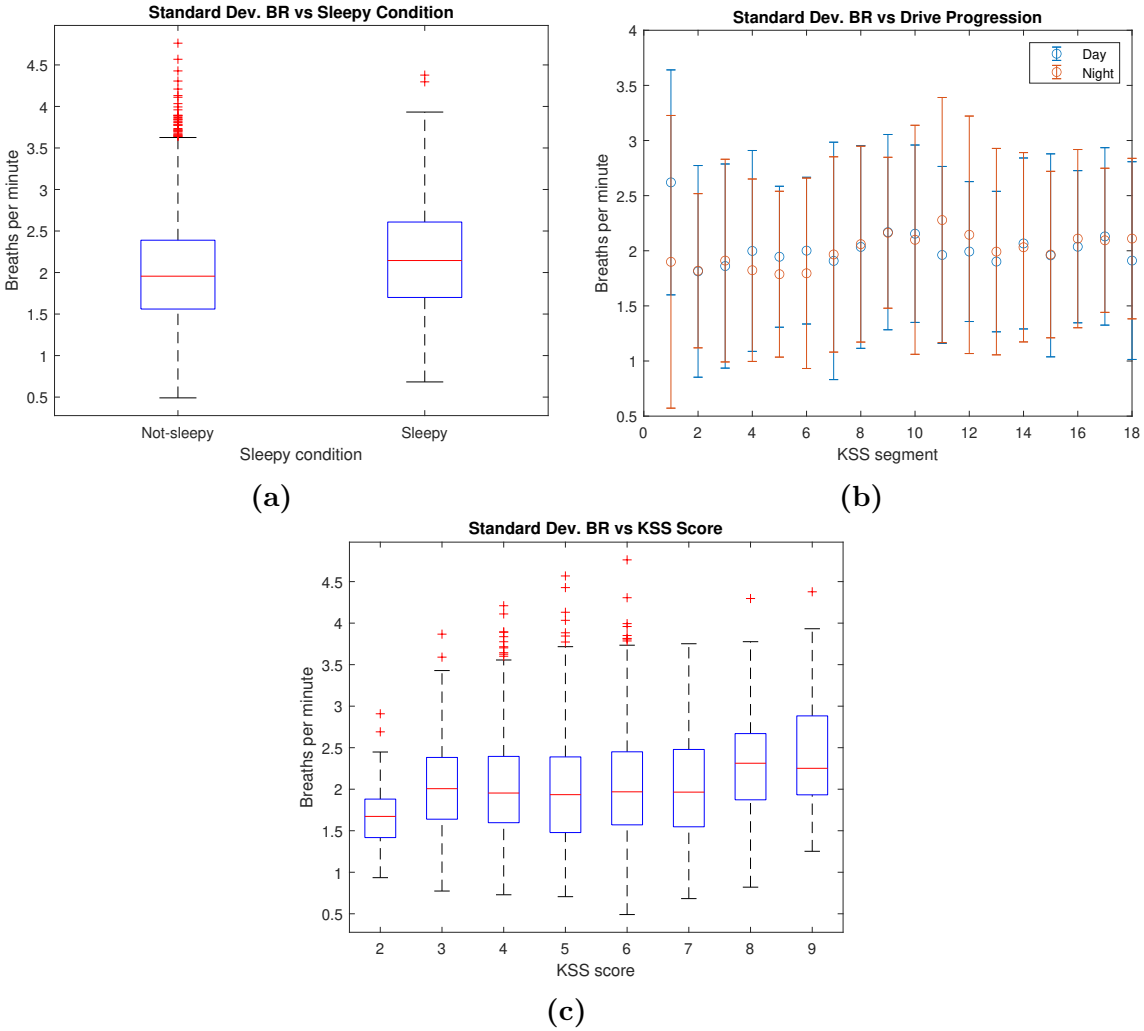


Figure B.2: The changes seen in the standard deviation of BR for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

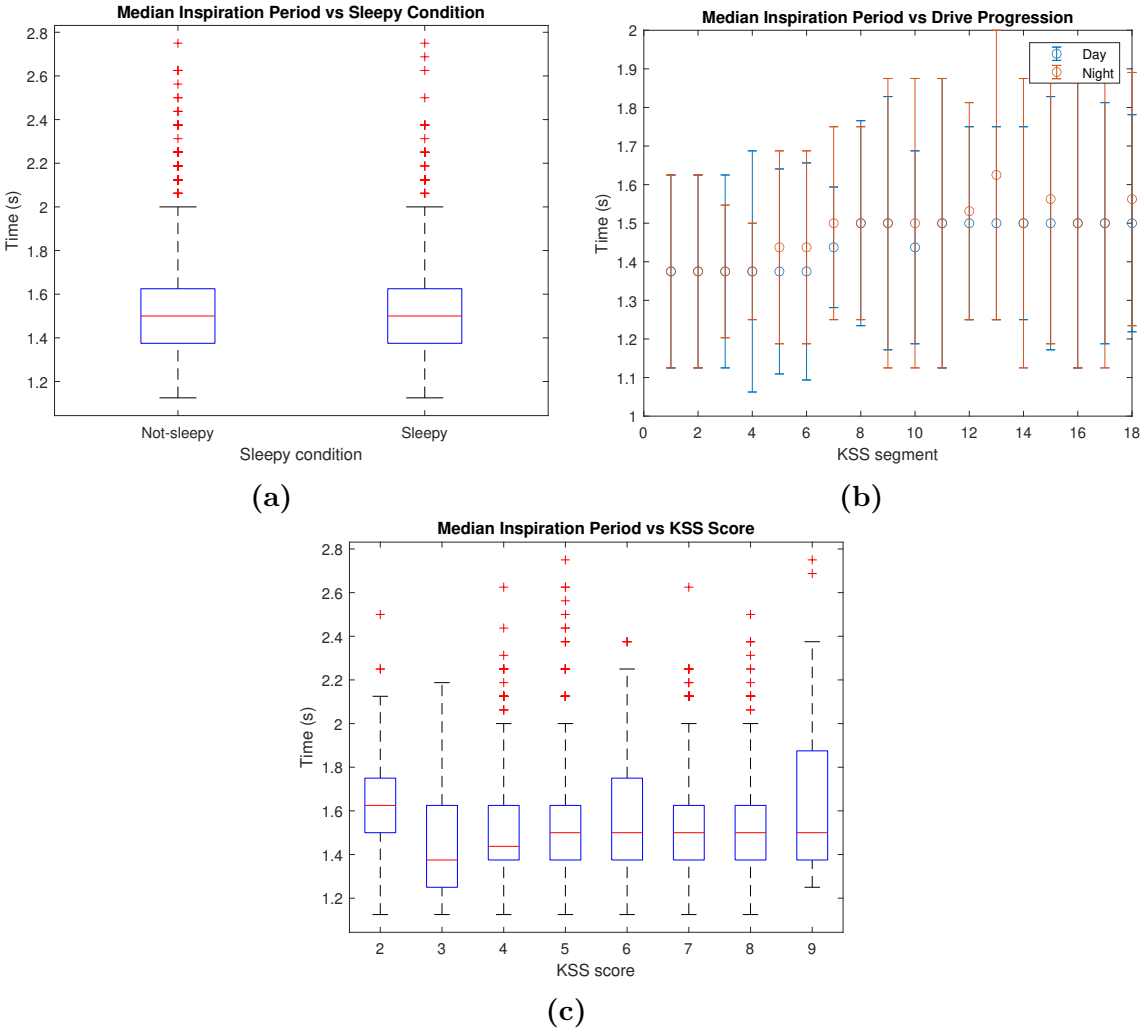


Figure B.3: The changes seen in median inspiration period for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

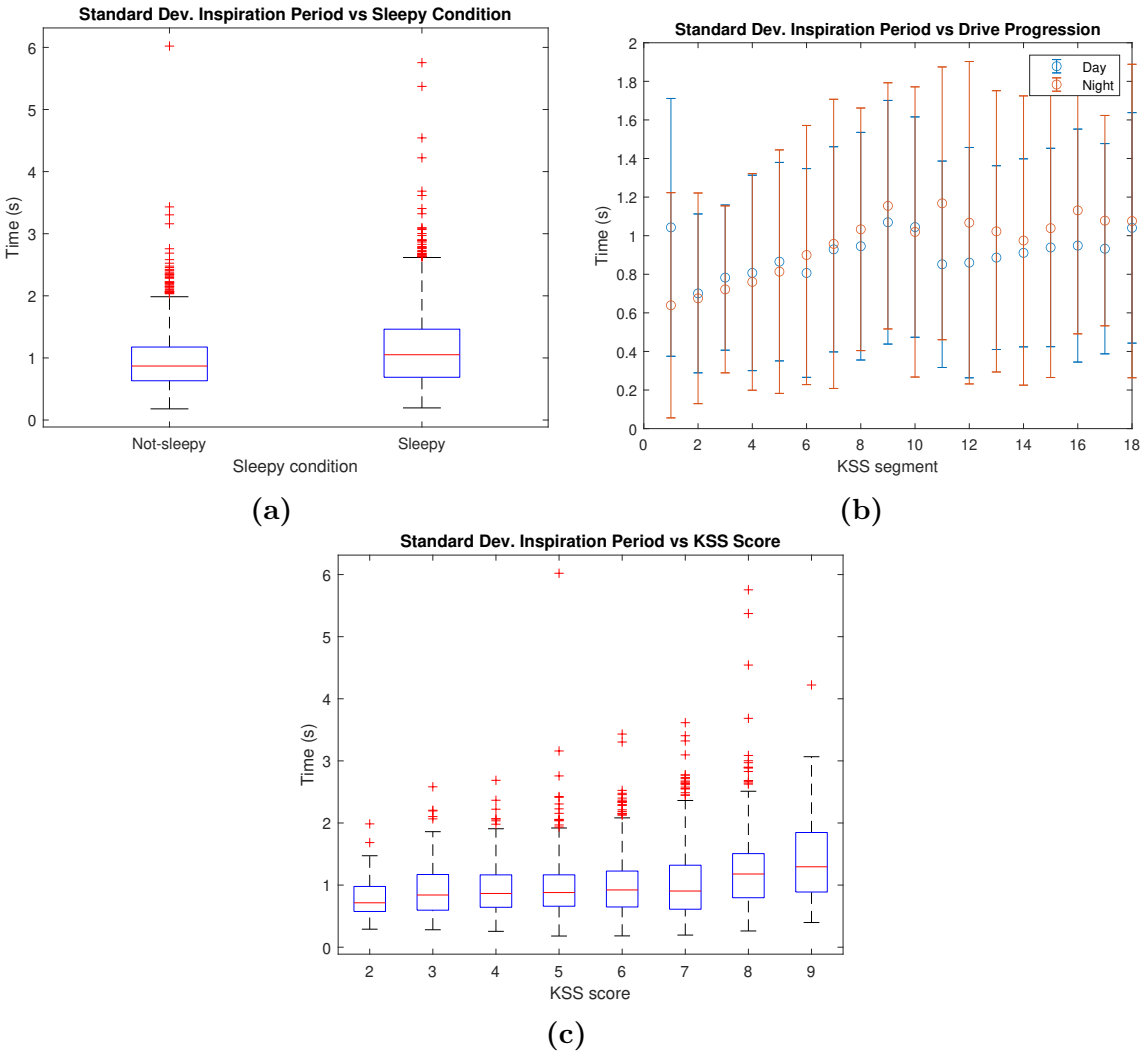


Figure B.4: The changes seen in the standard deviation of inspiration period for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

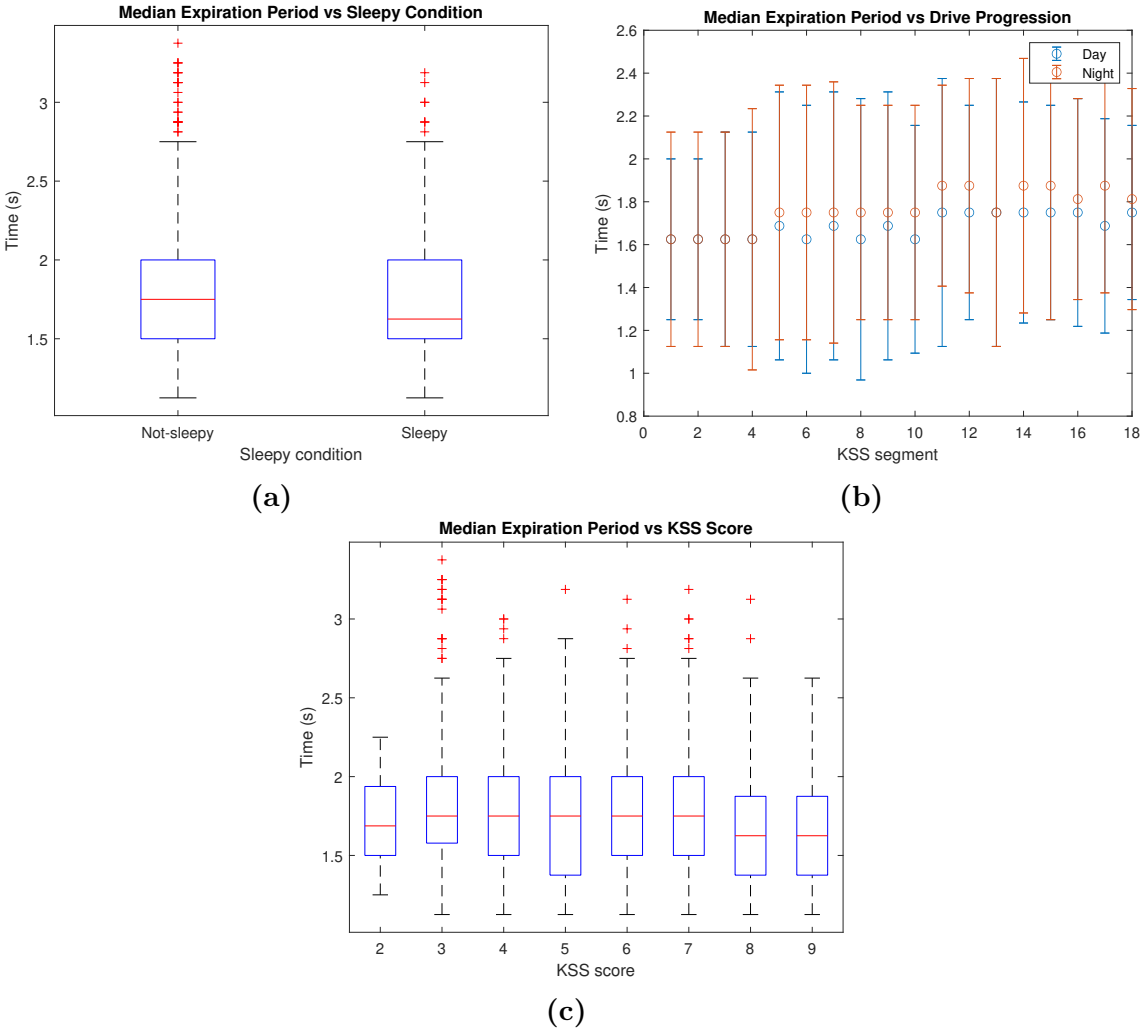


Figure B.5: The changes seen in median expiration period for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

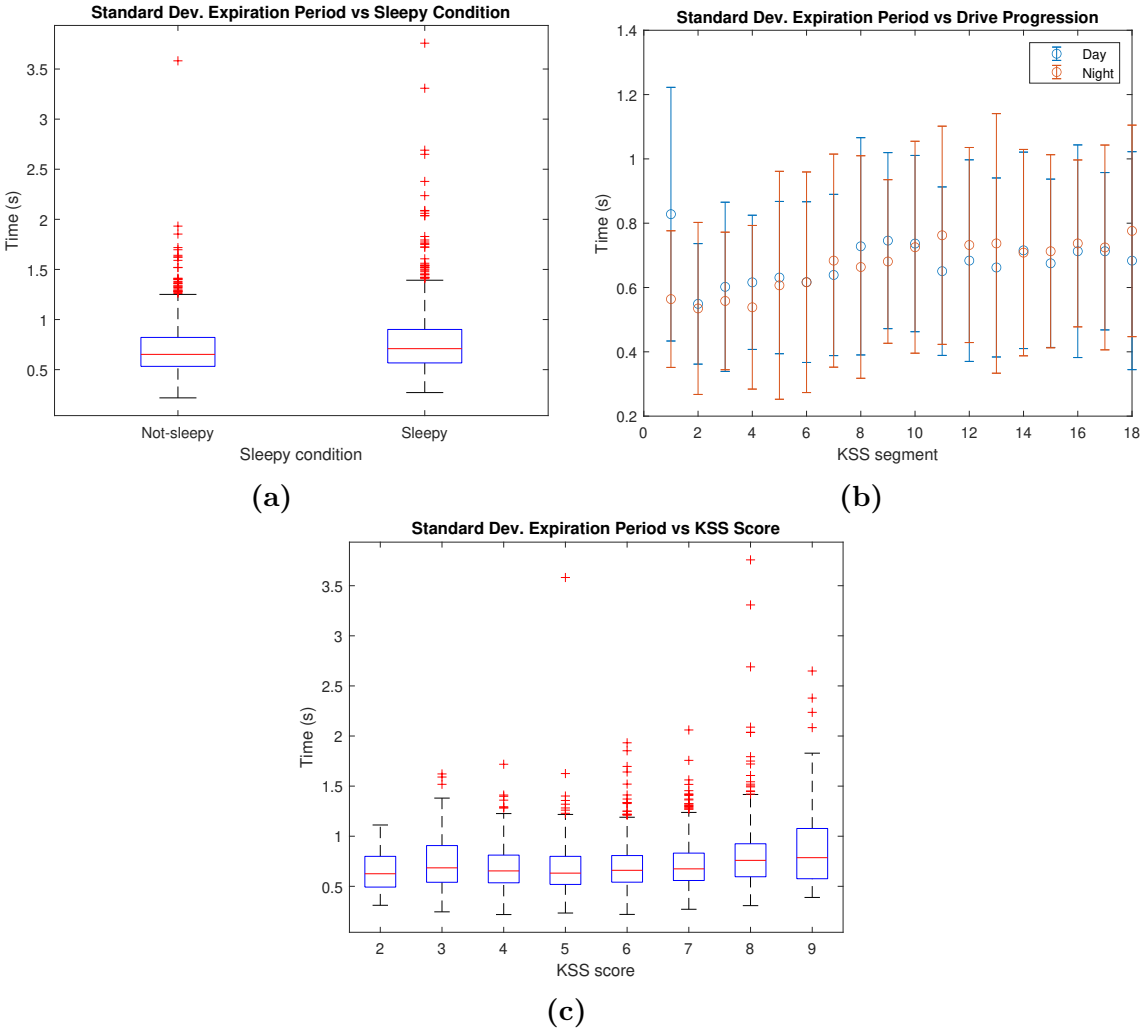


Figure B.6: The changes seen in the standard deviation of expiration period for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

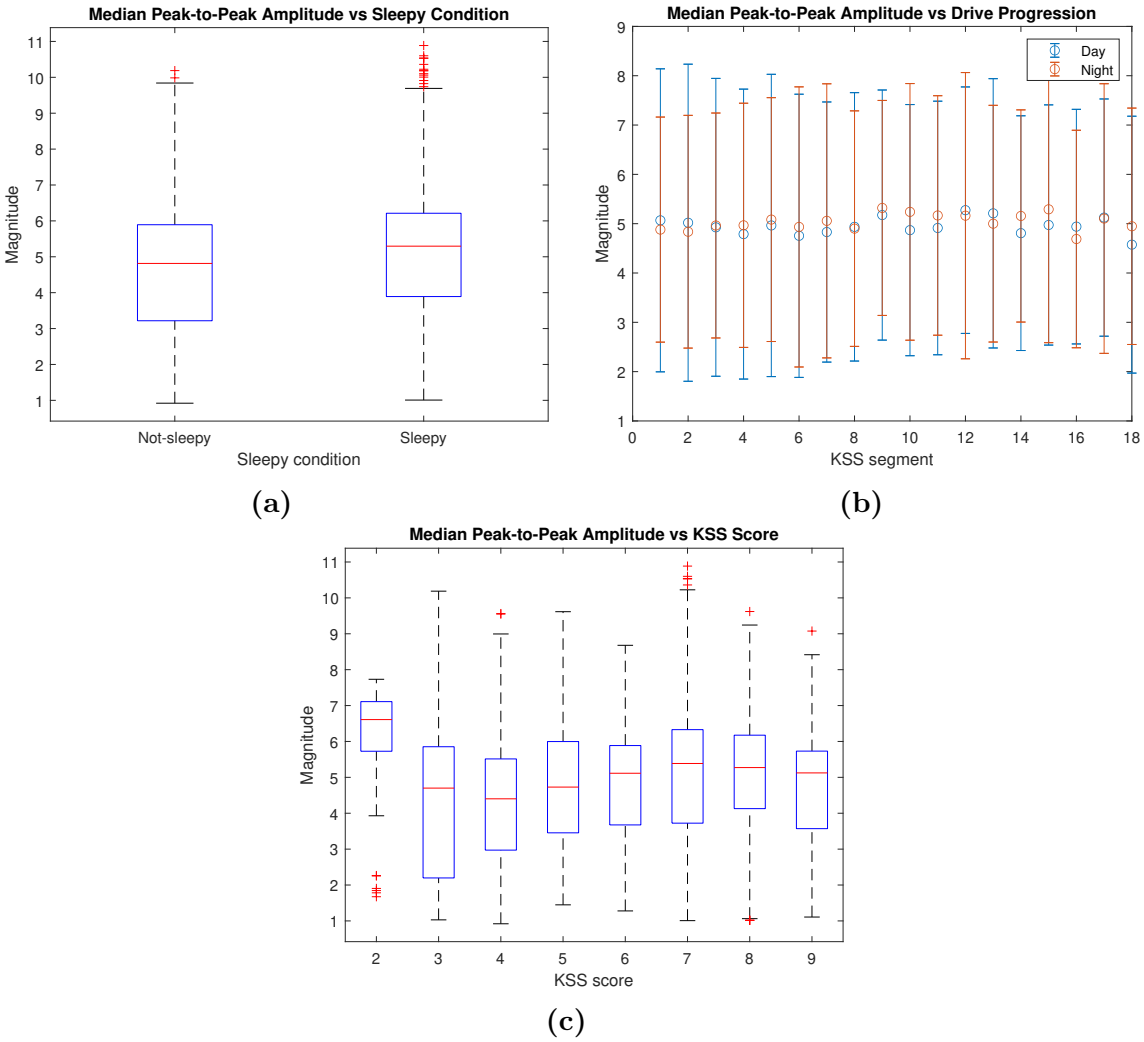


Figure B.7: The changes seen in median peak-to-peak amplitude for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

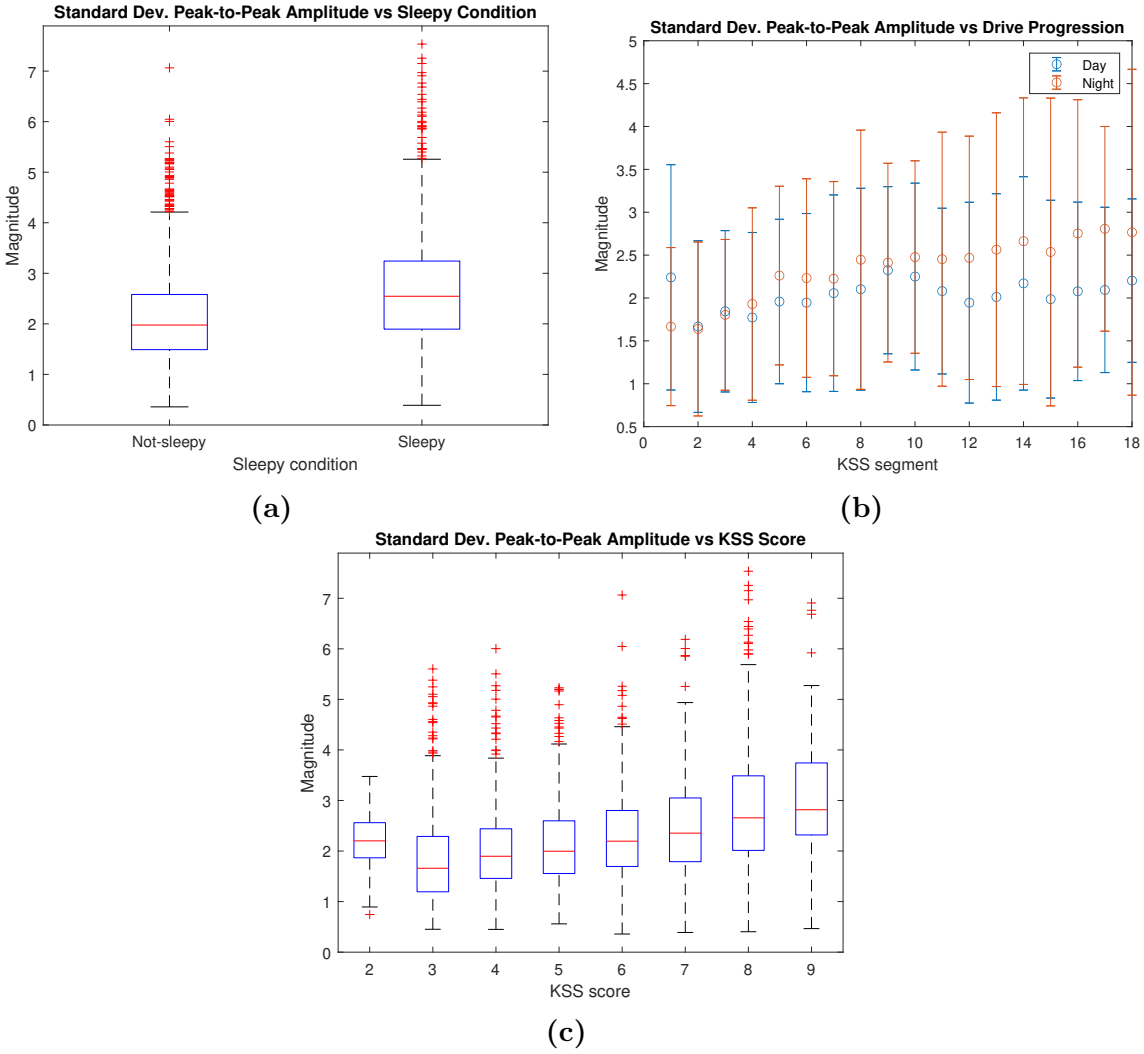


Figure B.8: The changes seen in the standard deviation of peak-to-peak amplitude for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

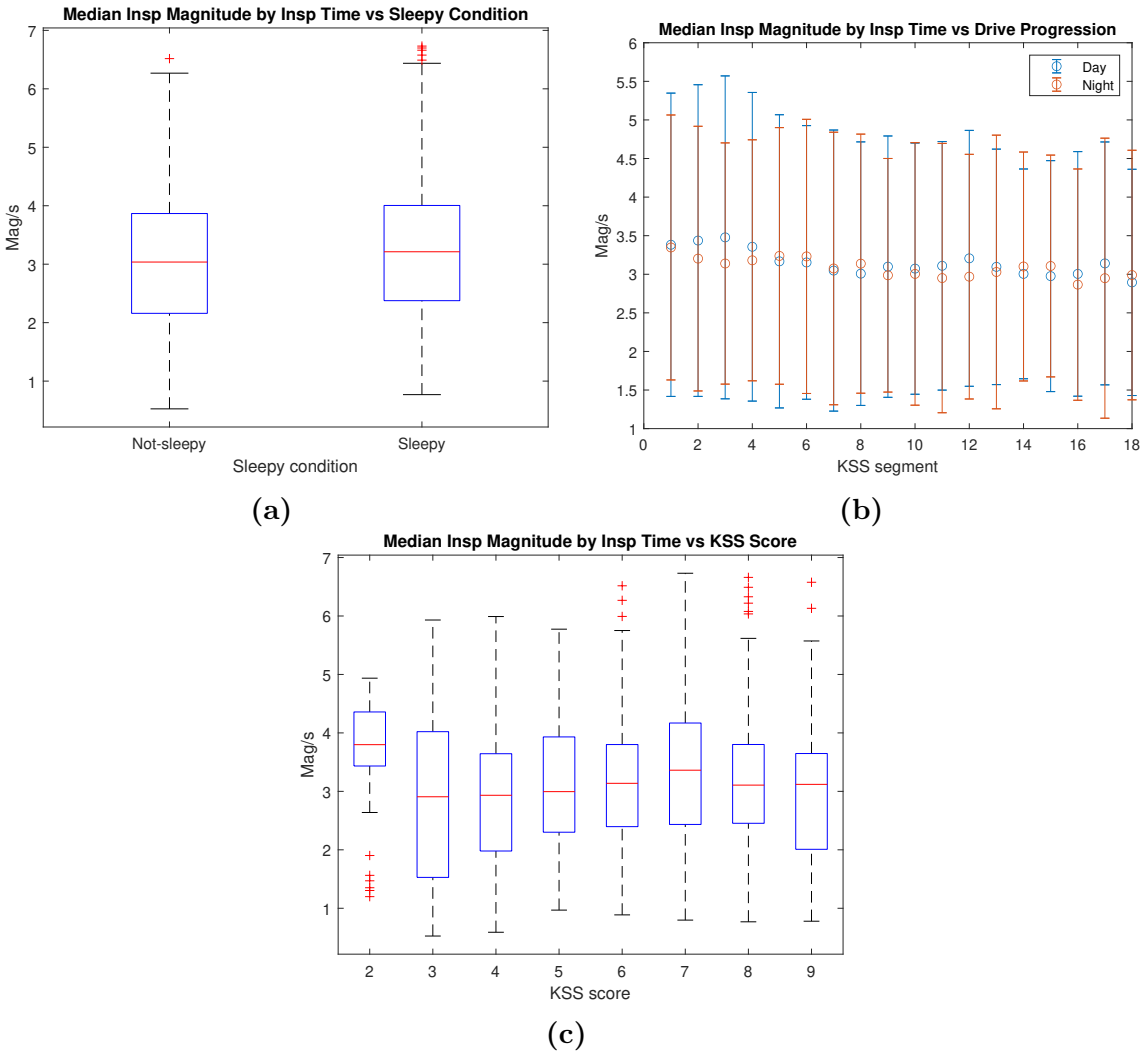


Figure B.9: The changes seen in the median inspiration amplitude to inspiration duration ratio for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

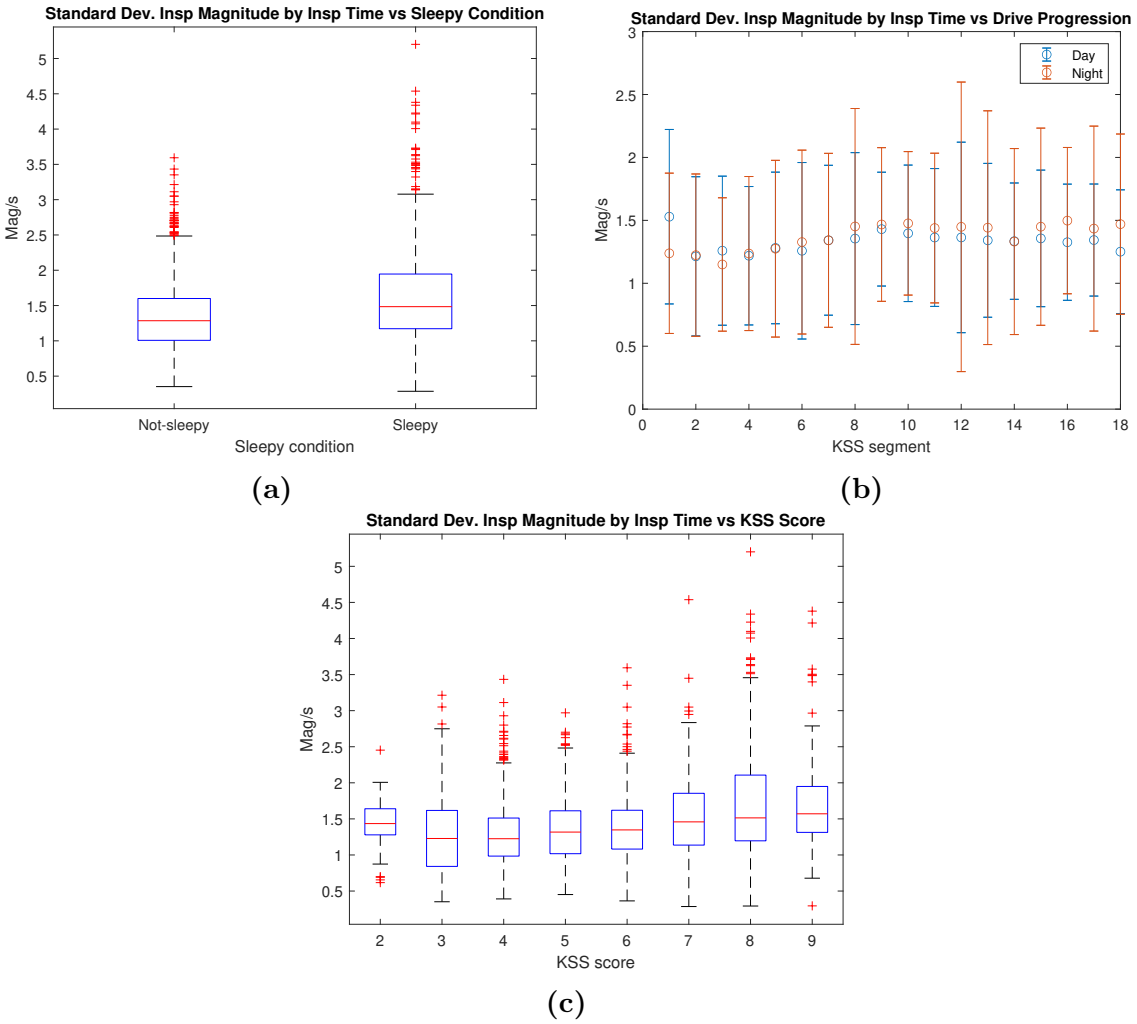


Figure B.10: The changes seen in the standard deviation of inspiration amplitude to inspiration duration ratio for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

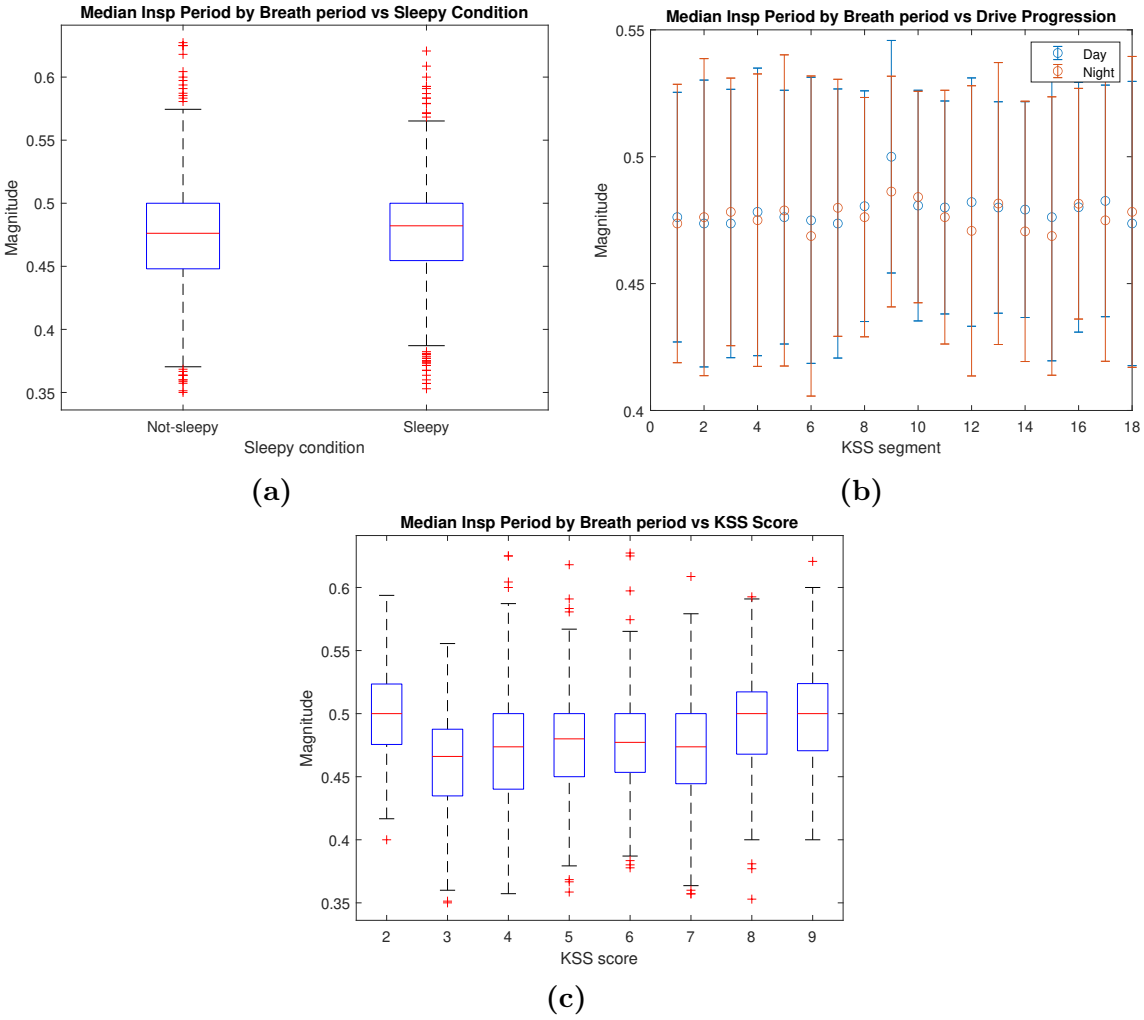


Figure B.11: The changes seen in median inspiration period to full breath period ratio for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

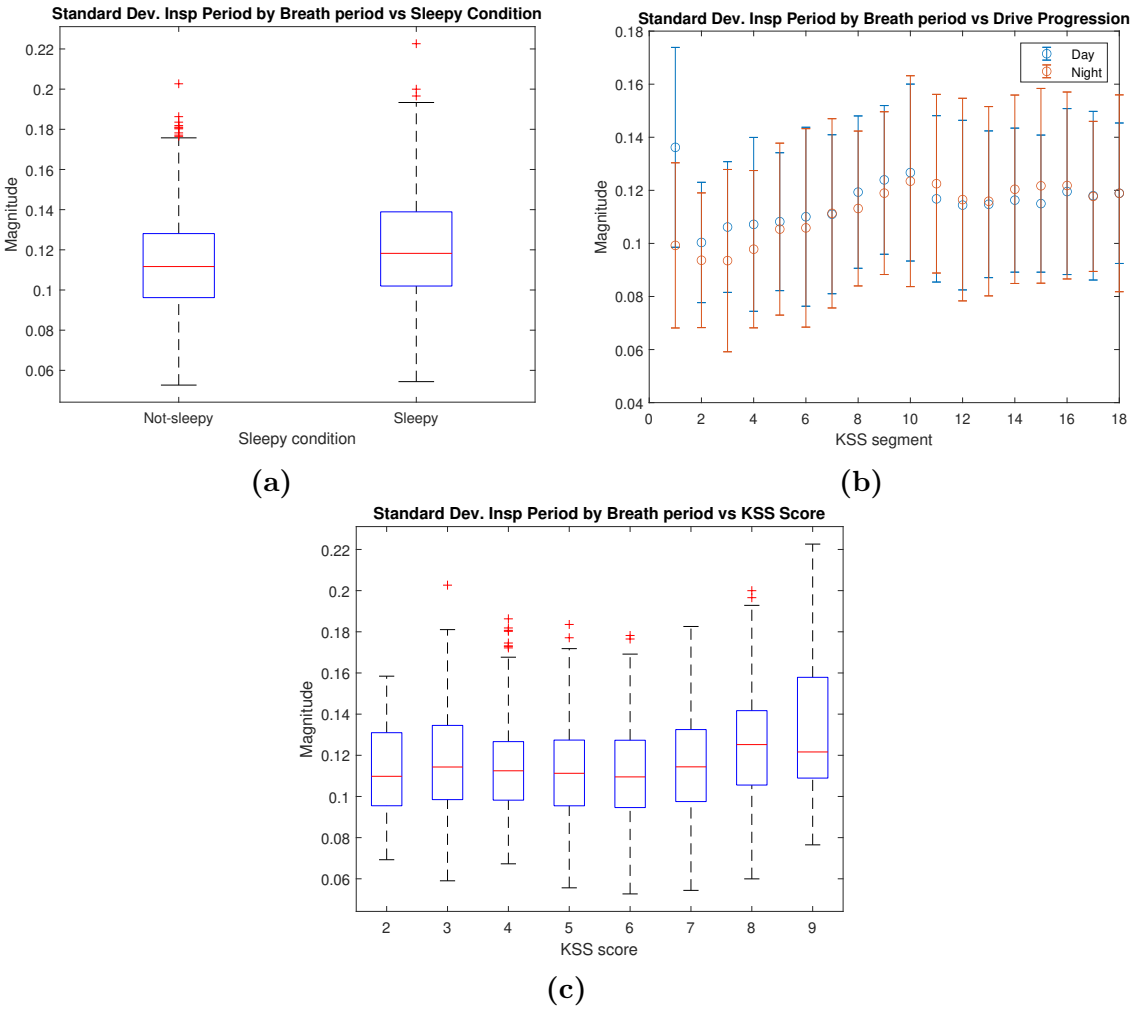


Figure B.12: The changes seen in the standard deviation of inspiration period to full breath period ratio for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

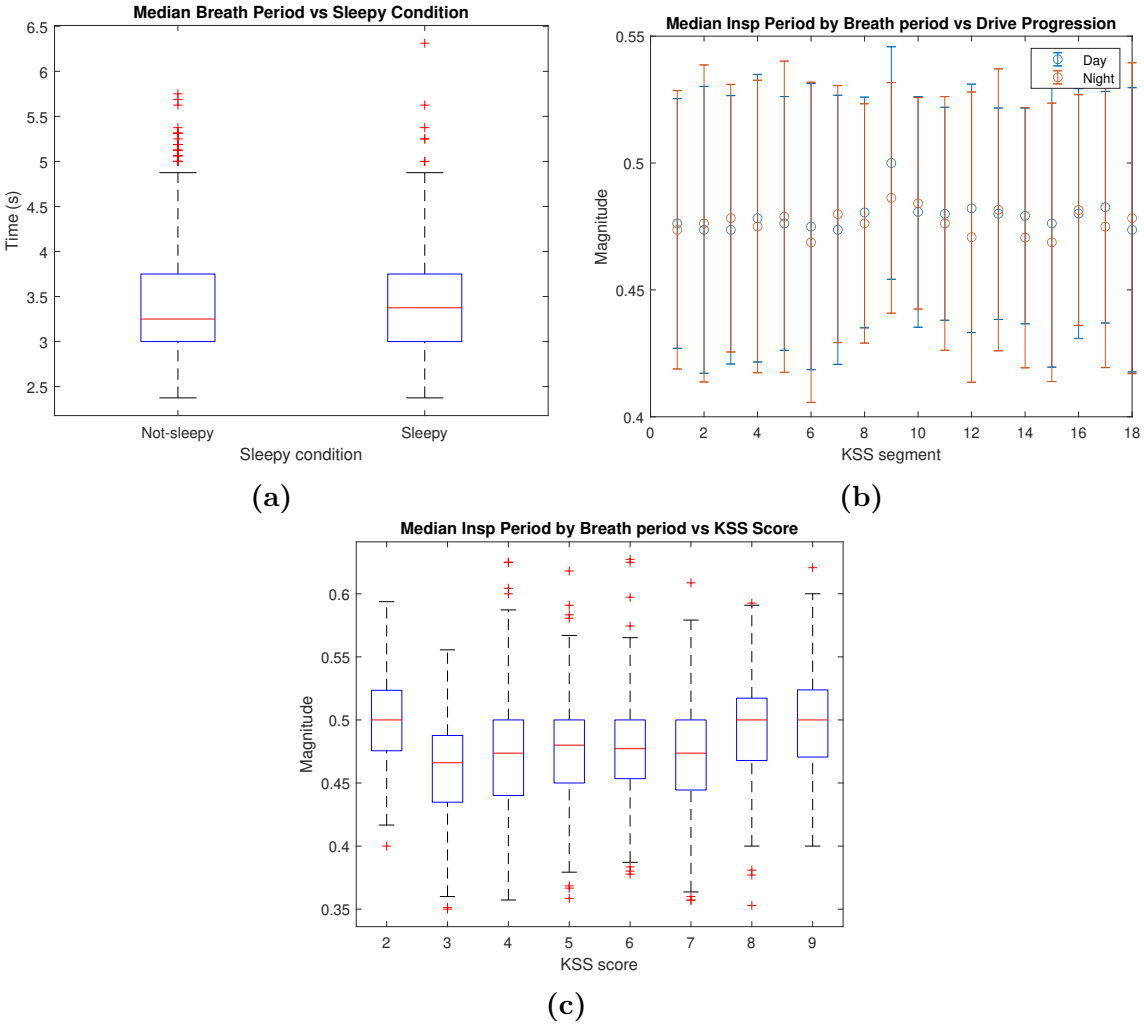


Figure B.13: The changes seen in median breath period for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

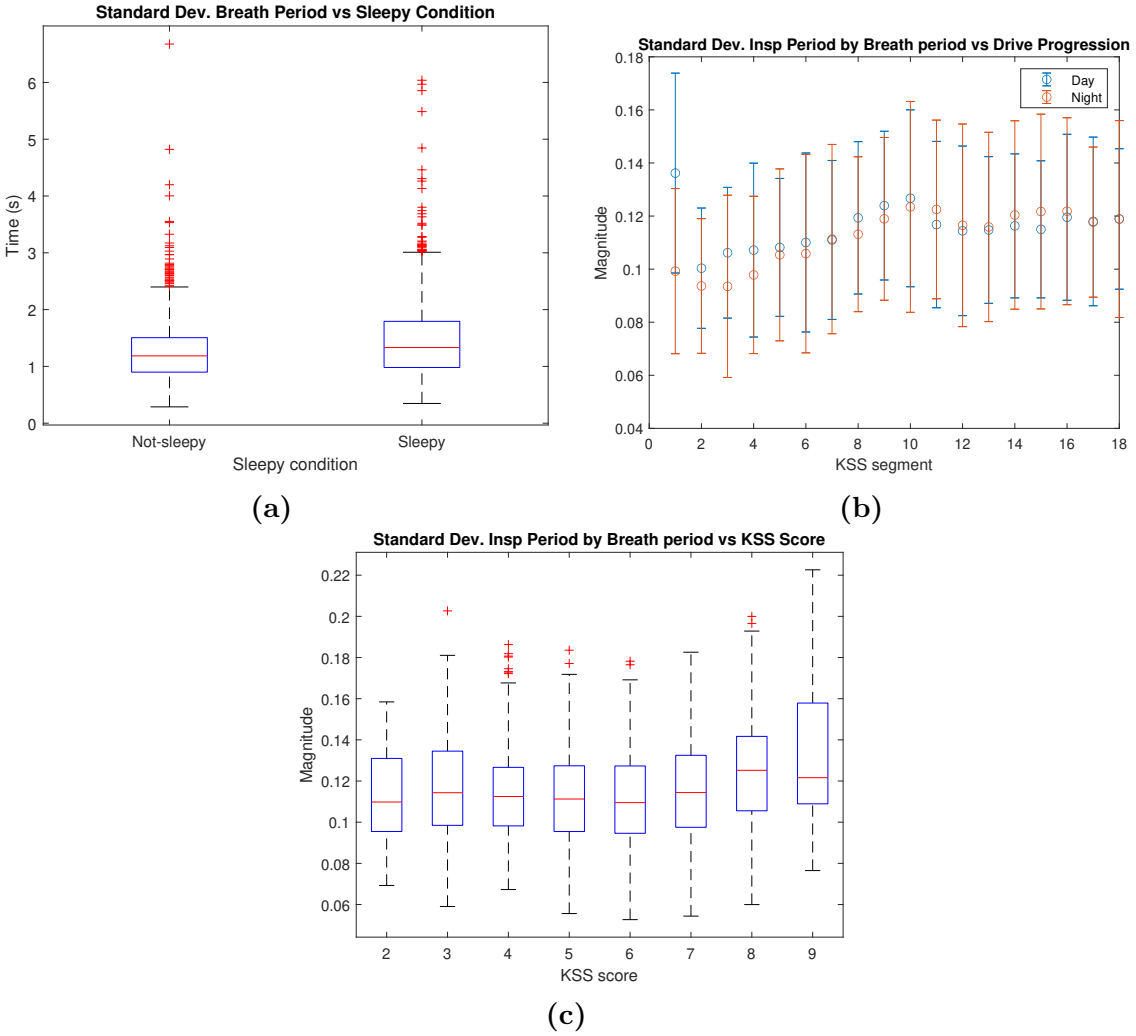


Figure B.14: The changes seen in the standard deviation of breath period for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

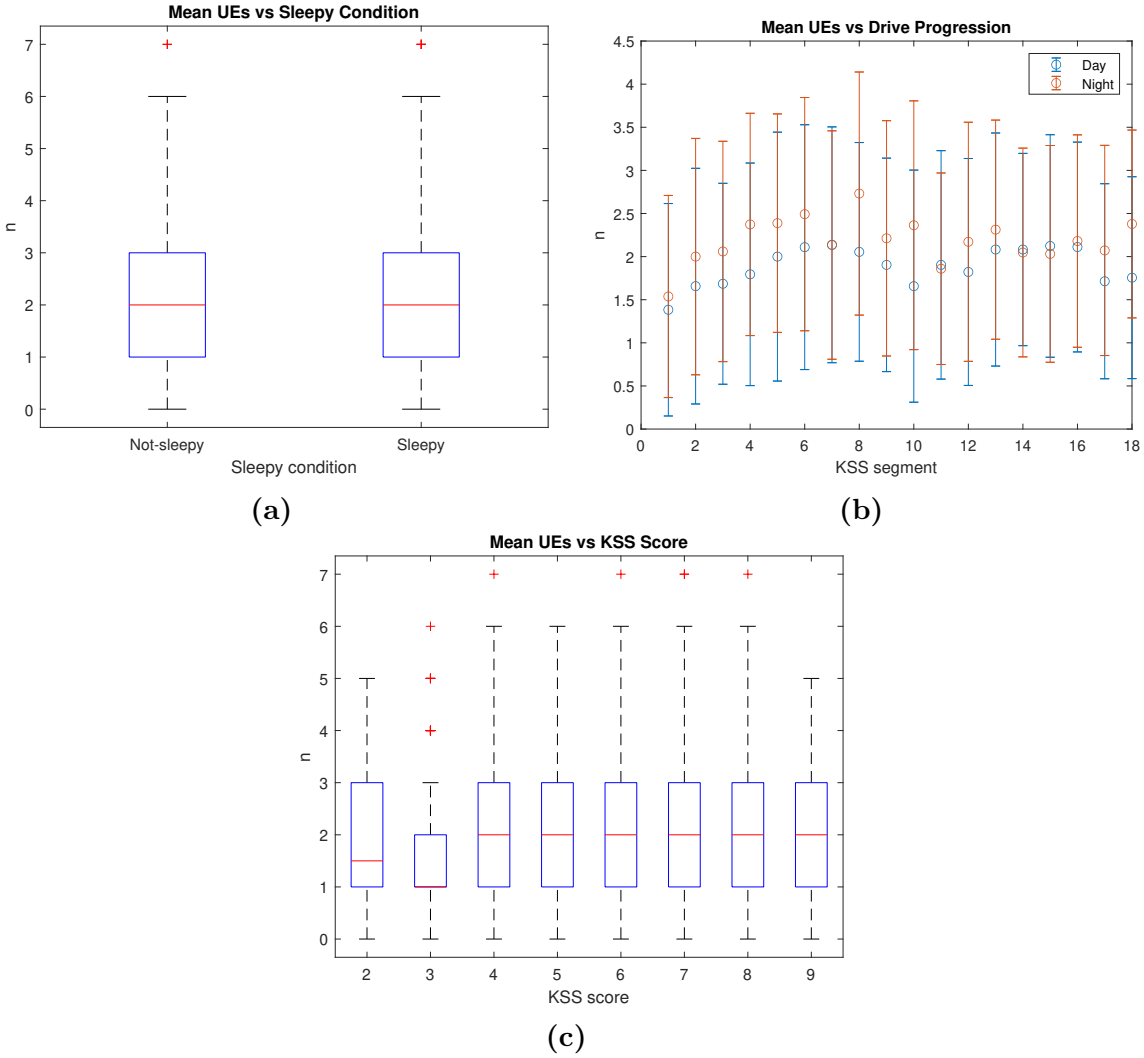


Figure B.15: The changes seen in mean unusual respiration count for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

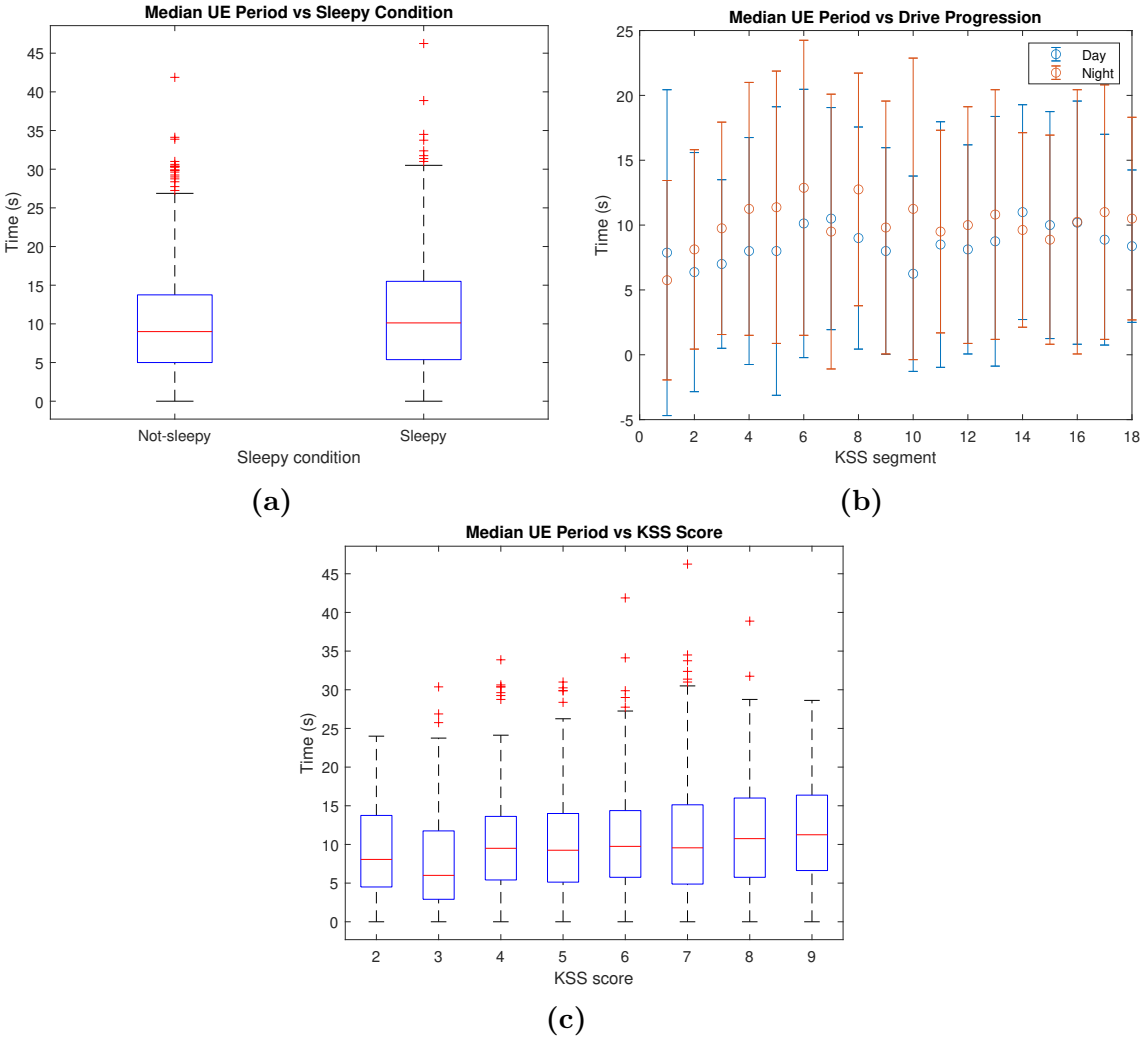


Figure B.16: The changes seen in median unusual respiration period for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

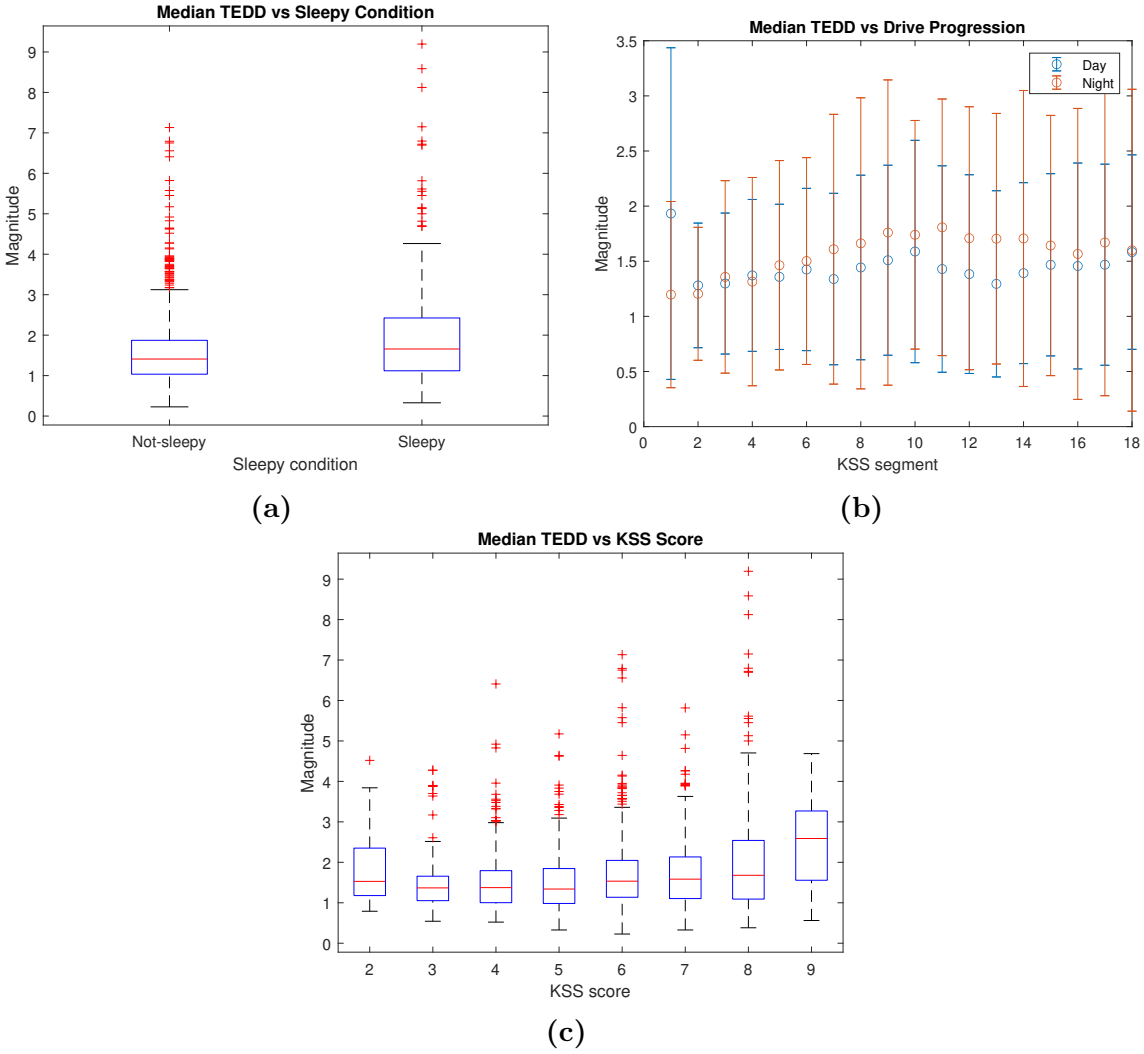


Figure B.17: The changes seen in median TEDD value for sleepy vs not sleepy segments (a), day/night condition as the drives progress (b) and for each bin of KSS score (c).

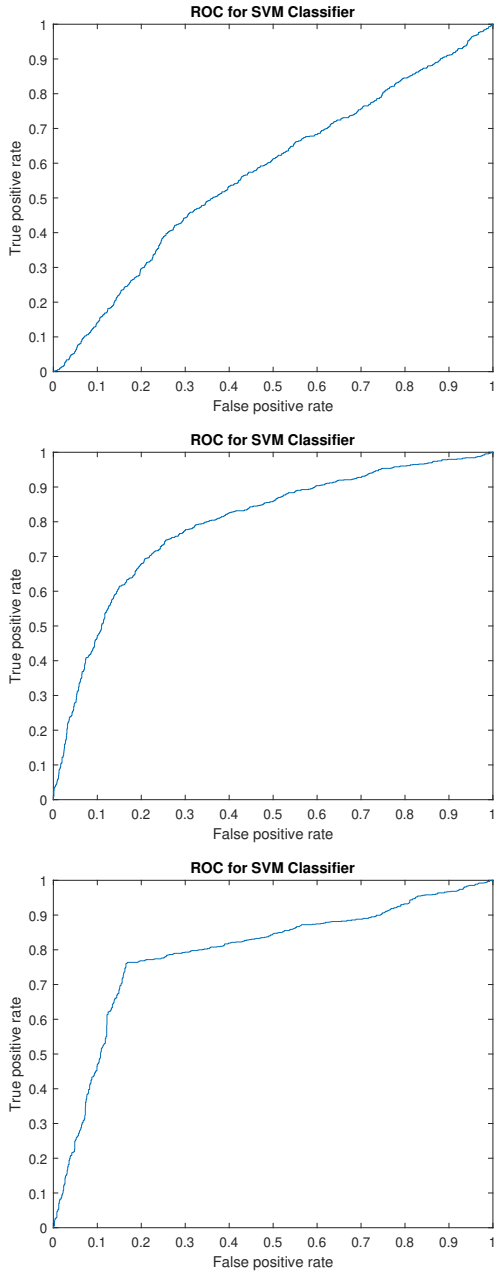


Figure B.18: Region under Curve for SVM classifier. Top: selected parameters, middle: selected parameters + automation and day/night condition, bottom: only automation and day/night condition.

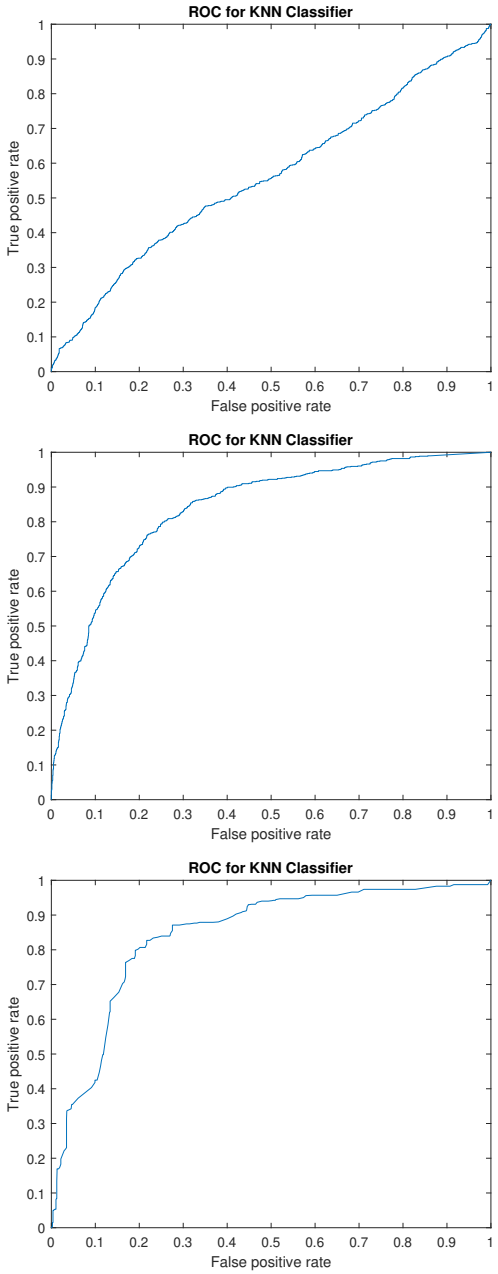


Figure B.19: Region under Curve for KNN classifier. Top: selected parameters, middle: selected parameters + automation and day/night condition, bottom: only automation and day/night condition.

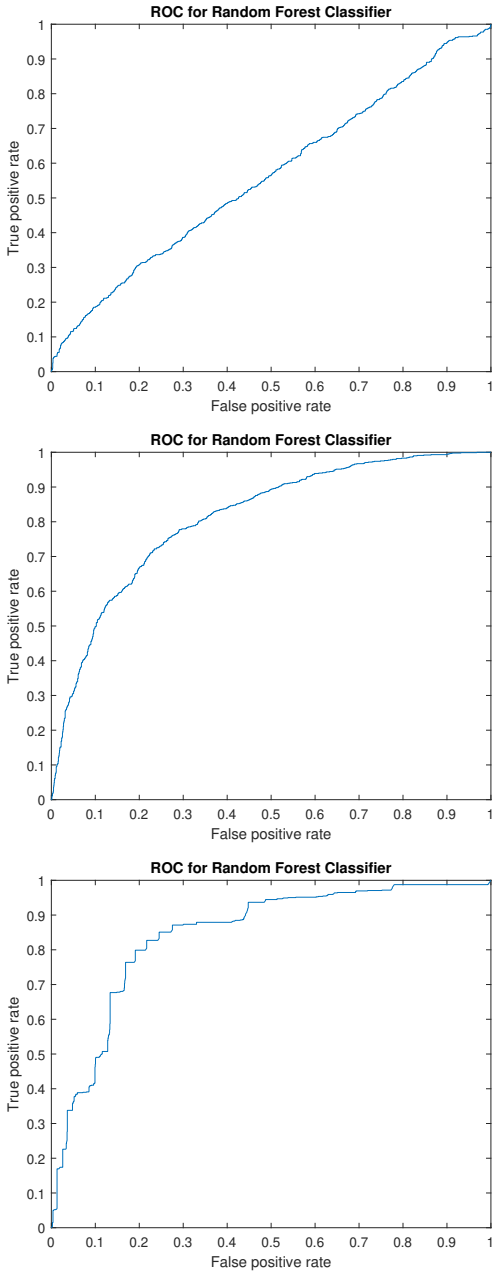


Figure B.20: Region under Curve for Random Forest classifier. Top: selected parameters, middle: selected parameters + automation and day/night condition, bottom: only automation and day/night condition.

DEPARTMENT OF ELECTRICAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY