Machine Learning in Driver Drowsiness Detection: A Focus on HRV, EDA, and Eye Tracking

by

Jose A. Alguindigue Ruiz

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2023

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Drowsy driving continues to be a significant cause of road traffic accidents, necessitating the development of robust drowsiness detection systems. This research enhances our understanding of driver drowsiness by analyzing physiological indicators – heart rate variability (HRV), the percentage of eyelid closure over the pupil over time (PERCLOS), blink rate, blink percentage, and electrodermal activity (EDA) signals. Data was collected from 40 participants in a controlled scenario, with half of the group driving in a non-monotonous scenario and the other half in a monotonous scenario. Participant fatigue was assessed twice using the Fatigue Assessment Scale (FAS).

The research developed three machine learning models: HRV-Based Model, EDA-Based Model, and Eye-Based Model, achieving accuracy rates of 98.28%, 96.32%, and 90% respectively. These models were trained on the aforementioned physiological data, and their effectiveness was evaluated against a range of advanced machine learning models including GRU, Transformers, Mogrifier LSTM, Momentum LSTM, Difference Target Propagation, and Decoupled Neural Interfaces Using Synthetic Gradients.

The HRV-Based Model and EDA-Based Model demonstrated robust performance in classifying driver drowsiness. However, the Eye-Based Model had some difficulty accurately identifying instances of drowsiness, likely due to the imbalanced dataset and underrepresentation of certain fatigue states. The study duration, which was confined to 45 minutes, could have contributed to this imbalance, suggesting that longer data collection periods might yield more balanced datasets.

The average fatigue scores obtained from the FAS before and after the experiment showed a relatively consistent level of reported fatigue among participants, highlighting the potential impact of external factors on fatigue levels.

By integrating the outcomes of these individual models, each demonstrating strong performance, this research establishes a comprehensive and robust drowsiness detection system. The HRV-Based Model displayed remarkable accuracy, while the EDA-Based Model and the Eye-Based Model contributed valuable insights despite some limitations. The research highlights the necessity of further optimization, including more balanced data collection and investigation of individual and external factors impacting drowsiness. Despite the challenges, this work significantly contributes to the ongoing efforts to improve road safety by laying the foundation for effective real-time drowsiness detection systems and intervention methods.

Acknowledgements

In the completion of this thesis, I am deeply indebted to many people who have made this journey possible and enriched my experience.

I would like to express my profound gratitude to my supervisors, Prof. Siby Samuel and Dr. Apurva Narayan, for their continuous support and guidance. Their insight and expertise have been invaluable, and their encouragement and faith in my abilities have greatly inspired me. I am privileged to have learned from their vast knowledge and to have been mentored by them.

My sincere appreciation also goes to my reading committee, Dr. Shi Cao and Prof. Oliver Schneider. Their critical feedback and constructive suggestions have been instrumental in shaping this thesis. I have greatly benefited from their expertise and keen understanding of the subject matter.

I extend my heartfelt thanks to all the participants who took part in the data collection process. Their time and contributions have played a pivotal role in making this research a reality.

My deepest gratitude goes to my family for their unwavering belief in me, their constant encouragement, and their unconditional love. Their emotional support and sacrifices have been my pillar of strength throughout this journey.

Special thanks to Marialissi, whose unwavering support, understanding, and love have been a beacon during this journey. Her patience and encouragement were constant sources of motivation, especially during the most challenging times.

Finally, I would like to acknowledge the larger community of scholars whose works have informed and guided my own research. Their contributions have significantly shaped my understanding and approach to this topic.

The journey of writing this thesis has been one of the most rewarding experiences of my life, and I am grateful for the opportunity to learn and grow in the process. All errors and omissions in this thesis remain my own.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Chapter 1

Introduction

1.1 Motivation

The motivation for this thesis is rooted in the substantial concern of driver fatigue, which continues to be a significant cause of road traffic accidents globally. Studies suggest that fatigue can significantly impair a driver's performance, leading to a higher risk of collisions, injuries, and fatalities [25]. Despite the growing body of evidence linking fatigue to poor driving performance, effective and reliable fatigue detection systems are not yet widespread [30]. The need for such systems becomes more pressing with the growing number of drivers and increased length of commutes.

Traditional fatigue detection methods often rely on indirect measures, such as tracking vehicle movements or relying on driver self-reports [10]. These measures have proven to be inadequate and unreliable as they either lack the sensitivity to detect early signs of fatigue or suffer from subjectivity [58]. For instance, drivers may not be fully aware of their fatigue levels or might underreport their drowsiness due to social desirability bias or underestimation of risks. Therefore, developing a system that directly measures physiological indicators of fatigue can overcome these limitations and significantly enhance our ability to detect driver fatigue accurately and promptly.

Emerging technologies, such as eye-tracking devices, heart rate monitors, and EDA sensors, provide a unique opportunity to capture direct physiological measures indicative of fatigue [22] [46] [15]. However, the complex and dynamic nature of these data requires advanced analysis techniques to accurately interpret and classify fatigue levels. Machine learning offers a promising solution to this challenge by providing sophisticated algorithms

capable of identifying intricate patterns and relationships in large and complex datasets [45] [28] [66].

Furthermore, the motivation for this study extends to the fact that previous research often employed an unimodal approach to fatigue detection, relying on a single data source. This approach overlooks the multifaceted nature of fatigue, which is likely to manifest through various physiological changes. Hence, a multimodal approach that combines different physiological measures can provide a more comprehensive and accurate detection of fatigue [30].

Ultimately, the driving motivation for this thesis is to significantly contribute to enhancing road safety. By developing a robust, comprehensive, and reliable drowsiness detection system, this research aims to enable timely interventions to prevent fatigue-related accidents, thereby potentially saving lives and reducing injuries on the roads [30] [10]. This thesis seeks to bring us closer to the goal of safer roads by providing more sophisticated and accurate tools for real-time drowsiness monitoring in various transportation settings.

1.2   Research objectives

1.2.1   General

Develop a comprehensive and robust drowsiness detection system that leverages machine learning models trained on diverse physiological indicators, including HRV, PERCLOS, BR, BP, and electrodermal activity (EDA) signals.

1.2.2   Specific

1. Develop a robust machine learning model for detecting driver drowsiness based on HRV data.

2. Design and train a separate machine learning model that can accurately analyze and interpret eye movement patterns (PERCLOS, BR, BP) for drowsiness detection.

3. Build a third machine learning model focused on interpreting EDA signals as an additional physiological indicator of drowsiness.

4. Compare the performance and reliability of each individual model in accurately detecting drowsiness in drivers.

5. Evaluate the correlation between the detected physiological indicators of drowsiness (HRV, eye movement patterns, EDA signals) and the self-reported fatigue levels using the FAS.

6. Explore the possibilities of integrating the outcomes from these individual models for a more comprehensive and robust drowsiness detection system.

7. Conduct rigorous performance assessments of each model, utilizing established evaluation metrics such as accuracy, precision, recall, and F1 score.

8. Identify the optimal machine learning models for each physiological data type, drawing from contemporary advancements such as gated recurrent unit (GRU), Transformers, Mogrifier LSTM, Momentum LSTM, Difference Target Propagation, and Decoupled Neural Interfaces Using Synthetic Gradients.

9. Contribute to the ongoing efforts to improve road safety by providing effective tools for real-time drowsiness detection and timely intervention.

1.3   Thesis overview

The remainder of this thesis is structured as follows:

- Chapter 2 offers a comprehensive review of existing literature on drowsiness, including its causes and effects, along with an overview of its measurement techniques, emphasizing the specific methods employed in this study.

- In Chapter 3, the research hypotheses are clearly defined, followed by a detailed outline of the experimental protocol. This section also delves into the specifics of the study methodology and the experimental procedure.

- Chapter 4 presents the key findings from the application of the drowsiness detection system, thoroughly analyzing and interpreting the results.

- Chapter 5 draws conclusions from the study and offers recommendations for future research, outlining potential avenues to continue exploring this important field of study.

Chapter 2

Background

2.1 Drowsiness causes and effects

Drowsiness can be described as the state in which a person is feeling excessively tired or sleepy during the day [3]. This sleeping disorder could arise by many factors, including medical conditions, lifestyle changes or medications. One of the most common medical conditions that can lead to drowsiness is diabetes. Long-term (chronic) pain, as well as hyponatremia or hypertremia (changes in blood sodium levels), can also impact metabolism and mental state, leading to drowsiness [39]. Lifestyle habits such as working for very long hours or in night shifts, not getting adequate sleep [3] can also contribute with this sleeping disorder. Medications which the user might be ingesting have drowsiness as a side effect (something common in modern-age medicine).

Drowsiness can pose a serious threat for drivers as it has been proven to have detrimental effects on their abilities. Specifically, these effects include, increased reaction times required for sudden brakes or steering changes, reduced attention levels on the road and affecting drivers ability to make sound decisions [25]. The likelihood of road accidents and loss of lives could be significantly amplified by these negative impacts.

2.2 Drowsiness measurement techniques

The measurement of drowsiness can be achieved through various techniques at different levels, which can be classified into five categories: subjective, physiological, vehicle-based, behavioural, and hybrid [71]. Currently, subjective tools are used to assess how

sleepy a subject is, and are based on questionnaires and electro-physiological measures of sleep. Some of the most well-known subjective tools are the Fatigue Assessment Scale (FAS) which is a 10-item scale evaluating symptoms of chronic fatigue [48]. The FAS treats fatigue as a uni-dimensional construct, and the questions represent a combination of physical and mental health symptoms. Additionally, the Epworth Sleepiness Scale (ESS) [35] is an eight item self assessment tool to measure the subject tendency to fall asleep in non-stressful situations (watching tv, reading, etc.). The Stanford Sleepiness Scale (SSS) [29] is another tool which has seven different statements where individuals mark their current level of alertness every two hours through the day.

Physiological techniques are more reliable and accurate in detecting drowsiness based on the driver's physical conditions as they are connected to what the driver is feeling physically at a particular time [58]. These signals have shown promising results as they start to change in the early stages of drowsiness [9]. This leads to drowsiness detection systems having extra time to notify drivers in the early stages of fatigue and therefore prevent accidents. Some of the most commonly used methods include Electrocardiogram (ECG), which records the electrical activity of the heart muscle from the body surface [41]. This data can be used to extract different heart features such as heart rate, P-R intervals, and QRS complex. Another technique used is Electroencephalogram (EEG), which records the human brain's activity and divides the signal into different bands depending on the frequency spectrum, and is able to distinguish between states of vigilance, such as wakefulness and sleep [9]. Additionally, Electrooculogram (EOG) measures the electrical potential difference between the cornea and retina of the eye [36]. However, the application of this technique is highly invasive, which is why it is not commonly used for research purposes.

Behavioural measurements are used to detect patterns in the driver's behaviour while in a drowsy state. These measurements mainly focus on identifying characteristics in facial expressions, such as constant blinking, nodding or swinging of the head, or frequent yawning [71]. Typically, these systems use video cameras and machine learning techniques to detect these patterns. Some of the most common methods are the detection of yawning and eye state. For the first, although evidence proves that yawning is a common stimulus for drowsiness, it can also be provoked by different actions such as boredom, thinking about it, reading, or listening to something [27]. Many studies have concluded that using yawning as the sole indicator for detecting drowsiness is inconclusive and may result in too many false positives.

On the other hand, eye state detection methods include, PERCLOS, blink percentage (BP), and blink rate (BR). The First one, is considered as a reliable measure of drowsiness by the National Highway Traffic Safety Administration (NHTSA) [65]. It measures the proportion of time in one minute or 30 seconds where the eye is at least 80 to 70 percent

closed. Blink Percentage (BP), refers to calculating the number of eye blinks over a specific period of time and calculating the percentage of time that the eyes are closed based on the total duration of the measurement period [17]. And the third, refers to the number of times a person blinks their eyes per minute. It is a measure of the frequency of eye blinking and can be easily obtained during a clinical examination or using a using high frame rate video analysis [20] [55].

Hybrid techniques are formed by combining other measurements techniques and leveraging their respective strengths and weaknesses. The goal is to merge different techniques to create a competitive drowsiness detection system [71]. In this research, a hybrid method will be used that integrates PERCLOS, BP, BR, HRV, and EDA. Moreover, the FAS will be applied to every participant before and after the study to assess their level of fatigue.

2.3   Overview of the variables to analyze

2.3.1   Percentage of eyelid closure over time (PERCLOS)

PERCLOS stands for Percentage of Eye Closure over the Pupil over time. It is a measure of the proportion of time that a person's eyes are closed or nearly closed during a given time interval [44]. PERCLOS is recognized as an important indicator of drowsiness and is widely used for the detection and evaluation of fatigue driving [45]. It can be calculated by dividing the total time where the eye is closed more than 80% with respect to some reference maximum opening by the window length [21]. PERCLOS can be measured using eye tracking glasses or even a dash cam and can be used to estimate the driver's state in real-time [8].

2.3.2   Blink Percentage (BP)

Blink percentage refers to the total time it takes for an eyelid to close and reopen during a blink on a determined amount of time. Studies have shown that longer blink durations are associated with increased drowsiness [40]. To measure BP, researchers often use automated algorithms to detect eye blinks in real-time. These algorithms analyze video recordings or images captured by cameras, the algorithms identify blinks based on eyelid movement and calculate the duration of the eye closed [16].

2.3.3 Blink Rate (BR)

Blink rate is defined as the number of times an individual blinks their eyes per minute. On average, a healthy human blinks approximately 15-20 times per minute [63]. However, the mean values can vary depending on factors such as age, gender, and the specific activity being performed. While it is a measure that can be easily obtained during a clinical examination, there is currently a lack of proper standardization [20]. BR can be measured either manually [47] or through high frame rate video analysis [55]. Manual measurement involves counting the number of blinks within a specific time frame, whereas high frame rate video analysis utilizes advanced imaging techniques to accurately capture and quantify BR.

2.3.4 Heart rate variability (HRV)

Heart rate variability (HRV) is a measure of the variation in time between consecutive heartbeats, which reflects the activity of the autonomic nervous system [32]. HRV analysis has been used to detect drowsiness in drivers, as changes in sleep condition affect the autonomic nervous system (ANS) and then HRV [23]. Previous studies demonstrated that driver drowsiness was successfully detected in seven out of eight cases before accidents occurred [22]. When reading HRV values to detect drowsiness, one should look for changes in both time domain and frequency domain. In the time domain, the root mean square of successive differences (Root Mean Square of Successive Differences between normal heart beats (RMSSD)) should increase because of the parasympathetic activity [61]. In the frequency domain, the high frequency (High frequency (HF)) band should increase and the low frequency (Low frequency (LF)) band should decrease because of the parasympathetic activity [67]. Therefore, HRV analysis can be used to detect drowsiness in drivers, and changes in RMSSD and HF band should be looked for when reading HRV values to detect drowsiness.

2.3.5 Overview of Electrodermal activity (EDA)

Electrodermal activity is a measure of the electrical changes that occur in the skin in response to various stimuli, including emotional or cognitive events [11]. It involves recording the electrical conductance or resistance of the skin to assess the activity of the sympathetic nervous system, making it a common physiological measure in research on emotion, stress, and other psychological processes [11]. While EDA was initially used as a

7

lie detector, it has also been found to be useful in assessing stress levels that can lead to sleep disorders and heart diseases [46].

While EDA has been shown to be useful in distinguishing between wake and sleep states, there is currently insufficient evidence to accurately identify sleep stages using EDA measures alone. Moreover, during periods of deep sleep, EDA signals may appear as elevated high frequency, which could potentially lead to misinterpretation [60].

2.3.6   Overview of Fatigue Assessment Scale (FAS)

The Fatigue Assessment Scale is a 10-item scale that was developed to evaluate symptoms of chronic fatigue [48]. It considers fatigue as a uni-dimensional construct and combines physical and mental health symptoms in its questions. Respondents are required to select one answer from a 5-point rating scale ranging from 1= Never to 5= Always. The scale has been shown to have good validity in assessing fatigue, emotional stability, and depression [48]. It is also noteworthy that the scale has demonstrated no variance related to gender or age. A score less than 22 indicates no fatigue, while scores between 22-34 suggest mild-to-moderate fatigue. Scores of 35 or higher indicate severe fatigue [18].

2.4   Drowsiness detection technologies:

2.4.1   PERCLOS for drowsiness detection

The use of PERCLOS as a measure of drowsiness detection has been widely studied and applied in various research studies. The effectiveness and reliability of PERCLOS for detecting drowsiness have been validated by the NHTSA. PERCLOS is calculated by dividing the number of images where the eyes are closed (pupil coverage of 70%, 80%, or the mean, square percentage of eyelid closure rating (EYEMAS), also known as P70, P80, and EM respectively) by the total number of images recorded within a specific time frame multiplied by 100 [45], [69], [70], [28], [15]. The time frame usually ranges from 15 to 60 seconds, and the total number of images analyzed depends on the acquisition frequency rate. For instance, if the camera records at 60Hz and the time period is 30 seconds, then the total number of images would be 1800.

The classification of PERCLOS is defined as PERCLOS less than 0.2 for an "Awake" state and PERCLOS greater than 0.2 for a "Drowsy" state, although the specific values may vary depending on the study. Some researchers have introduced a "Questionable

state" between the Awake and Drowsy states to improve the accuracy of the classification [45], [69], [70], [28], [15]. Moreover, the choice between P70, P80 or EM is still a matter of debate. However, the NHTSA recommends the use of P80 since it is considered to have the highest correlation with drowsiness [65]. Nonetheless, it should be noted that PERCLOS has some limitations such as sensitivity to lighting conditions, variations in the driver's eye shape, and the need for high processing power to analyze data in real-time [33], [42]. On the other hand, one of the main advantages of PERCLOS is that it can be implemented as a dash cam, making it a non-intrusive fatigue detection system.

2.4.2   Blink Percentage for drowsiness detection

The measurement of BP has been extensively investigated in drowsiness detection systems. The underlying principle is that as an individual becomes more fatigued or drowsy, their blinking patterns undergo changes. Notably, longer blink durations have been found to be correlated with increased drowsiness or sleepiness [12]. To ensure accurate measurement and calibration, preprocessing techniques are often employed, including the normalization of blink features using the mean and standard deviation [24]. However, there is currently no specific method universally adopted for approaching or calculating this variable.

The threshold for determining if a subject is drowsy varies across studies, but a common consensus suggests that when the proportion of closed-eye frames exceeds 30%, the driver is considered fatigued [19]. For instance, if a person's eyes remain closed for 3 minutes out of a 10-minute observation period, their BP would be 30%.

2.4.3   Blink Rate for drowsiness detection

Research has revealed that BR can vary in relation to the level of drowsiness experienced by individuals. For instance, one study observed an increase in BR as a characteristic of drowsy driving [51]. Similarly, another study found that the number of blinks tends to rise when an individual begins to feel drowsy, but decreases as wakefulness declines [4]. However, it is important to note that not all studies have shown significant differences in BR between alert and drowsy conditions [12].

Furthermore, BR is influenced by various factors, including the time of day. Spontaneous eye-blink rate tends to notably increase during the evening and during activities such as reading or driving a car. Additionally, BR can be affected by sleep deprivation [12] [62]. Therefore, these contextual elements should be taken into consideration when

interpreting BR data within the context of drowsiness detection. Moreover, it is worth mentioning that the specific BR thresholds for determining drowsiness may vary across studies and populations.

2.4.4   HRV for drowsiness detection

HRV has been investigated as a means of measuring drowsiness, a leading cause of fatal accidents, injuries, and property damage [13]. It is used to examine the Autonomic Nervous System (ANS) [32] by analyzing HRV signals, which typically exhibit changes during episodes of stress, extreme fatigue, and drowsiness [68]. Using HRV as a measure of drowsiness offers several benefits, including its non-invasiveness, ease of measurement, and ability to identify ANS activity changes [68]. HRV-based drowsiness detection methods have been suggested and validated with EEG (electroencephalography)-based sleep scoring [22], [43]. However, using HRV as a drowsiness measure could also have drawbacks, such as the potential impact of illuminance and color temperature on HRV measures [52], and the need for additional information, such as ECG-derived respiratory information, driver face images, or other physiological measures in some HRV-based drowsiness detection methods [22], [43], [64].

2.4.5   EDA for drowsiness detection

Electrodermal activity (EDA) has been utilized in several studies to measure drowsiness [60], [31], [46], [56]. EDA measurement can be classified into two types: Endosomatic method of measurement, which involves finding the amount of current passing through the skin while the potentials are kept constant, and Exosomatic method, which involves measuring the skin potential when the current flow is kept constant; this method measures the skin conductance of the user [46].

An advantage of EDA is that its signals tend to become unstable when participants become drowsy, making it a potential indicator of drowsiness [46]. Furthermore, EDA data can be collected non-invasively during normal day-to-day activities, without requiring controlled laboratory settings [46]. However, despite its potential usefulness in detecting drowsiness, little research has been conducted on implementing EDA as a feature to detect driving drowsiness, and its accuracy in detecting driving drowsiness remains unknown [31].

2.5 Combined use of PERCLOS, Blink Rate, Blink Percentage, EDA and HRV

Heart rate variability (HRV), electrodermal activity (EDA), and PERCLOS have served as variables in previous studies to detect drowsiness [22] [46] [15]. Technological advancements have enabled the non-invasive acquisition of these variables. Studies have indicated the effectiveness of HRV-based drowsiness detection systems [13] [68] [43] [53] [66]. A model that analyzes HRV in the HF band, LF band, and the LF/HF features has demonstrated promising results in detecting drowsiness in drivers [43]. Similarly, studies focusing on PERCLOS as the primary variable have successfully detected drowsiness in drivers across various scenarios [45] [69] [70] [28] [15]. In one study [69], the combination of a dashcam and facial recognition achieved close to 95% accuracy in identifying drowsiness. EDA studies have revealed significant differences in data obtained from different activities [31] [46] [56]. For instance, during sleep, EDA tends to exhibit a steady pattern, whereas during drowsiness, EDA displays high variability [46], suggesting that EDA could be a crucial variable for detecting drowsiness in drivers. BR and PERCLOS have also been utilized to detect drowsiness [38]. A study focused on real-time drowsiness detection combined these variables with the Haar algorithm for face detection, resulting in a system that achieved 98.55% accuracy in identifying drowsy drivers. Experimental evaluations of BR and BP [12] have also yielded meaningful insights as drowsiness measures, highlighting the significance of analyzing the duration of blink closures.

Several studies have demonstrated the potential benefits of combining PERCLOS and HRV [15] [66] [64], leading to a more robust algorithm for detecting drowsiness. Since PERCLOS can be influenced by lighting conditions during video capture, HRV can still serve as a viable variable for detecting drowsiness, and vice versa. For example, if the driver's watch battery dies, HRV may become unavailable, but PERCLOS continues to collect data and monitor the driver's fatigue state. HRV and EDA serve as useful tools for assessing the driver's sympathetic nervous system, although limited studies have examined their application in fatigue assessment among drivers. Nevertheless, one study demonstrated promising results in assessing drivers' drowsiness solely by utilizing these variables [56].

2.6 Previous studies on PERCLOS, Blink Rate, Blink Percentage, EDA, and HRV for drowsiness detection

In recent years, there has been increasing interest in using machine vision-based techniques to detect driver fatigue. One such approach involves fatigue driving detection based on facial multi-feature fusion, which combines different facial characteristics to identify signs of fatigue [45]. This method employs machine learning models such as YOLO-V3 convolutional neural network to capture every facial region. Additionally, the Dlib toolkit is used to introduce Eye Feature Vector (EFV) and Mouth Feature Vector (MFV), which evaluate the state of the eyes and mouth, respectively. The drivers' closed eyes time, blink frequency, and yawn frequency are then calculated to evaluate their fatigue state. By following this approach, researchers have achieved a 95.10% accuracy in detecting driver fatigue.

A real-time comprehensive algorithm has been developed based on drivers' facial and landmark expressions to detect fatigue [70]. This algorithm utilizes a deep convolutional neural network (DCNN) and extracts the eye feature of the eye aspect ratio (EAR), mouth aspect ratio (MAR), and PERCLOS from video recordings based on facial landmarks. The overall accuracy of the system has been achieved at 95.1%. Similarly, another study utilized the driver's facial features for real-time fatigue detection [69]. The algorithm comprises three parts: facial feature location, state recognition, and fatigue state estimation. Firstly, the multitask convolutional neural network (MTCNN) and practical Facial Landmark Detector (PFDL) classification networks are used to extract the eyes and mouth. Then, a sliding window model is applied to evaluate driver fatigue state by combining PERCLOS, blinking frequency, and yawning frequency. The team achieved a remarkable accuracy of 95% in detecting drowsiness with an average detection time of 38.72 ms.

Another study related to PERCLOS utilized a multivariate multiscale entropy (MMSE) method to analyze various features, including electroencephalogram (EEG) and electrooculogram (EOG) signals, to differentiate between drowsy, tired, and awake states [8]. The MMSE approach was also used to identify PERCLOS values. By combining the MMSE analysis with a Support Vector Machine (SVM) model and incorporating EEG signals and PERCLOS values, the authors were able to achieve a classification accuracy of 76.2%. Furthermore, PERCLOS and the Stanford Sleeping Scale (SSS) were used to determine the threshold of fatigue degree, depending on the monotonous scenario [28]. Data was collected from 36 participants who completed a driving simulation that included both roadway and low-grade rural areas as scenarios. A Long short-term memory network (LSTM) was uti-

lized to detect various fatigue levels. The results indicated that fatigue levels increased as driving time increased in both scenarios, with drivers more prone to experiencing fatigue on the expressway rather than in the rural area. The different states of fatigue were categorized into awake, mild fatigue, moderate fatigue, and severe fatigue, with a recognition rate of 97.8%.

A study implemented deep learning techniques with adaptive neural networks based on MobileNet-V2 and ResNet-50V2 [54]. The study focused on detecting blink patterns such as BR and BP by counting the duration and number of blinks and setting individual eye-opening thresholds. The study achieved a high accuracy of 97% in detecting driver drowsiness, demonstrating its practical potential in preventing accidents [55]. An algorithm was introduced in [55] that uses high frame rate video analysis to detect driver drowsiness. The algorithm incorporates blink features such as duration, PERCLOS, blink frequency, and amplitude-velocity ratio using fuzzy logic. It demonstrated over 80% accuracy in drowsiness detection and is independent of the driver, requiring no adjustments. Additionally, [34] develops a drowsiness prevention system using machine learning, facial recognition, and eye-blink recognition. The system tracks blink speed using a Haar classifier and CO2 concentration in the vehicle to detect drowsiness. If drowsiness is detected, the system activates music or guides the driver to a rest area. The study shows improved accuracy and reduced error rate in drowsiness detection.

In a study aimed at enhancing the robustness of drowsiness detection systems, the LF/HF ratio from heart rate variability (HRV) analysis and PERCLOS were incorporated to accurately and robustly determine fatigue levels in drivers [15]. To capture the data, a near-infrared webcam was used to obtain facial images under different light sources, and the appropriate RGB channel was utilized to obtain the LF/HF ratio from HRV. The study evaluated the sympathetic/parasympathetic balance index and PERCLOS of 10 wakeful participants and 30 drowsy participants, resulting in an accuracy rate of 92.5%. Likewise, using Standard deviation of the IBI intervals measured in ms (SDNN), RMSSD, LF, HF, and PERCLOS variables obtained from HRV analysis to detect drowsiness [66]. Back Propagation Neural Networks (BPNN) and long short-term memory (LSTM) were used for the analysis. By using LSTM on PERCLOS, a true positive rate of 75% and an accuracy of 88% were achieved. Using BPNN on HRV, a true positive rate of 80% and an accuracy of 88% were obtained.

Furthermore, a study utilized eyelid closure and HRV to estimate drowsiness levels [64]. Two multi-class classifiers, namely, error-correcting output coding (ECOC) and loss-based error-correcting output coding (LD-ECOC), were employed for this purpose. The data was collected from five subjects across ten different experiments, each lasting for 20 minutes. The ECOC classifier achieved an accuracy of 90.69%, while the LD-ECOC

classifier achieved an accuracy of 88.78%.

On the other hand, it has been suggested in some studies that focusing only on HRV analysis might be sufficient [22]. In this proposed method, validation was performed by comparing it with an electroencephalography (EEG)-based sleep scoring system. Eight HRV features were extracted, and the results demonstrated that drowsiness was detected in 12 out of 13 cases, with a false positive rate of 1.7 times per hour. Also, a comparison between machine learning algorithms and deep learning methods was made using HRV analysis [13]. The machine learning classifiers used in the study were Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), and Naive Bayes (NB). Additionally, a one-dimensional convolutional neural network (1D CNN) was used for deep learning. The study achieved a 99.94% accuracy using the 1D CNN method.

Moreover, a paper presents a neural network-based approach that utilizes HRV analysis to detect drowsiness in its early stages [53]. The analysis focuses on three HRV features, namely, HF which represents parasympathetic activity, LF which represents sympathetic activity, and the ratio of both (LF/HF). The study collected data from 12 participants and achieved a 90% accuracy on the model. In contrast to this, a paper researched into developing a Samsung Gear S smartwatch application to detect drowsiness using HRV time domain features R-R and RMSSD [59]. While the idea is promising, the study did not provide any evidence of its effectiveness or real-life application. Further research and testing are required to validate the potential of this application.

Subsequently, a system was designed to detect the driver's heartbeat using a pulse sensor [57]. The system was connected to an Arduino, and the variables captured from HRV included LF, HF, and LF/HF. The study also reported that the LF/HF ratio demonstrated decreasing trends when the driver transitioned from an awake and alert state to a state of drowsiness. In comparison, a study implemented fabric electrodes mounted on the steering wheel, the researchers monitored the driver's heart rate and obtained HR, SDNN, RMSSD, proportion of NN50 divided by the total number of NN (R-R) intervals (pNN50), LF, and HF from the original ECG signals [37]. With this system, they were able to detect drowsiness and fatigue. Although the accuracy of the results was not very promising, the authors noted that they were able to capture HRV in a non-invasive and convenient manner.

Even though HRV and PERCLOS for drowsiness detection has been deeply studied, EDA which hasn't has shown some interest in some studies. A hybrid model was created to detect drowsiness by capturing Electro Dermal Activity (EDA), respiration (RESP), and Photoplethysmography (PPG) data [31]. Different models were compared including k-nearest neighbours (KNN), back propagation neural networks (BPNN), and Cascade

14

forward neural networks (CFNN). The best average accuracy was obtained from the CFNN, reaching 97%, with an average training time of 0.3 s. Additionally, An EDA-based wearable device was developed to detect drowsiness in individuals [46]. EDA was selected as the variable to capture, as it is related to sweat, which directly corresponds to the subject's mental state. Although no machine learning or additional studies were conducted to create a drowsiness detection system, the author demonstrated the relationship between EDA activity in a participant's active, asleep, and drowsy states using this device.

Table 2.1: Summary of Fatigue Detection Methods

No.	Method	Techniques	Variables	Accuracy	Additional Information	Ref.
1	Fatigue detection based on facial multi-feature fusion	YOLO-V3 CNN, Dlib toolkit	EFV, MFV, closed eyes time, blink frequency, yawn frequency	95.10%	-	[45]
2	Real-time fatigue detection using facial and landmark expressions	DCNN	EAR, MAR, PERCLOS	95.1%	-	[70]
3	Real-time fatigue detection using driver's facial features	MTCNN, PFDL	PERCLOS, blinking frequency, yawning frequency	95%	Average detection time: 38.72 ms	[69]

Continued on next page

Table 2.1 – *Continued from previous page*

No.	Method	Techniques	Variables	Accuracy	Additional Information	Ref.
4	Fatigue analysis using EEG and EOG signals	MMSE, SVM	EEG, EOG, PERCLOS	76.2%	-	[8]
5	Fatigue detection in driving simulation using PERCLOS and SSS	LSTM	PERCLOS, SSS	97.8%	Awake to severe fatigue categorization	[28]
6	Drowsiness detection using deep learning techniques and adaptive neural networks	MobileNet-V2, ResNet-50V2	Blink patterns (BR, BP), duration, number of blinks, eye-opening thresholds	97%	-	[54], [55]
7	High frame rate video analysis for drowsiness detection	Fuzzy logic	Blink duration, PERCLOS, blink frequency, amplitude-velocity ratio	80%	Independent of the driver	[55]
8	Drowsiness prevention system using machine learning and facial recognition	Haar classifier	Blink speed, CO_2 concentration	Not specified	Activates music or guides to rest area	[34]

Table 2.1 – *Continued from previous page*

No.	Method	Techniques	Variables	Accuracy	Additional Information	Ref.
9	Drowsiness detection using LF/HF ratio from HRV analysis and PERCLOS	Near-infrared webcam, RGB channel	LF/HF ratio from HRV, PERCLOS	92.5%	Evaluated on 10 wakeful and 30 drowsy participants	[15]
10	Drowsiness detection using HRV analysis variables	BPNN, LSTM	SDNN, RMSSD, LF, HF, PERCLOS	88% (LSTM), 88% (BPNN)	True positive rate: 75% (LSTM), 80% (BPNN)	[66]
11	Estimation of drowsiness levels using eyelid closure and HRV	ECOC, LD-ECOC	Eyelid closure, HRV	90.69% (ECOC), 88.78% (LD-ECOC)	Data from 5 subjects across 10 experiments	[64]
12	Drowsiness detection using HRV analysis compared with EEG-based sleep scoring	Not specified	HRV features	Not specified	False positive rate: 1.7 times per hour	[22]

Table 2.1 – *Continued from previous page*

No.	Method	Techniques	Variables	Accuracy	Additional Information	Ref.
13	Comparison of machine learning and deep learning methods using HRV analysis	RF, SVM, DT, NB, 1D CNN	HRV analysis	99.94% (1D CNN)	-	[13]
14	Neural network-based approach using HRV analysis for early drowsiness detection	Neural Network	HF, LF, LF/HF	90%	Data from 12 participants	[53]
15	Samsung Gear S smartwatch application using HRV time domain features	Not specified	R-R, RMSSD	Not specified	No evidence of effectiveness provided	[59]
16	Driver's heartbeat detection using a pulse sensor connected to Arduino	Pulse sensor, Arduino	LF, HF, LF/HF	Not specified	LF/HF ratio showed decreasing trends	[57]

Continued on next page

Table 2.1 – *Continued from previous page*

No.	Method	Techniques	Variables	Accuracy	Additional Information	Ref.
17	Drowsiness detection using fabric electrodes mounted on the steering wheel	Fabric electrodes	HR, SDNN, RMSSD, pNN50, LF, HF	Not specified	Non-invasive HRV capture	[37]
18	Drowsiness detection using EDA, RESP, and PPG data	KNN, BPNN, CFNN	EDA, RESP, PPG	97% (CFNN)	Average training time: 0.3 s	[31]
19	EDA-based wearable device for drowsiness detection	Not specified	EDA	Not specified	Demonstrated relationship between EDA states	[46]

Chapter 3

Current Study

Distinct machine learning models, trained individually on eye movement data, heart rate variability (HRV) features, and electrodermal activity (EDA) signals, will each capture unique manifestations of drowsiness, thereby providing complementary information. This integrated approach will contribute to a comprehensive and robust drowsiness detection system.

This thesis aims to develop an advanced, comprehensive, and robust drowsiness detection system by leveraging distinct machine learning models specifically tailored to eye movement data, HRV features, and EDA signals. Each data source captures unique physiological changes associated with drowsiness, and their combined analysis can yield a more comprehensive understanding of this state. This study will involve collecting these physiological data from a diverse group of participants under controlled conditions. Three separate machine learning models, specifically optimized for each data source, will be developed and trained. The performance of each model will be rigorously evaluated using appropriate evaluation metrics, including accuracy, precision, recall, and F1 score. This study also seeks to determine the potential benefits of integrating the outputs from these individual models in enhancing the overall accuracy and reliability of the drowsiness detection system. Additionally, this thesis will aim to address potential confounding factors, adhere to strict ethical guidelines, and underscore the need for a multimodal approach to build a comprehensive and effective drowsiness detection system.

3.1   Materials

3.1.1   Questionnaires and Activity Log

The participants completed various questionnaires, including the MSSQ-Short, FAS, demographic survey, and an Activity Log.

The Motion Sickness Susceptibility Questionnaire-Short form (MSSQ-Short [26], see Appendix A) was used to determine an individual's susceptibility to motion sickness and the specific types of motion that cause discomfort, such as queasiness, nausea, or vomiting. Participants rated their level of sickness on a scale of 0-3 for various modes of transportation and entertainment attractions, including cars, trains, aircraft, and funfair rides.

In addition, the participants were required to fill out an activity log that collected information about their recent activities. This included data about the amount of sleep they had and details about their meals, such as what they had for dinner, breakfast, and lunch (if applicable). This additional information aimed to provide a more comprehensive understanding of the participants' physiological state and potential factors that could influence their drowsiness levels during the test.

To evaluate fatigue levels, the Fatigue Assessment Scale (FAS) [48] (see Appendix B) was employed. This questionnaire assesses fatigue as a unidimensional construct, incorporating both physical and mental health symptoms. Participants rate their typical experiences on a scale from 1-5 for various statements, such as feeling tired quickly, having difficulty thinking clearly, and experiencing mental exhaustion.

To collect participant responses, questionnaires including the demographic, the MMSQ-Short, FAS, and the Activity Log, as well as consent forms, were hosted on the Qualtrics online platform. This allowed participants to conveniently access and complete the necessary forms before the experiment.

3.1.2   Drowsiness Detection Sensors

Drowsiness was measured using the Ergoneers Dikablis Glasses 3 eye-tracker (See figure 3.1) and the Empatica E4 wristband (See figure 3.2).

Figure 3.1: Ergoneers Dikablis Glasses 3 eye-tracker.

The Ergoneers Dikablis 3 is a wearable wired binocular eye-tracker with a 60 Hz eye camera tracking frequency, 0.05° visual angle pupil tracking accuracy, and 0.1°-0.3° visual angle glance direction accuracy. It includes an eye camera with a resolution of up to 648×488 pixels and a field camera with a resolution of 1920×1080 pixels at 30 fps (frames per second). Calibration is performed to ensure accurate tracking by having the participant fixate on markers placed at four corner points of the screen. The eye-tracker also captures the wearer's pupils using a red circular cross-hair.

Figure 3.2: Empatica E4 wristband.

The Empatica E4 wristband combines EDA and PPG sensors, allowing simultaneous

measurement of sympathetic nervous system activity and heart rate at a sampling rate of 64Hz. The PPG sensor measures Blood Volume Pressure (BVP), Interbeat Interval (IBI), and heart rate (HR). The GSR sensor, also known as the EDA sensor, detects fluctuating changes in certain electrical properties of the skin. The wristband also includes a 3-axis accelerometer to capture motion-based activity and a skin temperature sensor displayed on Figure 3.3. The device connects via Bluetooth to a smartphone, where the data is stored in the cloud and automatically synchronized before recording.
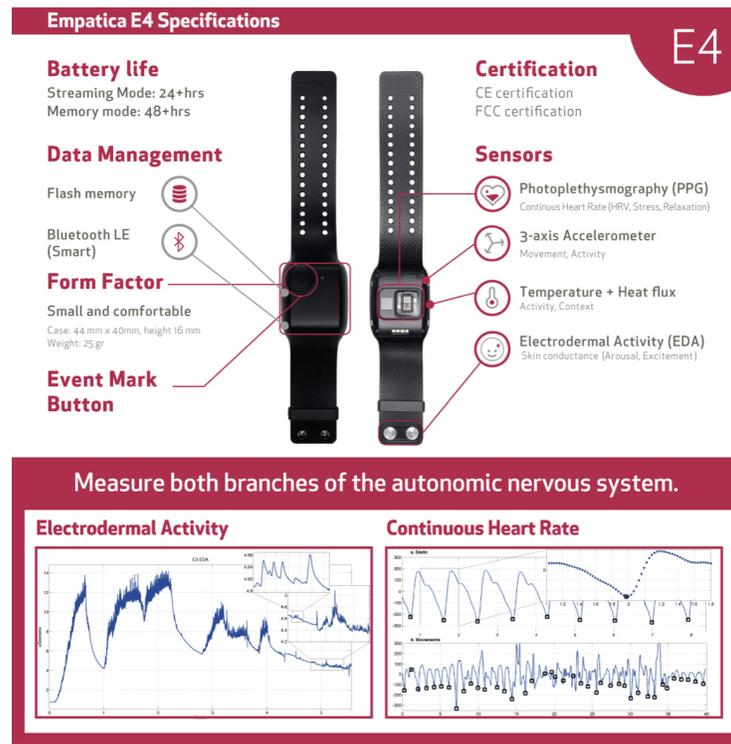
Figure 3.3: Empatica E4 features and placement of the sensors.

3.1.3   Simulation Setup

The driving scenario was implemented using CARLA [2], an open-source autonomous driving simulator (see Figure 4). The simulation and eye tracking software were installed on a computer equipped with an AMD Ryzen 9 5950Xi processor, 32 GB of RGB RAM (4x8 GB), an NVIDIA GeForce RTX 3080 LHR Graphics Card, and GIGABYTE AORUS Gen 4 7000s M.2 1TB SSD. The setup also included a Logitech G29 steering wheel controller

with force feedback and a set of pedals for vehicle control. The driving scenarios were displayed on a single 27-inch Full HD 1080p LED monitor, providing participants with a high-resolution and wide-angle view of the road ahead.

Figure 3.4: Sample driving scenario.

3.2 Participants

The experimental setup, including the arrangement of the sensors on the participants, is shown in Figure 3.5. Participants eligible for the research were required to be over 18 years of age and hold a valid Canadian Driver's License (e.g., Ontario Class G2 or G) with at least one year of driving experience. Moreover, they had to be active drivers during the time of the experiment.

Participants were screened using a cut-off score of 23 on the MSSQ-Short, and those lacking full or corrected eye vision were excluded from the experiment. Participants were randomly selected to participate in either the monotonous or non-monotonous scenario. The group chosen for the non-monotonous scenario served as a control group and was compared against the monotonous participants. Recruitment occurred following ethical approval to conduct the study, through channels such as social media, flyers, and email lists. The participant pool primarily comprised students from the University of Waterloo, as well as neighboring universities and colleges.

Figure 3.5: Sample driving scenario.

Through the advertisements, participants were requested to complete a demographic questionnaire. Those who met the requirements were subsequently contacted for participation in the study. A sample of 33 participants was recruited, but due to issues with the eye tracker sensor, incomplete data was generated for a total of 30 participants. Table 3.1 presents a summary of the participant characteristics in the monotonous and non-monotonous scenarios. The age of the participants ranged from 18 to 43, with a median of 25, and a standard deviation of 4.93. There were 21 male and 12 female participants. 19 participants held a valid Canadian G driver's license or its equivalent, while 14 had a G2 driver's license or its equivalent. The driving experience of participants varied from 1 to 26 years, with an average of 6.56 years.

Participants reported driving between 2000 and over 30000 kilometers on average per year. Furthermore, 20 participants self-reported daily consumption of caffeine-related products, with the most common product being Coffee. They consumed an average of 1.8 cups per day, but all participants declared that they did not consume on the day of the test. When asked about their tendency to drive while feeling drowsy or sleepy on a scale of 0-10, they reported an average score of 3.78.

Table 3.1: Participant Characteristics

Characteristic	Monotonous	Non-Monotonous
Age range (years)	20-30	18-43
Average age (years)	24.53	25.86
Age standard deviation	3.46	6.36
Number of Males	9	11
Number of Females	6	4
G Driving license	7	11
G2 Driving license	8	5
Driving experience range (years)	1-13	1.5-26
Average driving experience (years)	4.93	8.23
Average Km driven per year	9.13	13.43
Consumption of caffeine daily	7	11
Average consumption daily (cups)	1.8	1.83
Most consumed product	Coffee	Coffee
Average tendency to drive while feeling drowsy (1-10)	3.79	4

3.3   Experimental design

The study implemented a mixed factorial design ($2 \times 2 \times 2$; scenario type $\times$ driving license type $\times$ daily caffeine consumption), with age as a covariate. The independent variables between subjects included scenario type (monotonous vs non-monotonous), driving license type (G vs G2), and daily caffeine consumption (yes vs no).

Additionally, the study conducted multiple within-subject measurements related to drowsiness. These measurements comprised the Percentage of Eyelid Closure Over Time (PERCLOS), Blink Percentage (BP), Blink Rate (BR), Heart Rate Variability (HRV), and Electrodermal Activity (EDA). Notably, all these metrics were continuously tracked during the course of the experiment.

3.3.1   Dependent variables

The dependent variable for this investigation is the "Level of Drowsiness" in participating subjects. Drowsiness is quantified using a three-tiered ordinal scale: "0" indicating "No Drowsiness," "1" indicating "Possible Drowsiness," and "2" indicating "Definite

Drowsiness." This classification is made based on a combination of established physiological and visual indicators, including specific values and patterns in the independent variables

3.3.2   Independent variables

Several independent variables have been identified for this study, each offering different measures related to the physiological and visual cues associated with drowsiness.

Percentage of eyelid closure over time (PERCLOS): it is a robust indicator of drowsiness, gauging the percentage of time a subject's eyes are at least 80% closed over a given observation window.

Blink Percentage (BP): it measures the proportion of time during which a participant is observed to be in a blink state as compared to an open-eye state in a defined observation window.

Blink Rate (BR): it measures the number of complete blink cycles (from the onset of eye closure to the reopening of the eye) that a participant executes per minute. An increased BR can be indicative of heightened stress, fatigue, or other factors associated with drowsiness.

Heart Rate Variability (HRV): it is a measure of the variance in time elapsed between successive heartbeats. High levels of HRV can signal higher stress or fatigue levels, potentially indicative of a drowsy state. For this study the variables analyzed in HRV are Root Mean Square of Successive Differences between normal heartbeats (RMSSD), Low Frequency (LF), and High Frequency (HF).

Electrodermal Activity (EDA): also known as galvanic skin response, quantifies the variations in the electrical characteristics of the skin. Changes in EDA can indicate heightened levels of stress or arousal, both of which may correlate with drowsiness.

3.3.3   Control variables

To ensure the validity of the study, two control variables have been identified:

Caffeine Consumption: As caffeine is a known stimulant that can impact physiological indicators associated with drowsiness, all subjects will be instructed to abstain from consuming caffeine-related products on the day of the test.

Driving Scenario: To control for environmental influences on drowsiness, the driving scenarios used in the simulator will be classified as either 'Monotonous' or 'Non-Monotonous'. This classification will be accounted for during the data analysis phase to control for its potential effects on the state of drowsiness.

3.4   Procedure

Before participating in the study, consent was obtained from all participants. They filled out a brief demographics questionnaire, as presented in Appendix C, to provide personal information like age and driving experience, as well as the Motion Sickness Susceptibility Questionnaire Short-Form (MSSQ-Short) [26]. A second, more detailed form was also completed (see Appendix A) that asked about their driving experience, sleep schedules, and caffeine consumption rates. This form included questions regarding average yearly kilometers driven, weekly driving duration, their caffeine intake habits, and the duration of their typical continuous work hours.

The questionnaire segment concluded with the participants filling out the Fatigue Assessment Scale (FAS) (refer to Appendix B) and an Activity Log (see Appendix D). After adjusting their seat positions as needed, the Ergoneers Dikablis Glasses 3 eye-tracker and the Empatica E4 were fitted and calibrated for each participant. Before initiating the experiment, participants were familiarized with the simulator and its controls. A preliminary test run on the driving simulator allowed them to acclimate to the controls and speed while simultaneously recording baseline HRV data.

Once the participants expressed comfort and confidence in managing the driving task and vehicle functions, typically achieved within 5-10 minutes of driving, the primary phase of the experiment commenced. This involved a 45-minute drive on a specific scenario (either monotonous or non-monotonous). Both during the actual experiment and in the test run, participants were instructed to drive close to the posted speed limit of 110 km/h. Those in the monotonous group were additionally instructed to maintain their lane throughout the test.

To conclude, participants completed the FAS a second time to facilitate a comparative analysis of their results. Each participant was then thanked for their contribution, compensated accordingly, and provided with a verbal debriefing to conclude the experiment. The collection of data became possible with the approval of ethics board clearance ORE#44726, as indicated in Appendix E.

3.5 Driving scenarios

The study encompassed two distinct driving scenarios: monotonous and non-monotonous. Participants were randomly assigned to one of these scenarios. To maintain a level of realism, each scenario included AI vehicles on the road, primarily composed of sedans and SUVs. This was achieved through a script designed to spawn a specific set of vehicle types.

Each scenario's main differences are outlined below:

1. In the monotonous scenario, 50 vehicles were present on the road, as opposed to 100 in the non-monotonous scenario. Although a certain level of realism was aimed for in the monotonous scenario, the number of vehicles on the road was kept to a minimum to avoid causing significant distraction.

2. The non-monotonous scenario featured trees and surrounding environment close to the road, whereas in the monotonous scenario, the environment was dispersed and kept at a distance from the road to minimize potential distractions.

Participants drove a blue Dodge Charger for 45 minutes at a speed between 90 km/h and 120 km/h on a four-lane closed loop highway of 8.1 km long. To ensure consistent conditions across both groups, the road conditions, climate, and the highway itself remained the same for both scenarios, with a sunny, clear day prevailing in each. Speed limit signs were posted every 900 meters as stated on [6], indicating a maximum speed of 110 km/h. Please refer to figures 3.6 and 3.7 for visual representation of the monotonous and non-monotonous scenarios, respectively.

Figure 3.6: Monotonous Scenario. Road View.

Figure 3.7: Non-Monotonous Scenario. Road View.

3.6   Model Selection

This thesis involved the construction of three distinct machine learning models, each designed to handle a specific type of variable: Heart data, Eye data, and Electrodermal Activity data. Text-based data commonly employs various models, including Linear Regression, Logistic Regression, Decision Tree, sequential layers, Support Vector Machines, and even 1-Dimensional Convolutional Neural Networks. Given that HRV data consists of text-based numeric values, a Sequential neural network model with multiple hidden layers was selected. This model possesses the ability to analyze non-linear relationships within the data, recognize hidden patterns (Generalization), make accurate predictions on unseen data, and exhibit strong performance and flexibility by allowing the addition of any number of hidden layers and different optimizers. Additionally, it excels in capturing complex data patterns [50].

For Electrodermal Activity data, which also involves text-based numeric values, a 1-Dimensional Convolutional Neural Network (1D-CNN) classification model was utilized. Due to the limited amount of research on using EDA for drowsiness detection, a more robust model was chosen. This model can effectively capture complex patterns and identify them as potential drowsy scenarios [14]. It demonstrates proficiency in handling sequential data, possesses strong generalization abilities [7], and is capable of analyzing spacial patterns within the data, such as temporal changes in mean, standard deviation, and range values [1], which are associated with patterns related to drowsiness.

For the eye data, which primarily consists of images and text used to identify labels, a Convolutional Recurrent Neural Network (CRNN) is employed for analysis. The convo-

lutional part of the model is primarily utilized to analyze the frames of videos and extract spatial features from them. On the other hand, the recurrent neural network is utilized to capture temporal dependencies within the sequence. This model was chosen due to its capacity to effectively extract spatial features from eye images while considering the temporal dynamics [2], handle variability in eye images such as changes in lighting conditions, different eye shapes, and varying image resolutions, and its ability to automatically learn relevant features from the eye data [49]. The convolutional layers progressively learn hierarchical representations of eye images, extracting more abstract features over time.

3.7   Model Architecture

3.7.1   HRV-Model

The HRV model employed a Sequential neural network architecture with multiple hidden layers, as depicted in Figure 3.8. The initial layer, Dense(64, input dim = 3, activation = tanh, kernel regularizer = regularizers.l2(0.001)), consisted of 64 units and accepted a 3-dimensional input. It utilized the hyperbolic tangent (tanh) activation function and incorporated L2 regularization with a coefficient of 0.001. Following this, subsequent layers were added, such as Dense(32, activation = relu, kernel regularizer = regularizers.l2(0.001)), Dense(16, activation = relu, kernel regularizer = regularizers.l2(0.001)), Dense(8, activation = relu, kernel regularizer = regularizers.l2(0.001)), and Dense(4, activation = relu, kernel regularizer = regularizers.l2(0.001)). These layers followed a similar pattern but with decreasing numbers of units. They utilized the rectified linear unit (ReLU) activation function and included L2 regularization with a coefficient of 0.001.

The final layer, Dense(3, activation='softmax'), consisted of 3 units, representing the number of classes in the classification task. It employed the softmax activation function to generate a probability distribution across the classes.

For model compilation, the Adamax optimizer was used with a learning rate of 0.002, and the categorical cross-entropy loss function was employed.
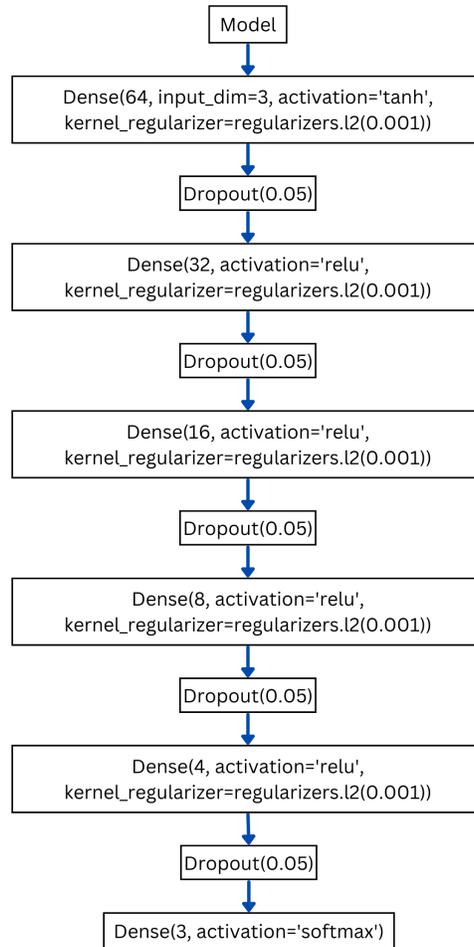
Figure 3.8: HRV Sequential neural network model.

3.7.2 EDA-Model

For the EDA data, a 1-Dimensional Convolutional Neural Network (1D-CNN) model was employed for classification, as depicted in Figure 3.9. The model consisted of multiple layers that worked together to process the data and make predictions.

The first layer, Conv1D, applied 64 filters of size 3 to the input data. It utilized the Rectified Linear Unit (ReLU) activation function and expected input shapes of (X_train.shape[1], 1). This layer performed convolutional operations on the input data,

extracting relevant features.

The output from the Conv1D layer was then passed to the Flatten layer. This layer transformed the multi-dimensional feature maps obtained from the previous layer into a 1-dimensional vector. By doing so, it prepared the data for input into a fully connected layer.

The subsequent layer, Dense, consisted of 32 neurons with a ReLU activation function. It served as a fully connected layer, processing the flattened features from the previous layer. To mitigate overfitting, a kernel regularizer with L2 regularization (weight decay) of 0.01 was applied.

To prevent overfitting further, a Dropout layer was included in the model. This layer randomly set a fraction of the input units to 0 during training, specifically 0.5% or 0.005. By doing so, it encouraged the model to learn more robust and generalized representations.

The final layer, Dense, had 3 neurons representing the output classes. It employed the softmax activation function, which provided a probability distribution across the classes. The predicted class was determined by selecting the class with the highest probability.

By utilizing this architecture, the 1D-CNN model effectively processed the EDA data, extracting relevant features, and making accurate predictions for the classification task.
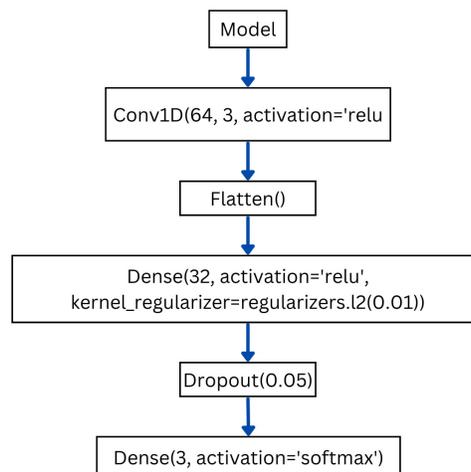

Figure 3.9: EDA 1-Dimensional Convolutional Neural Network (1D-CNN) model.

3.7.3   Eye-Model

The eye data was processed using a Convolutional Recurrent Neural Network (CRNN) model. Initially, the model incorporates two input layers: left_eye_input and right_eye_input, which represent the input sequences for the left and right eye images, respectively. These inputs are structured as (None, 64, 64, 1), denoting a variable-length sequence of grayscale images with dimensions of 64x64.

Following that, a function is implemented to define a sub-model responsible for processing each eye input. This sub-model consists of a series of TimeDistributed layers, which perform convolutional operations, batch normalization, max pooling, and dropout. The convolutional layers progressively increase the number of neurons employed (16, 32, 64, 126), while the dropout rate is set at 0.25. The final outcome of this function involves flattening the last convolutional layer, generating a feature representation for each eye. These representations are then concatenated using the Concatenate layer, resulting in a combined representation of both eyes.

Subsequently, three LSTM layers are stacked. Each LSTM layer comprises 128 units and incorporates a dropout rate of 0.25. It returns both the output sequences and the final states, with the output of the preceding LSTM layer serving as the input for the subsequent layer. Additionally, the decoder LSTM layer contains 128 units and applies dropout at a rate of 0.2. It takes the repeated encoder output as input and generates output sequences.

To produce the final predictions for each time step, the decoder_dense layer employs a TimeDistributed dense operation with num_classes units and softmax activation. The overall model is then defined using the Model class from Keras, with the left and right eye inputs as inputs and the decoder output as the output.

During the compilation phase, the model is configured with the Categorical Crossentropy loss function, the Adamax optimizer with a learning rate determined by the ExponentialDecay schedule, and various evaluation metrics including accuracy, precision, and recall.

Throughout the training process, several callbacks such as ModelCheckpoint and EarlyStopping are implemented. These callbacks serve to save the best model and halt training when the validation loss ceases to improve. The model is trained using the fit method, which involves providing the appropriate data generators, the number of steps per epoch, and the validation data.

In summary, the architectural design involves the following steps: first, the left and right eye images are taken as input; then, they are processed independently, producing separate representations; these representations are combined using concatenation; LSTM

layers are employed for sequence modeling; and finally, the model generates the final predictions. This process is depicted in Figure 3.10.
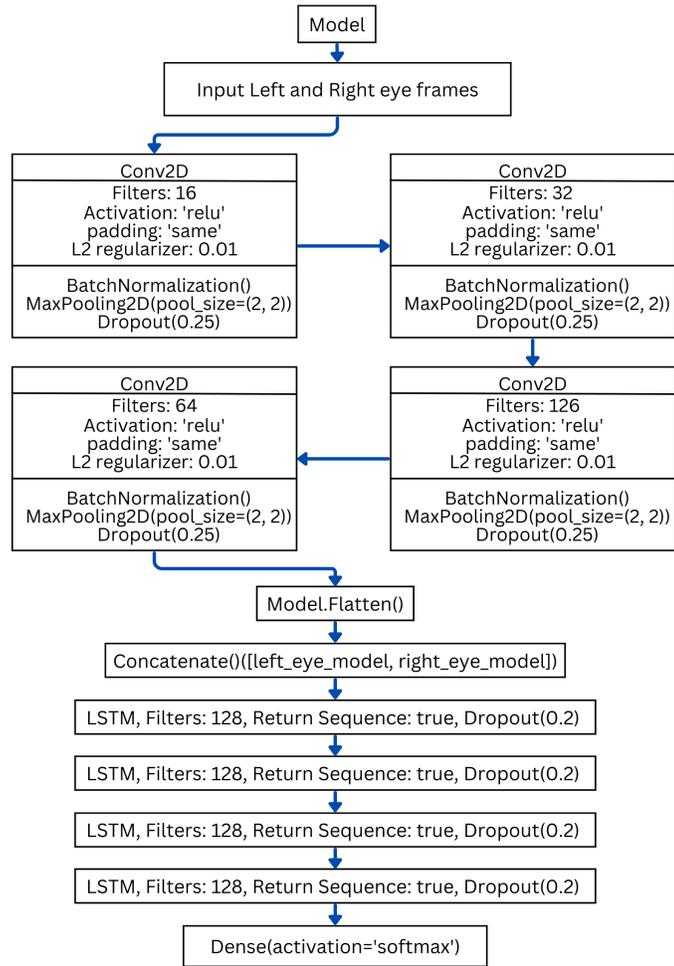

Figure 3.10: Eye-Model Convolutional Recurrent Neural Network (CRNN) Model.

3.8 Data Preprocessing

During this phase, various data types acquired from the experiments were processed utilizing distinct methodologies.

Initially, the BVP and IBI data, harvested from the Empatica E4 device, were transformed into HRV segments lasting 1.5 minutes (90 seconds) each. This conversion was facilitated through a Python script specifically crafted for this purpose. The process commenced with the application of a bandpass Butterworth filter to mitigate noise and continued with successive differences, culminating in a frequency-domain HRV analysis. The subsequent computational steps involved the extraction of time-domain HRV features (RMSSD, SDNN, pNN50) from the filtered BVP readings, followed by the determination of frequency-domain attributes (VLF, LF, HF, LF/HF) using the Welch method for power spectral density estimation.

After the HRV data transformation, a baseline normalization was conducted using the initial 5 minutes of participant recordings as they acquainted themselves with the system. The data was then labelled using a threshold method to gauge the drowsiness levels. The thresholds were established based on the following criteria: RMSSD at 1.5, HF at 140, and LF at -1. The labelling was as follows:

- Label 0: None of the thresholds were met, indicating alertness.

- Label 1: Any one of the thresholds was met, signalling potential drowsiness.

- Label 2: Two or more thresholds were met, suggesting a definitive state of drowsiness or sleepiness.

The EDA data was segmented into 90-second intervals, with various parameters such as mean, standard deviation, and range being extracted. The data underwent baseline normalization and labelling based on specific thresholds: mean (1), standard deviation (0.05), and range (1).

For the eye data, a pupil detection system integrated into the eye tracker software was utilized, ensuring a minimum accuracy rate of 95% in pupil detection for each participant. This system generated a CSV file for each session, documenting metrics like pupil length, height, and area. The pupil area data facilitated the creation of labels indicating the state of the eyes (open or closed). Moreover, PERCLOS was computed at intervals of 30 seconds, while BP and BR were calculated every 60 seconds. The drowsiness levels were determined using a threshold system with the following criteria: PERCLOS (8), BP (30), and BR (20). The labelling scheme was as follows:

- Label 0: No thresholds were met, indicating alertness.

- Label 1: Any one of the thresholds was met, indicating potential drowsiness.

- Label 2: Two or more thresholds were met, indicating a definitive state of drowsiness.

To synchronize with the eye tracker software's recording rate of approximately 130 fps, the labels underwent downsampling to match the video sampling rate of 60 fps. This process entailed the removal of even rows from the CSV file, given that the calculated variables remained constant over time frames of at least 30 seconds, yielding a single value for each 30-second label interval. Lastly, the frames were resized to 64x64 pixels prior to their integration into the model.

3.9   Training Process

The training process for the HRV model consists of several steps. Initially, the dataset is split into training and testing sets using the train_test_split function from scikit-learn. Specifically, 80% of the data is assigned to the training set, represented by X_train and y_train, while the remaining 20% is allocated to the testing set, denoted by X_test and y_test. To prepare the target variable y_train for training, which is currently in one-hot encoded format, it is converted back to class labels using the np.argmax function. Additionally, class weights are computed to handle any class imbalance, based on the distribution of class labels in the training set, employing the class_weight.compute_sample_weight function.

The next step involves defining the Keras model using the Sequential API. The desired architecture is specified by selecting appropriate layers and activation functions. To compile the model, the Adamax optimizer is utilized with a learning rate of 0.002, and the categorical cross-entropy loss function is employed. During the training process, an EarlyStopping callback is used to monitor the validation loss and halt training if there is no improvement for a specified number of epochs (in this case, 5). Additionally, a ModelCheckpoint callback is implemented to save the best model based on validation accuracy. Moving on, the model is trained using the fit function, where the training data (X_train and y_train) is used, and 20% of the training data is reserved for validation (validation_split=0.2). The training process occurs over 100 epochs, with a batch size of 10. The progress of training is displayed (verbose=1), and both the early stopping and model checkpoint callbacks are applied.

As for the training process for the EDA model, the data is split into training and testing sets using the train_test_split function from the sklearn.model_selection module. Similarly, 80% of the data is allocated for training (X_train, y_train), while the remaining

20% is dedicated to testing (X_test, y_test). To ensure compatibility with the expected input shape of the Conv1D layer, the input data is reshaped using the reshape function.

Subsequently, the model is compiled using the model.compile method. This entails specifying the loss function, optimizer, and metrics to be used. In addition, an early stopping callback is set up to monitor the validation loss and halt training if it fails to improve within a specified number of epochs. The training of the model is then carried out using the model.fit function, utilizing the training data (X_train, y_train), and specifying parameters such as the number of epochs, batch size, and validation split. Throughout the training process, the early stopping callback is utilized. Finally, the trained model is evaluated on the test data using the model.evaluate method, enabling the calculation of accuracy.

For the Eye model, the data is also split by 80% for training and 20% for testing. Afterwards, a custom data generator code is utilized to provide the model with batches of frames based on the specified window size (in this case, batches of 1800 frames corresponding to 30 seconds) since the PERCLOS is measured every 30 seconds and BR and BP are calculated every 1 minute. After defining the architecture, the model is compiled using Adamax as the optimizer and categorical cross-entropy as the loss function. The training process begins by calling the fit function on the model. The train_data_generator is used as the input, specifying the number of steps per epoch (determined by the number of available 1800-frame windows from the data generator code) and the total number of epochs. Additionally, callbacks are included for saving the best model based on validation loss (ModelCheckpoint) and stopping early if the validation loss does not improve (EarlyStopping). The verbose parameter is set to 1 to print the training progress during each epoch.

In the development of all three models the division of data was a crucial decision. The exploration of different splits, including a 70/30 split, 60/40 split, and a random split, was conducted to determine the most effective distribution. However, in each case, the 80/20 split consistently provided the most accurate results. Thus, this ratio was chosen for the final model training process, reflecting the balance of maintaining a sufficient training set while allowing for comprehensive model testing.

3.10 Model Evaluation

Following the training process, all models underwent evaluation using common performance metrics. For the HRV and EDA models, accuracy was assessed on the test

dataset. The models were utilized to make predictions on the test dataset, enabling a comparison between the predicted class labels and the actual labels. This comparison led to the generation of a confusion matrix, which provides a comprehensive overview of the model's performance by presenting the counts of true positive, true negative, false positive, and false negative predictions. Additionally, a classification report was generated, offering a detailed evaluation of the model's performance. Precision, recall, and F1-score were calculated to provide a more in-depth analysis of the model's effectiveness.

Upon completion of training for the Eye model, evaluation was conducted using the evaluate function. The test_data_generator was employed as the input, and various metrics were computed, including test accuracy, precision, recall, and loss. Furthermore, a confusion matrix was generated to analyze the predictions made by the model on the test data.

3.11 Hyperparameter Tuning and Model Sensitivity Analysis

The performance and generalizability of each model were enhanced through an intricate process of hyperparameter tuning and model sensitivity analysis. This process involved examining different configurations and parameter values to identify the most optimal model architecture.

In the realm of machine learning, hyperparameters are parameters whose values are defined prior to the initiation of the learning process, and directly govern the learning procedure itself. These, in our context, include, but are not limited to, the number of layers, the quantity of neurons in each layer, activation functions, optimization algorithms, learning rates, and regularization techniques.

Given the complex interactions among these hyperparameters, their optimal values cannot be theoretically deduced, but must instead be ascertained through practical experimentation. In this research, hyperparameter tuning was manually conducted. A variety of configurations were tested on the data, with their performance assessed based on pertinent evaluation metrics such as accuracy, precision, recall, and the F1 score.

For example, within the HRV model, various numbers of layers and neurons were tested, leading to the observation that the optimal architecture comprised five hidden layers with 64, 32, 16, 8, and 4 neurons respectively. Similarly, the hyperbolic tangent (tanh) function was chosen for the input layer owing to its superior performance, while

the rectified linear unit (ReLU) was selected for the hidden layers in terms of activation functions.

In the EDA model, a sequence of Conv1D layers with differing filter sizes and quantities were examined. The conclusion was drawn that a filter size of 3 with 64 filters produced the most beneficial results. Additionally, the dropout rate was adjusted, revealing the optimal value to be 0.5

For the Eye model, the number of LSTM layers and the quantity of units in each layer were adjusted. After thorough testing, it was decided to employ three LSTM layers, each with 128 units. The dropout rates varied, with 0.25 being pinpointed as the optimal value for both the encoder and intermediate LSTM layers, and 0.2 for the decoder LSTM layer.

The hyperparameter tuning and model sensitivity analysis process assured that the constructed models were robust and performed optimally on unseen data. It is important to note that the chosen configurations were selected based not only on their immediate performance but also their ability to offer effective generalization and avoid overfitting. This ensures that the models are well-suited for practical deployment in real-world scenarios.

Chapter 4

Analysis & Results

The research presented in this thesis can potentially be a cornerstone for the development of robust and effective drowsiness detection systems. It addresses several critical aspects that have been understudied in previous research. These include exploring the complementary nature of different types of physiological data in drowsiness detection, and utilizing specific machine learning models tailored to each data source to extract relevant features and patterns.

The expected outcome of this study is threefold. First, it is anticipated that the HRV-based, EDA-based, and eye data-based models will each exhibit strong performance in detecting drowsiness, reflecting their ability to capture distinct aspects of drowsiness relevant to each data source. Second, by integrating the outputs of the three models, it is expected that the overall drowsiness detection system will achieve higher accuracy and reliability compared to systems based on a single data source or model. This reflects the complementary nature of the three types of data used, providing a comprehensive picture of drowsiness in drivers.

Finally, the thesis aims to provide insights and recommendations for future research and practical applications in the field of drowsiness detection. These include the potential of using a multimodal approach for more accurate and reliable drowsiness detection, the importance of balancing the dataset for training machine learning models, the exploration of alternative machine learning models and architectures, the need for reliable and non-intrusive data collection methods, and the necessity of validating the developed models under real-world driving conditions.

Given the critical role that drowsiness detection plays in enhancing road safety and preventing fatigue-related accidents, this thesis is of great importance. It is expected to

make significant contributions to the existing body of knowledge in the field, and provide a foundation for future studies. By advancing the understanding and technological capabilities of drowsiness detection, this research ultimately aims to help create safer driving environments and save lives.

4.1   Overview of the Data

The data used in the study was collected through various devices and sensors, including the Carla Simulator for driving scenarios, Ergoneers Dikablis Glasses 3 eye-tracker for eye data, and Empatica E4 wristband for HRV and EDA measurements.

1. HRV Data:

   - HRV data was obtained from the Empatica E4 wristband.
   - Variables such as Blood Volume Pressure (BVP) and Interbeat Interval (IBI) were recorded.
   - Custom Python code was used to transform the variables into HRV segments of 1.5 minutes (90 seconds).
   - Time and frequency variables, including RMSSD, SDNN, pNN50, VLF, LF, HF, and LF/HF ratio, were extracted from the HRV segments.
   - Baseline normalization was performed using the first 5 minutes of recording.
   - Labels indicating drowsiness were assigned based on specific threshold values for RMSSD, HF, and LF.

2. EDA Data:

   - EDA data was collected using the Empatica E4 wristband.
   - The data was divided into 90-second intervals.
   - Various EDA variables, such as mean, standard deviation, minimum value, maximum value, range, number of peaks, mean peak amplitude, and mean peak duration, were extracted.
   - Baseline normalization was performed using the first 5 minutes of recording.
   - Labels indicating drowsiness were assigned based on specific threshold values for mean, standard deviation, and range.

3. Eye Data:

- Eye data was captured using the Ergoneers Dikablis Glasses 3 eye-tracker.
- Pupil detection system from the eye tracker software was used to detect the pupil of each participant with at least 95% accuracy.
- CSV files were generated, containing values such as pupil length, height, and area.
- Labels indicating eye status (open or closed) were created based on the pupil area.
- PERCLOS, BP, and BR were calculated at specific intervals.
- Labels indicating drowsiness were assigned based on threshold values for PERCLOS, BP, and BR.
- Downsampling was performed to match the eye tracker software's recording rate, resulting in labels at 30-second intervals.
- The frames were downscaled to 64x64 pixels before feeding them into the model.

4.2 Machine Learning Models

4.2.1 Model 1: HRV-Based Model

The HRV model leverages features extracted from heart rate variability (HRV) data, specifically RMSSD, HF, and LF, to categorize driver drowsiness. The model's architecture consists of multiple dense layers with dropout regularization and L2 kernel regularization implemented to counteract overfitting.

The HRV-Based Model was trained and evaluated using a dataset that consists of 30 CSV files, yielding a combined total of 867 samples, consisting of 669 data points with label 0, 165 with label 1, and 33 with label 2. The dataset underwent preprocessing, where the input features (RMSSD, HF, and LF) and the corresponding output labels (drowsy, alert, etc.) were separated. To facilitate categorical classification, the output labels were subjected to one-hot encoding.

Upon evaluation, the HRV-Based Model achieved a test set accuracy of 98.28%, demonstrating its efficiency in accurately classifying driver drowsiness (refer to Figure 4.1 and Figure 4.2). Various performance metrics, including precision, recall, and F1-score, were utilized to evaluate the model's predictions. The classification report presents

43

a detailed analysis of the model's performance for each class. The precision, recall, and F1-score values for different classes (drowsy, alert, etc.) illustrate the model's ability to accurately classify instances. Across all classes, the model demonstrates high precision, recall, and F1-score, achieving a score of 98% (refer to Figure 4.3).
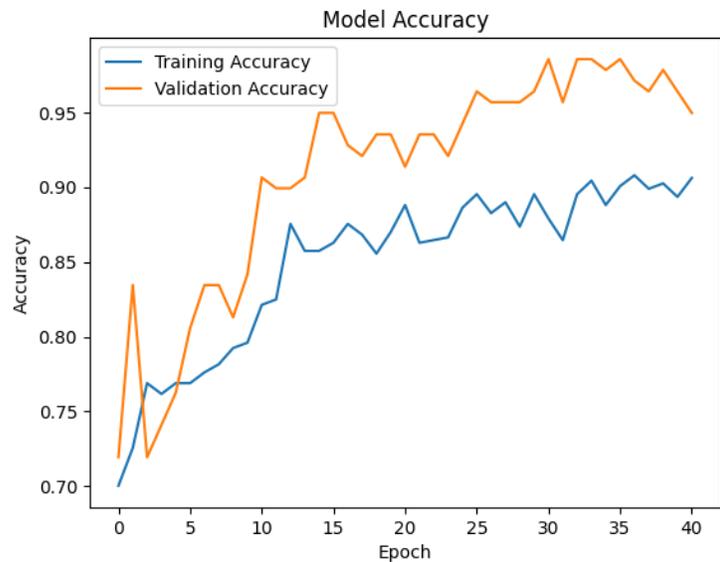

Figure 4.1: Accuracy for the HRV-Model.
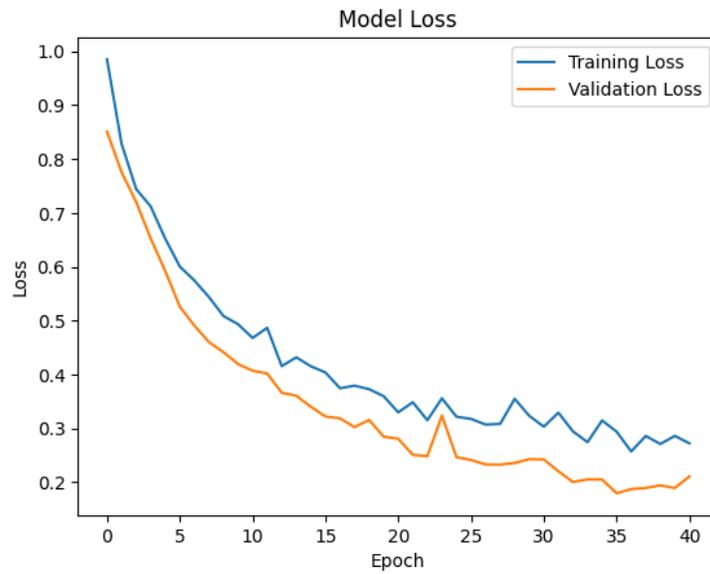
Figure 4.2: Loss for the HRV-Model.
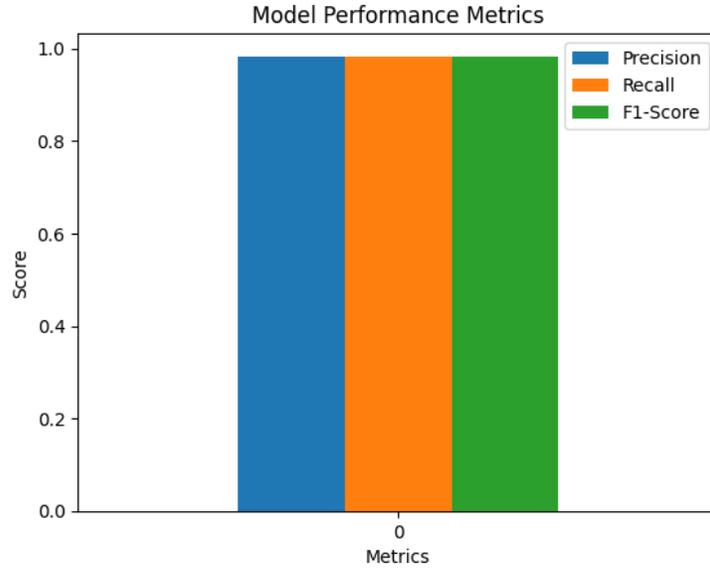
Figure 4.3: Performance for the HRV-Model.

The confusion matrix further illustrates the model's performance by showing the

distribution of actual and predicted labels for each class (refer to Figure 4.4). The majority of instances were correctly classified, leading to high accuracy and validating the model's efficiency in detecting driver drowsiness.
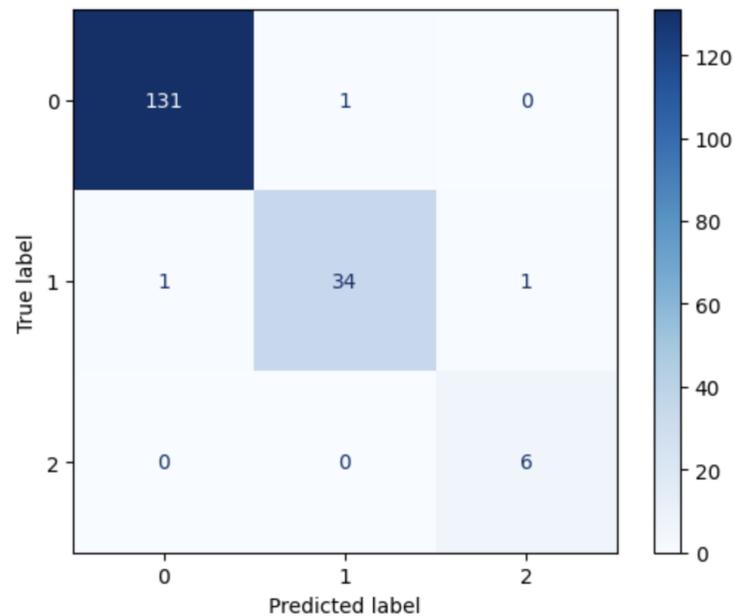
Figure 4.4: Confusion Matrix for the HRV-Model.

The precision score implies a low false-positive rate, indicating that the model displays high accuracy when predicting driver drowsiness or alertness. The recall score denotes a low false-negative rate, showcasing the model's capacity to identify instances of drowsiness in drivers. The F1-score, which accounts for both precision and recall, provides a balanced measure of the model's overall performance.

4.2.2 Model 2: EDA-Based Model

The EDA-Based Model leverages features extracted from Electrodermal Activity (EDA) data to categorize driver drowsiness. The model's architecture is composed of a 1-Dimensional Convolutional Neural Network (1D-CNN), followed by fully connected layers. Dropout regularization and L2 kernel regularization are applied to mitigate overfitting, as illustrated in Figure 4.5 and 4.6.
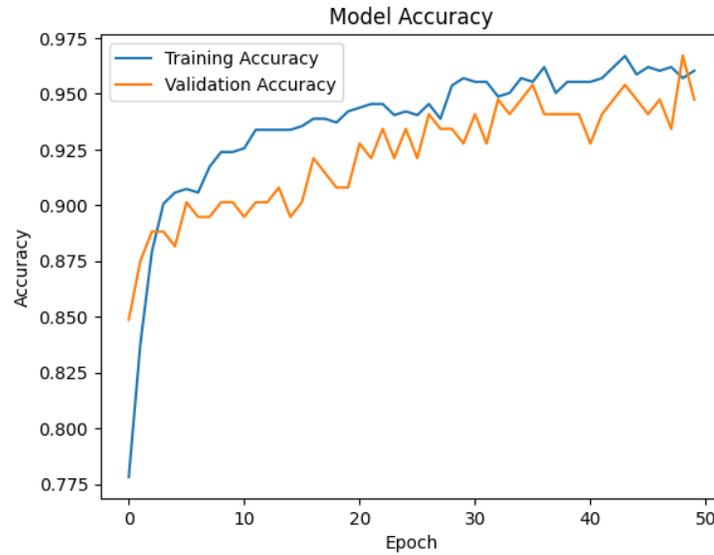
Figure 4.5: Accuracy obtained for the EDA-Model.

This model was trained and evaluated using a dataset consisting of 30 CSV files, yielding a combined total of 946 samples, consisting of 690 data points with label 0, 147 with label 1, and 109 with label 2. The dataset underwent preprocessing, during which input features (mean, standard deviation, and range) and corresponding output labels were separated. Output labels were encoded using one-hot encoding to facilitate categorical classification.
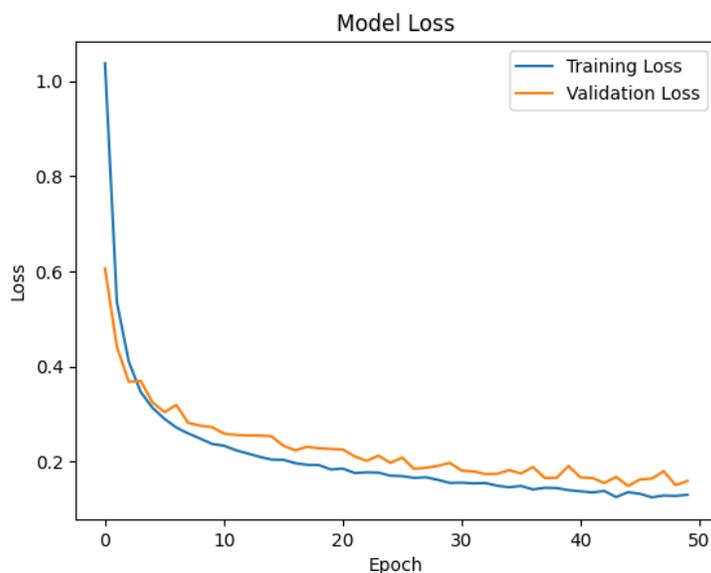
Figure 4.6: Loss obtained for the EDA-Model.

Upon evaluation, the EDA-Based Model achieved a test set accuracy of 96.32%, thereby demonstrating its capability to accurately classify driver drowsiness. Various performance metrics, including precision, recall, and the F1-score, were utilized to assess the model's predictions. The classification report provides a comprehensive analysis of the model's performance for each class. The precision, recall, and F1-score values shed light on the model's ability to accurately distinguish between instances of drowsiness and alertness. The EDA-Based Model exhibits high precision, recall, and F1-score across all classes, achieving an overall weighted average of 0.96, refers to Figure 4.7.
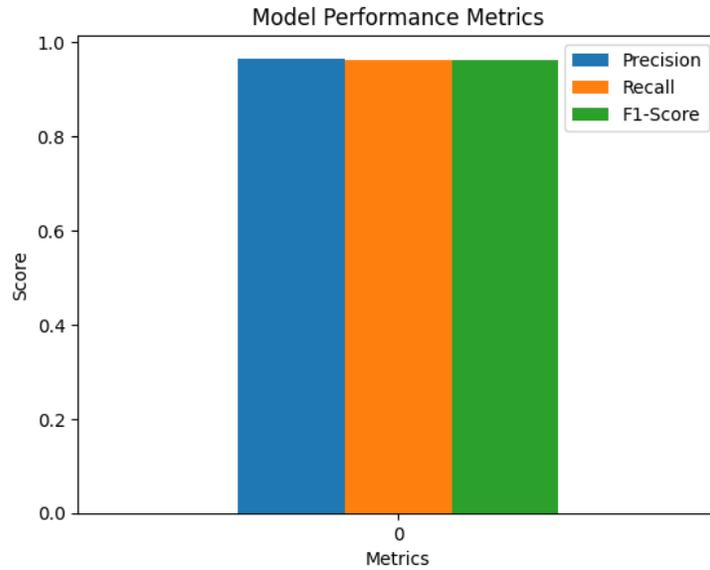
Figure 4.7: Performance for the EDA-Model.

The confusion matrix further portrays the model's performance by highlighting the distribution of actual and predicted labels for each class. The majority of instances were correctly classified, leading to high accuracy and validating the model's effectiveness in detecting driver drowsiness , detailed in Figure 4.8. The precision score implies a low false-positive rate, signifying the model's high accuracy in predicting drowsiness or alertness. The recall score denotes a low false-negative rate, indicating the model's capability to identify instances of drowsiness in drivers. The F1-score, which takes both precision and recall into account, provides a balanced measure of the model's overall performance.
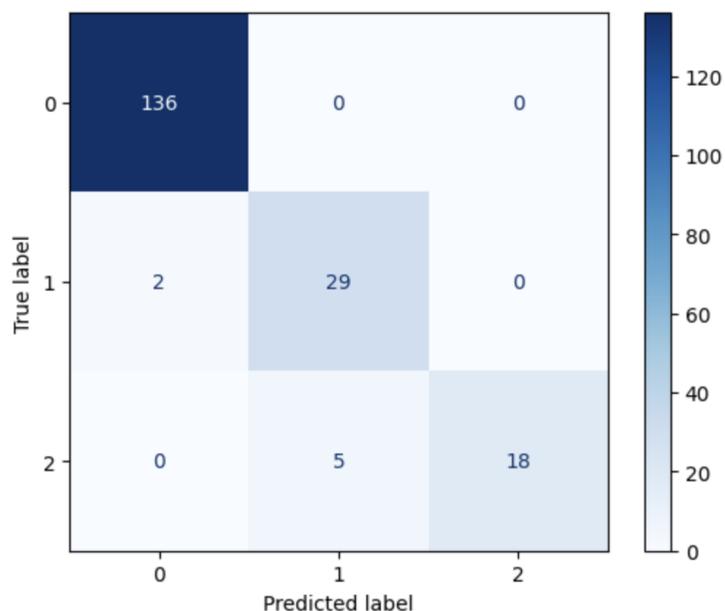
Figure 4.8: Confusion Matrix for the EDA-Model.

4.2.3 Model 3: PERCLOS, Blink Percentage, and Blink Rate based Model

The PERCLOS, BP, and BR based Eye Model leverages key ocular features to classify driver drowsiness. This architecture is structured as a Convolutional Recurrent Neural Network (CRNN). Both left and right eye sequences are independently processed through a succession of time-distributed convolutional, normalization, pooling, and dropout layers. These sequences are subsequently flattened and merged using a concatenation layer. A trio of stacked LSTM layers processes the concatenated output with dropout, and this output is then employed as the initial state for the decoder LSTM. The final outputs are generated by applying a softmax activation function to a dense layer that is connected to the decoder outputs.

A comprehensive dataset, including a combined total of 3,951,000 frames for training and 973,800 frames for testing collected from 30 participants, was utilized for training and evaluation. The preprocessing phase encompassed pupil recognition identification optimization, PERCLOS, BP, and BR calculations, as well as drowsiness labeling. Training was conducted using the Adamax optimizer, which started with an initial learning rate of

50

0.01, that was subsequently reduced gradually. The model employed Categorical Cross-Entropy as the loss function, and the primary evaluation metrics were accuracy, precision, and recall. The model with the lowest validation loss was saved during the training, and early stopping was implemented if the validation loss failed to improve after five epochs.

Upon completion of 100 epochs, the model accomplished an approximate accuracy, precision, and recall of 90% on the test set, refer to Figure 4.9. The test loss was documented at about 50%. However, despite these notable figures, the model did not correctly predict cases with labels 1 or 2 according to the confusion matrix, as shown in Figure 4.10. This outcome could be attributed to a highly unbalanced dataset, with 1416 windows of 1800 frames per eye labeled as 0 (active), 396 windows labeled as 1 (possible drowsiness), and 116 windows labeled as 2 (definitely drowsy). The highly unbalanced dataset and the results obtained underscore a significant weakness of the model: its inability to predict accurately all classes of drowsiness.
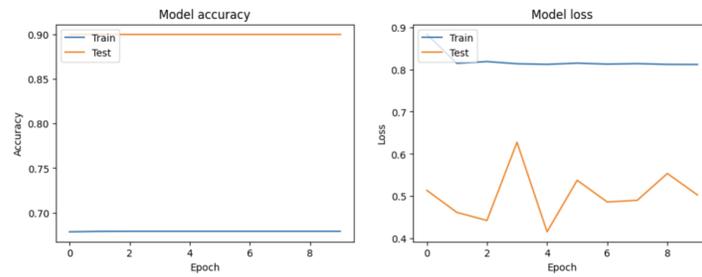
Figure 4.9: Accuracy and Lost results for the Eye-Model.
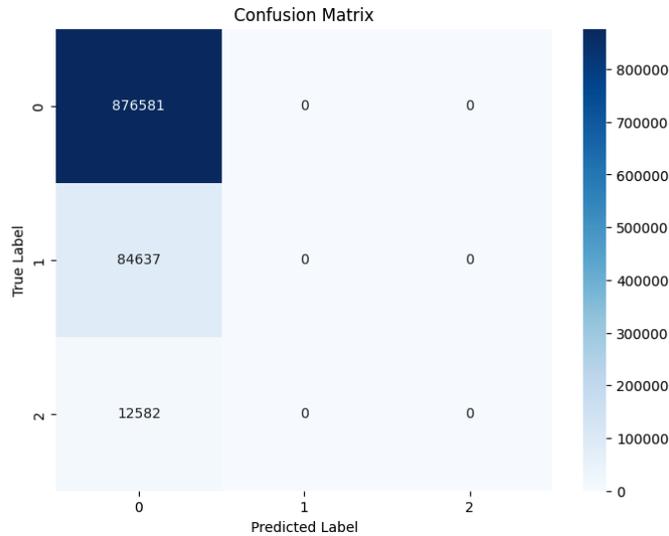
Figure 4.10: Confusion Matrix for the Eye-Model.

4.3 Fatigue Assessment Scale (FAS) Analysis

This section presents an analysis of the results obtained from the Fatigue Assessment Scale (FAS), administered before and after data collection, referred to as FAS 1 and FAS 2, respectively.

A thorough understanding of the fatigue levels reported by the participants was pursued through the calculation of the descriptive statistics for both FAS 1 and FAS 2. The mean fatigue scores were recorded as 25.7 for FAS 1 and 25.8 for FAS 2, with standard deviations of 4.3003608509375 and 4.58182315768252, respectively. The maximum fatigue score observed in FAS 1 was 35, while in FAS 2, it reached 38. Conversely, the minimum fatigue score was 17 for FAS 1 and 20 for FAS 2.

Alterations in fatigue scores between FAS 1 and FAS 2 were examined. Out of all the participants, 10 individuals showed an increase in their fatigue scores, 13 reported a decrease, and 7 showed no change, having equal scores in both assessments.

To assess whether the changes in fatigue scores were statistically significant, the Wilcoxon Signed-Rank Test was employed [5]. This non-parametric statistical hypothesis test is utilized when comparing two related samples or repeated measurements on a single sample to determine whether their population mean ranks differ. It serves as an

alternative to the paired Student's t-test when the normal distribution of the population cannot be assumed.

A null hypothesis (H0) and alternative hypothesis (H1) were set for the Wilcoxon Signed-Rank Test. The null hypothesis assumed that there was no significant difference in fatigue levels before and after performing the driving simulator task. Conversely, the alternative hypothesis assumed a significant difference in fatigue levels.

The Wilcoxon Signed-Rank Test resulted in a statistic of 134.0 and a p-value of approximately 0.90. The high p-value led to a failure to reject the null hypothesis, suggesting no statistically significant difference in the median scores of FAS 1 and FAS 2. Therefore, the variations in individual scores do not provide strong evidence that the driving simulator task had a significant effect on the participants' drowsiness as measured by FAS.

The distribution of fatigue levels based on the FAS scores was further analyzed. In FAS 1, one participant reported no fatigue, while 27 participants fell into the category of mild-to-moderate fatigue. Additionally, two participants indicated severe fatigue. In FAS 2, the number of participants reporting no fatigue rose to four, while 24 participants fell into the category of mild-to-moderate fatigue. However, the number of participants experiencing severe fatigue remained constant at two.

The descriptive analysis of the FAS scores revealed some interesting insights into the participants' fatigue levels. The mean fatigue scores remained relatively stable between FAS 1 and FAS 2, with a negligible difference of 0.1 between the two assessments. The standard deviation values suggest a moderate variability in the reported fatigue levels for both FAS 1 and FAS 2.

A minor shift was observed in the distribution of fatigue levels between the two assessments. FAS 2 showed an increase in the number of participants reporting no fatigue and a decrease in those with mild-to-moderate fatigue, while the number of participants experiencing severe fatigue stayed constant.The findings suggest that the data collection period may not have significantly impacted the overall fatigue levels of the participants on average.

Chapter 5

Conclusions

5.1 Discussion of the results

In conclusion, the evaluation of the HRV-Based Model, the EDA-Based Model, and the Eye-Based Model underscores varying levels of performance in detecting and classifying driver drowsiness.

The HRV-Based Model demonstrated an impressive accuracy rate of 98.28% in effectively classifying driver drowsiness through heart rate variability features. Its remarkable performance is further accentuated by exceptional precision, recall, and F1-score results. Such results emphasize its real-world applicability in distinguishing between alert and drowsy drivers.

Similarly, the EDA-Based Model exhibited substantial performance, achieving an accuracy of 96.32%. However, this model revealed certain limitations. While proficient in predicting instances of non-drowsiness (class 0) and pronounced drowsiness (class 2), it faced challenges discerning instances of mild drowsiness (class 1). This challenge is possibly due to the complexities of distinguishing between mild drowsiness and non-drowsiness based solely on EDA features.

Conversely, the Eye-Based Model demonstrated a promising overall performance, achieving a commendable accuracy of 90%. Nevertheless, this model displayed shortcomings in accurately identifying instances of drowsiness. This deficiency became particularly evident during the confusion matrix analysis, wherein the model encountered difficulties in correctly identifying positive cases of drowsiness. Despite its capability to recognize the majority class (non-drowsy cases), the model struggled with the minority class (drowsy

cases), a common predicament in imbalanced datasets. Despite employing various strategies to address this imbalance, substantial improvements were elusive.

It is crucial to recognize that the datasets utilized for training these models were characterized by a significant disparity in class distribution, favoring non-drowsy states. This skewed distribution significantly impacted the models' achieved accuracies, as they exhibited a propensity to predict the dominant class due to this disproportionate representation. Furthermore, the overall state of drowsiness across the datasets was consistently low, further contributing to the imbalanced nature of the datasets.

The scarcity of minority classes (labels 1 and 2) indicates that the data collection timeframe might not have been optimal for inducing drowsiness in participants. Given the study's duration (45 minutes) and the observation that participants primarily began experiencing drowsiness towards the study's conclusion or afterward, an extended data collection period could potentially yield a more balanced dataset. Nonetheless, the model's elevated accuracy, precision, and recall values for specific classes, in conjunction with the proficient feature extraction capacity of CNNs and LSTMs, attest to its robust performance.

Regarding the Fatigue Assessment Scale (FAS), the consistent average fatigue scores derived from both FAS 1 and FAS 2, alongside moderate score variability, suggest a relatively stable level of fatigue among participants throughout the data collection period. Supported by the Wilcoxon Signed-Rank Test, these observations imply no statistically significant difference in participants' fatigue levels before and after the driving simulator task. While this suggests that the task itself might not have significantly impacted overall fatigue levels, the slight disparities between FAS 1 and FAS 2 underscore the importance of considering external influences and individual factors that might have influenced reported fatigue levels.

While the evaluated models present notable accuracy in detecting driver drowsiness based on physiological indicators, it's essential to acknowledge potential external factors that may influence these measures. Factors such as fatigue, stress, workload, and mind wandering, which were not controlled or measured in this study, could affect physiological responses. Recognizing these variables is crucial for further refinement and applicability of the models in real-world scenarios.

Employing multiple machine learning models for driver drowsiness detection, each tailored to distinct physiological indicators, constitutes a robust approach that comprehensively addresses the intricate and multifaceted nature of driver drowsiness. These findings offer invaluable insights and guidance for future research endeavors aimed at enhancing drowsiness detection systems, thereby advancing road safety. The exploration and validation of these models pave the way for the development of a comprehensive and resilient

drowsiness detection system, which holds the potential to significantly mitigate drowsy driving-related accidents and enhance driver safety.

5.2   Limitations

Despite the efforts made to conduct a comprehensive study, several limitations were encountered during the data collection and analysis process. These limitations may have implications for the interpretation of the results and the generalizability of the findings. The main limitations are as follows:

1. Reliability of Eye-Tracking Equipment: The reliability of the Ergoneers Dikablis Glasses 3 eye-tracker used for eye movement data collection presented a significant challenge. Repeated system malfunctions and program crashes during data collection sessions resulted in lost data and incomplete recordings. These technical issues could have compromised the quality and comprehensiveness of the eye movement data, thereby potentially influencing the accuracy and reliability of the eye movement-based model.

2. Time Constraints for Data Collection: The study was further limited by the time constraints imposed by the ethics board. The approved data collection window, between noon and 5 pm, may not fully represent the range of fatigue experiences drivers could encounter at different times of the day, especially during evening or nighttime driving when drowsiness can be a critical issue. This limitation potentially impacts the generalizability of the study findings.

3. Data Collection Session Length: The duration of the data collection sessions, set at 45 minutes, may not have been adequate to capture the full range of fatigue experiences, particularly given that some participants reported feeling drowsy or experiencing fatigue changes only towards the end of the session. This limitation could have led to an underrepresentation of more pronounced fatigue states (labels 1 and 2) in the collected data, thereby affecting the ability of the models to accurately detect and predict these specific fatigue states.

4. Influence of External Factors on Physiological Measures: While the evaluated models demonstrate notable accuracy in detecting driver drowsiness based on physiological indicators, external factors that were neither controlled nor measured in this study, such as fatigue, stress, workload, and mind wandering, might have influenced these

measures. The presence of these unmeasured factors could affect the models' physiological responses, introducing additional variability and potential error in drowsiness classification.

5.3 Comparison of prior work with the current study

In the field of drowsiness detection, a detailed review of earlier studies reveals certain limitations that the current research aims to address. A primary limitation observed in previous works was their reliance on existing data. Instead of generating new data, many of these studies utilized available datasets. Such an approach might introduce bias or reduce the accuracy of their models due to a lack of control over the data collection conditions.

Additionally, much of the earlier research focused on analyzing individual physiological variables. An integrated assessment that considered multiple variables simultaneously was often missing. In this context, the current study stands out by concurrently analyzing a unique combination of physiological variables: HRV, PERCLOS, blink rate, blink percentage, and EDA. To the best of our knowledge, no other study has explored this specific combination of variables, making this research pioneering in its approach.

The potential of Electrodermal Activity (EDA) as a predictor for drowsiness detection was not extensively explored in prior works, despite its evident significance. Furthermore, the specific driving scenarios employed in many of these studies were often not clearly defined, leading to potential gaps in understanding the effects of monotonous versus non-monotonous driving.

In contrast, the current study introduces several advancements. A distinguishing feature of this research is the development of three separate models to detect drowsiness based on HRV, EDA, and eye-based data (including PERCLOS, blink rate, and blink percentage). This multi-model approach adds robustness to the system, ensuring that it doesn't rely solely on one type of variable, thereby enhancing its reliability and accuracy.

Another significant aspect of this research is the use of a custom-developed simulator. This tool, designed with controlled variables and specific scenarios, ensures enhanced accuracy, relevance, and consistency in data collection. By providing a more complete perspective on driver fatigue through its integrated approach, the study is expected to yield deeper insights than investigations limited to single-variable analyses.

Moreover, this study offers insights into EDA, presenting new findings on the differences between monotonous and non-monotonous subjects. This exploration emphasizes

EDA's potential as a reliable indicator for drowsiness detection. By categorizing participants based on two distinct driving scenarios, the research provides a detailed understanding of how environmental conditions might influence drowsiness.

In conclusion, while earlier studies have provided a foundational understanding in drowsiness detection, the current research builds upon this foundation. By addressing the limitations of previous works, leveraging the strengths of multiple physiological indicators, introducing a novel combination of variables, and developing multiple models, this study contributes significantly to the development of more accurate and reliable drowsiness detection systems in the future.

5.4   Future work and improvements

In the complex field of drowsiness detection, previous research efforts have primarily been characterized by a predominant reliance on pre-existing data, possibly initiating biases or hindering the precision of the developed models due to a lack of stringent control over the conditions of data acquisition. This study seeks to transcend these limitations, pioneering a methodical approach that integrates a unique constellation of physiological variables: HRV, PERCLOS, blink rate, blink percentage, and EDA, a combination which, to our knowledge, remains unexplored, marking this venture as groundbreaking.

Furthermore, the Electrodermal Activity (EDA) as a credible predictive marker for drowsiness detection has remained relatively under-explored, despite showcasing substantial potential. Additionally, the driving scenarios delineated in many preceding studies often lacked explicit clarity, thereby leaving a gap in comprehending the dynamics of monotonous versus non-monotonous driving.

This study inaugurates a series of advancements in the field. Notably, it introduces the development of three distinct models focussed on HRV, EDA, and eye-based data (comprising PERCLOS, blink rate, and blink percentage) to delineate drowsiness, enhancing the system's robustness and thereby amplifying its reliability and accuracy. This multiple-model approach aids in circumventing an over-dependence on a singular variable type.

Furthermore, this research utilizes a specially developed simulator, designed to encapsulate controlled variables and distinct scenarios, fostering heightened accuracy, relevance, and consistency in data acquisition. This tool promises to impart a more comprehensive perspective on driver fatigue, transcending the insights garnered from single-variable analysis studies prevalent before.

Moreover, the study delves deeper into the potentials of EDA, uncovering novel insights into the differences between monotonous and non-monotonous subjects, thereby accentuating EDA's reliability as an index for drowsiness detection. By distinguishing participants based on two specific driving environments, the study grants a nuanced understanding of how environmental factors could potentially modulate drowsiness.

In summation, this study not only builds upon the foundational understanding established by prior studies but also ventures further, addressing known deficiencies and leveraging the strengths of multiple physiological markers. The incorporation of a novel variable amalgamation, coupled with the development of multiple models, stands to significantly influence the progression towards more accurate and dependable drowsiness detection systems in the forthcoming years.

References

[1] Automated classification system for drowsiness detection using convolutional neural network and electroencephalogram - Balam - 2021 - IET Intelligent Transport Systems - Wiley Online Library.

[2] A combined convolutional and recurrent neural network for enhanced glaucoma detection | Scientific Reports.

[3] Drowsiness Information | Mount Sinai - New York.

[4] Trends and Future Prospects of the Drowsiness Detection and Estimation Technology.

[5] Wilcoxon Signed Ranks Test - an overview | ScienceDirect Topics.

[6] Law Document English View, July 2014.

[7] 1D convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing*, 151:107398, April 2021. Publisher: Academic Press.

[8] Kawser Ahammed and Mosabber Ahmed. *Multivariate Multiscale Entropy: An Approach to Estimating Vigilance of Driver*. May 2020.

[9] Mehmet Akin, Muhammed B. Kurt, Necmettin Sezgin, and Muhittin Bayram. Estimating vigilance level by using EEG and EMG signals. *Neural Computing and Applications*, 17(3):227–236, June 2008.

[10] Yaman Albadawi, Maen Takruri, and Mohammed Awad. A Review of Recent Developments in Driver Drowsiness Detection Systems. *Sensors*, 22(5):2069, January 2022. Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

[11] Wolfram Boucsein. *Electrodermal Activity*. Springer US, Boston, MA, 2012.

[12] Philipp P. Caffier, Udo Erdmann, and Peter Ullsperger. Experimental evaluation of eye-blink parameters as a drowsiness measure. *European Journal of Applied Physiology*, 89(3):319–325, May 2003.

[13] Aswathi CD, Nimmy Ann Mathew, K S Riyas, and Renu Jose. Comparison Of Machine Learning Algorithms For Heart Rate Variability Based Driver Drowsiness Detection. In *2021 2nd Global Conference for Advancement in Technology (GCAT)*, pages 1–7, October 2021.

[14] Siwar Chaabene, Bassem Bouaziz, Amal Boudaya, Anita Hökelmann, Achraf Ammar, and Lotfi Chaari. Convolutional Neural Network for Drowsiness Detection Using EEG Signals. *Sensors (Basel, Switzerland)*, 21(5):1734, March 2021.

[15] Robert Chen-Hao Chang, Chia-Yu Wang, Wei-Ting Chen, and Cheng-Di Chiu. Drowsiness Detection System Based on PERCLOS and Facial Physiological Signal. *Sensors (Basel, Switzerland)*, 22(14):5380, July 2022.

[16] Hanna Chouchane, Hiroki Nakamura, Kenji Sato, Jacobo Antona-Makoshi, Genya Abe, and Makoto Itoh. Identifying the Out Of The Loop phenomenon during driving automation using spontaneous gaze behavior. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 66(1):305–309, September 2022. Publisher: SAGE Publications Inc.

[17] Kwok Tai Chui, Miltiadis D. Lytras, and Ryan Wen Liu. A Generic Design of Driver Drowsiness and Stress Recognition Using MOGA Optimized Deep MKL-SVM. *Sensors*, 20(5):1474, January 2020. Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

[18] Brenda L. Den Oudsten, Guus L. Van Heck, Alida F. W. Van der Steeg, Jan A. Roukema, and Jolanda De Vries. Second operation is not related to psychological outcome in breast cancer patients. *International Journal of Cancer*, 126(6):1487–1493, March 2010.

[19] Wanghua Deng and Ruoxue Wu. Real-Time Driver-Drowsiness Detection System Using Facial Features. *IEEE Access*, 7:118727–118738, 2019. Conference Name: IEEE Access.

[20] Günther Deuschl and Christof Goddemeier. Spontaneous and reflex activity of facial muscles in dystonia, Parkinson's disease, and in normal subjects. *Journal of Neurology, Neurosurgery & Psychiatry*, 64(3):320–324, March 1998. Publisher: BMJ Publishing Group Ltd Section: Paper.

[21] P. Ebrahimbabaie Varnosfaderani and J. G. Verly. Geometric Brownian motion (GBM) random process model appears to be an excellent choice for modeling realizations of perclos signals. *Sleep Medicine*, 40:e86, December 2017.

[22] Koichi Fujiwara, Erika Abe, Keisuke Kamata, Chikao Nakayama, Yoko Suzuki, Toshitaka Yamakawa, Toshihiro Hiraoka, Manabu Kano, Yukiyoshi Sumi, Fumi Masuda, Masahiro Matsuo, and Hiroshi Kadotani. Heart Rate Variability-Based Driver Drowsiness Detection and Its Validation With EEG. *IEEE Transactions on Biomedical Engineering*, 66(6):1769–1778, June 2019. Conference Name: IEEE Transactions on Biomedical Engineering.

[23] Heidrun Gehlen, Johanna Loschelder, Roswitha Merle, and Maike Walther. Evaluation of Stress Response under a Standard Euthanasia Protocol in Horses Using Analysis of Heart Rate Variability. *Animals*, 10(3):485, March 2020. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

[24] Reza Ghoddoosian, Marnim Galib, and Vassilis Athitsos. A Realistic Dataset and Baseline Temporal Model for Early Drowsiness Detection. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 178–187, June 2019. ISSN: 2160-7516.

[25] Namni Goel, Hengyi Rao, Jeffrey Durmer, and David Dinges. Neurocognitive Consequences of Sleep Deprivation. *Seminars in Neurology*, 29(04):320–339, September 2009.

[26] John F Golding. Motion sickness susceptibility questionnaire revised and its relationship to other forms of sickness. *Brain Research Bulletin*, 47(5):507–516, November 1998.

[27] Sharat Gupta and Shallu Mittal. Yawning and its physiological significance. *International Journal of Applied and Basic Medical Research*, 3(1):11–15, 2013.

[28] Hao Han, Kejie Li, and Yi Li. Monitoring Driving in a Monotonous Environment: Classification and Recognition of Driving Fatigue Based on Long Short-Term Memory Network. *Journal of Advanced Transportation*, 2022:e6897781, March 2022. Publisher: Hindawi.

[29] E. Hoddes, V. Zarcone, H. Smythe, R. Phillips, and W. C. Dement. Quantification of sleepiness: a new approach. *Psychophysiology*, 10(4):431–436, July 1973.

[30] Rohit Hooda, Vedant Joshi, and Manan Shah. A comprehensive review of approaches to detect fatigue using machine learning techniques. *Chronic Diseases and Translational Medicine*, August 2021.

[31] Yanwen Huang and Yuanchang Deng. A Hybrid Model Utilizing Principal Component Analysis and Artificial Neural Networks for Driving Drowsiness Detection. *Applied Sciences*, 12(12):6007, January 2022. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.

[32] Syem Ishaque, Naimul Khan, and Sri Krishnan. Trends in Heart-Rate Variability Signal Analysis. *Frontiers in Digital Health*, 3, 2021.

[33] Melinda L. Jackson, Susan Raj, Rodney J. Croft, Amie C. Hayley, Luke A. Downey, Gerard A. Kennedy, and Mark E. Howard. Slow eyelid closure as a measure of driver drowsiness and its relationship to performance. *Traffic Injury Prevention*, 17(3):251–257, 2016.

[34] Seok-Woo Jang and Byeongtae Ahn. Implementation of Detection System for Drowsy Driving Prevention Using Image Recognition and IoT. *Sustainability*, 12(7):3037, January 2020. Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.

[35] M. W. Johns. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep*, 14(6):540–545, December 1991.

[36] D. P. Joseph and S. S. Miller. Apical and basal membrane ion transport mechanisms in bovine retinal pigment epithelium. *The Journal of Physiology*, 435:439–463, April 1991.

[37] Sang-Joong Jung, Heung-Sub Shin, and Wan-Young Chung. Driver fatigue and drowsiness monitoring system with embedded electrocardiogram sensor on steering wheel. *IET Intelligent Transport Systems*, 8(1):43–50, 2014. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1049/iet-its.2012.0032.

[38] Muhammad Kamazlan, Wan Khairunizam, Abdul Hafiz Halin, M. Rudzuan M. Nor, Azian Abdullah, and Norrima Mokhtar. Investigation of A Real-Time Driver Eye-Closeness for the Application of Drowsiness Detection. *Proceedings of International Conference on Artificial Life and Robotics*, 26:747–753, January 2021.

[39] Kati Blake. Everything You Need to Know About Drowsiness, October 2015.

[40] Muhammad Qasim Khan and Sukhan Lee. A Comprehensive Survey of Driving Monitoring and Assistance Systems. *Sensors*, 19(11):2574, January 2019. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

[41] Andrei Khorovets. What Is An Electrocardiogram (ECG)? *The Internet Journal of Advanced Nursing Practice*, 4(2), December 1999. Publisher: Internet Scientific Publications.

[42] Samuel Kim. Detecting Fatigue Driving Through PERCLOS: A Review.

[43] A.K. Kokonozi, E.M. Michail, I.C. Chouvarda, and N.M. Maglaveras. A study of heart rate and brain system complexity and their interaction in sleep-deprived subjects. In *2008 Computers in Cardiology*, pages 969–971, Bologna, Italy, September 2008. IEEE.

[44] Lawrence C. Barr, C. Y. David Yang, Richard J. Hanowski, and Rebecca Olson. Assessment of Driver Fatigue, Distraction, and Performance in a Naturalistic Setting, 2005.

[45] Kening Li, Yunbo Gong, and Ziliang Ren. A Fatigue Driving Detection Algorithm Based on Facial Multi-Feature Fusion. *IEEE Access*, 8:101244–101259, 2020. Conference Name: IEEE Access.

[46] D. Malathi, JD Dorathi Jayaseeli, S. Madhuri, and K. Senthilkumar. Electrodermal Activity Based Wearable Device for Drowsy Drivers. *Journal of Physics: Conference Series*, 1000(1):012048, April 2018. Publisher: IOP Publishing.

[47] Mayamiko Mbamba, Thokozani Mzumara, Precious Chisale, and Joseph Afonne. The distribution of blinkrate among Malawian young adults: a cross-sectional study. *Scientific Reports*, 13(1):2039, February 2023. Number: 1 Publisher: Nature Publishing Group.

[48] Helen J. Michielsen, Jolanda De Vries, and Guus L. Van Heck. Psychometric qualities of a brief self-rated fatigue measure: The Fatigue Assessment Scale. *Journal of Psychosomatic Research*, 54(4):345–352, April 2003.

[49] Imran Khan Mirani, Chen Tianhua, Malak Abid Ali Khan, Syed Muhammad Aamir, and Waseef Menhaj. Object Recognition in Different Lighting Conditions at Various Angles by Deep Learning Method, October 2022. arXiv:2210.09618 [cs].

[50] Lichao Mou, Pedram Ghamisi, and Xiao Xiang Zhu. Deep Recurrent Neural Networks for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and*

Remote Sensing, 55(7):3639–3655, July 2017. Conference Name: IEEE Transactions on Geoscience and Remote Sensing.

[51] Thien Nguyen, Sangtae Ahn, Hyojung Jang, Sung Chan Jun, and Jae Gwan Kim. Utilization of a combined EEG/NIRS system to predict driver drowsiness. *Scientific Reports*, 7(1):43933, March 2017. Number: 1 Publisher: Nature Publishing Group.

[52] Hiroki Noguchi and Toshihiko Sakaguchi. Effect of Illuminance and Color Temperature on Lowering of Physiological Activity. *Applied Human Science*, 18(4):117–123, 1999.

[53] M. Patel, S. K. L. Lal, D. Kavanagh, and P. Rossiter. Applying neural network analysis on heart rate variability data to assess driver fatigue. *Expert Systems with Applications*, 38(6):7235–7242, June 2011.

[54] Anh-Cang Phan, Ngoc-Hoang-Quyen Nguyen, Thanh-Ngoan Trieu, and Thuong-Cang Phan. An Efficient Approach for Detecting Driver Drowsiness Based on Deep Learning. *Applied Sciences*, 11(18):8441, January 2021. Number: 18 Publisher: Multidisciplinary Digital Publishing Institute.

[55] Antoine Picot, Sylvie Charbonnier, and Alice Caplier. Drowsiness detection based on visual signs: blinking analysis based on high frame rate video. In *2010 IEEE Instrumentation & Measurement Technology Conference Proceedings*, pages 801–804, May 2010. ISSN: 1091-5281.

[56] Hugo F. Posada-Quintero, Tanya Dimitrov, Aurelie Moutran, Seohyoung Park, and Ki H. Chon. Analysis of Reproducibility of Noninvasive Measures of Sympathetic Autonomic Control Based on Electrodermal Activity and Heart Rate Variability. *IEEE Access*, 7:22523–22531, 2019. Conference Name: IEEE Access.

[57] Herlina Abdul Rahim, Ahmad Dalimi, and Haliza Jaafar. Detecting Drowsy Driver Using Pulse Sensor. *Jurnal Teknologi*, 73(3), March 2015. Number: 3.

[58] Muhammad Ramzan, Hikmat Ullah Khan, Shahid Mahmood Awan, Amina Ismail, Mahwish Ilyas, and Ahsan Mahmood. A Survey on State-of-the-Art Drowsiness Detection Techniques. *IEEE Access*, 7:61904–61919, 2019. Conference Name: IEEE Access.

[59] Sergio Rios-Aguilar, José Luis Miguel Merino, Andrés Millán Sánchez, and Álvaro 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*, 3(Regular Issue), 2015.

[60] Akane Sano and Rosalind W. Picard. Toward a taxonomy of autonomic sleep patterns with electrodermal activity. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 777–780, August 2011. ISSN: 1558-4615.

[61] Laurent Schmitt, Jacques Regnard, and Grégoire P. Millet. Monitoring Fatigue Status with HRV Measures in Elite Athletes: An Avenue Beyond RMSSD? *Frontiers in Physiology*, 6, 2015.

[62] Brook A. Shiferaw, Luke A. Downey, Justine Westlake, Bronwyn Stevens, Shantha M. W. Rajaratnam, David J. Berlowitz, Phillip Swann, and Mark E. Howard. Stationary gaze entropy predicts lane departure events in sleep-deprived drivers. *Scientific Reports*, 8(1):2220, February 2018. Number: 1 Publisher: Nature Publishing Group.

[63] Kazuo Tsubota, Seiichiro Hata, Yukio Okusawa, Fuminobu Egami, Tomohiro Ohtsuki, and Katsu Nakamori. Quantitative Videographic Analysis of Blinking in Normal Subjects and Patients With Dry Eye. *Archives of Ophthalmology*, 114(6):715–720, June 1996.

[64] Ayumi Tsuchida, Md. Shoaib Bhuiyan, and Koji Oguri. Estimation of drowsiness level based on eyelid closure and heart rate variability. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2543–2546, September 2009. ISSN: 1558-4615.

[65] United States. Federal Motor Carrier Safety Administration. Technology Division. PERCLOS: A Valid Psychophysiological Measure of Alertness As Assessed by Psychomotor Vigilance. Technical Report FHWA-MCRT-98-006, October 1998.

[66] Darmawan Utomo, Tzu-Hsuan Yang, Dao Thi Thanh, and Pao-Ann Hsiung. Driver Fatigue Prediction Using Different Sensor Data with Deep Learning. In *2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)*, pages 242–247, May 2019.

[67] Emilio Vanoli, Philip B. Adamson, null Ba-Lin, Gian D. Pinna, Ralph Lazzara, and William C. Orr. Heart Rate Variability During Specific Sleep Stages. *Circulation*, 91(7):1918–1922, April 1995. Publisher: American Heart Association.

[68] José Vicente, Pablo Laguna, Ariadna Bartra, and Raquel Bailón. Drowsiness detection using heart rate variability. *Medical & Biological Engineering & Computing*, 54(6):927–937, June 2016.

[69] Fangting Zhang, Gaochao Guo, Peng Yan, Chengwei Zhang, and Xiaojun Hei. A Real-time Fatigue Driving Detection Method Based on Facial Multi-Feature Fusion. In *2021 4th International Conference on Algorithms, Computing and Artificial Intelligence*, ACAI'21, pages 1–8, New York, NY, USA, February 2022. Association for Computing Machinery.

[70] Tianjun Zhu, Chuang Zhang, Tunglung Wu, Zhuang Ouyang, Houzhi Li, Xiaoxiang Na, Jianguo Liang, and Weihao Li. Research on a Real-Time Driver Fatigue Detection Algorithm Based on Facial Video Sequences. *Applied Sciences*, 12(4):2224, January 2022. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

[71] Aleksandar Čolić, Oge Marques, and Borko Furht. Driver Drowsiness Detection and Measurement Methods. In Aleksandar Čolić, Oge Marques, and Borko Furht, editors, *Driver Drowsiness Detection: Systems and Solutions*, SpringerBriefs in Computer Science, pages 7–18. Springer International Publishing, Cham, 2014.

APPENDICES

Appendix A

Motion Sickness Susceptibility Questionnaire Short-form (MSSQ-Short) [26]

Motion Sickness Susceptibility Questionnaire Short-form (MSSQ-Short)

1. Please State Your **Age** .... Years.   *2.* Please State Your Sex (tick box) **Male   Female**

[ ]₁   [ ]₂

This questionnaire is designed to find out how susceptible to motion sickness you are, and what sorts of motion are most effective in causing that sickness. Sickness here means feeling queasy or nauseated or actually vomiting.

Your CHILDHOOD Experience Only (before 12 years of age), for each of the following types of transport or entertainment please indicate:

3. **As a CHILD (before age 12),** how often you **Felt Sick or Nauseated** (tick boxes):

	Not Applicable - Never Travelled	Never Felt Sick	Rarely Felt Sick	Sometimes Felt Sick	Frequently Felt Sick
Cars					
Buses or Coaches					
Trains					
Aircraft					
Small Boats					
Ships, e.g. Channel Ferries					
Swings in playgrounds					
Roundabouts in playgrounds					
Big Dippers, Funfair Rides					
	t	0	1	2	3

Your Experience over the LAST 10 YEARS (approximately), for each of the following types of transport or entertainment please indicate:

4. **Over the LAST 10 YEARS,** how often you **Felt Sick or Nauseated** (tick boxes):

	Not Applicable - Never Travelled	Never Felt Sick	Rarely Felt Sick	Sometimes Felt Sick	Frequently Felt Sick
Cars					
Buses or Coaches					
Trains					
Aircraft					
Small Boats					
Ships, e.g. Channel Ferries					
Swings in playgrounds					
Roundabouts in playgrounds					
Big Dippers, Funfair Rides					
	t	0	1	2	3

70

Scoring the MSSQ- Short

Section A (Child) (Question 3)

Score the number of types of transportation <u>not</u> experienced (i.e., total the number of ticks in the 't' column, maximum is 9).

Total the sickness scores for each mode of transportation, i.e. the nine types from 'cars' to 'big dippers' (use the 0-3 number score key at bottom, those scores in the 't' column count as zeroes).

MSA = (total sickness score child) x (9) / (9 - number of types not experienced as a child)

Note 1. Where a subject has not experienced any forms of transport a division by zero error occurs. It is not possible to estimate this subject's motion sickness susceptibility in the absence of any relevant motion exposure.

Note 2. The Section A (Child) score can be used as a pre-morbid indicator of motion sickness susceptibility in patients with vestibular disease.

Section B (Adult) (Question 4)

Repeat as for section A but using the data from section B.

MSB = (total sickness score adult) x (9) / (9 - number of types not experienced as an adult)

Raw Score MSSQ-Short

Total the section A (Child) MSA score and the section B (Adult) MSB score to give the MSSQ-Short raw score (possible range from minimum 0 to maximum 54, the maximum being unlikely)

MSSQ raw score = MSA + MSB

Percentile Score MSSQ-Short

The raw to percentile conversions are given below in the Table 1 of Statistics & Figure 1. Use interpolation where necessary.

Alternatively a close approximation is given by the fitted polynomial where y is percentile; x is raw score
$y = a.x + b.x^2 + c.x^3 + d.x^4$
a = 5.1160923   b = -0.055169904
c = -0.00067784495   d = 1.0714752e-005

Table 1. Means and Percentile Conversion Statistics for the MSSQ-Short (n=257)

| Percentiles Conversion	Raw Scores MSSQ-Short		
	Child Section A	Adult Section B	Total A+B
0	0	0	0
10	.0	.0	.8
20	2.0	1.0	3.0
30	4.0	1.3	7.0
40	5.6	2.6	9.0
50	7.0	3.7	11.3
60	9.0	6.0	14.1
70	11.0	7.0	17.9
80	13.0	9.0	21.6
90	16.0	12.0	25.9
95	20.0	15.0	30.4
100	23.6	21.0	44.6
Mean	7.75	5.11	12.90
Std. Deviation	5.94	4.84	9.90

Table note: numbers are rounded
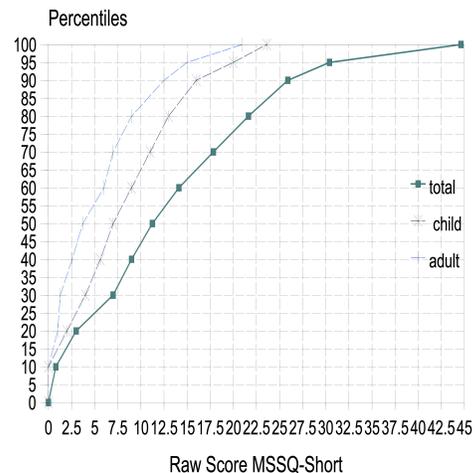
Figure 1. Cumulative distribution Percentiles of the Raw Scores of the MSSQ-Short (n=257 subjects).

71

Development, Normalisation & Validation of the MSSQ-Short (Golding, 2006)

Background   Motion sickness susceptibility questionnaires (MSSQ), sometimes called motion history questionnaires, predict individual differences in motion sickness caused by a variety of stimuli. The original "Reason & Brand MSSQ" (Reason, 1968; Reason & Brand, 1975) had perhaps the best proven track record in the world for motion sickness research. It was subsequently revised, renormalized & revalidated (Golding, 1998). The aim was to develop a short version of the MSSQ, denoted "MSSQ-Short".

Methods   Development used repeated item analysis, and various scoring methods of the MSSQ (Golding, 1998). Retained were: motion types (cars, boats, planes, trains, funfair rides, etc); corrections for motion type exposure with a much simplified format; sickness severity weightings; childhood versus adult experiences. New items such as visual/optokinetic items (cinerama, virtual reality, etc), were introduced but then excluded since they had low sickness prevalence & added little information. However they could become important in the future. Norms and percentiles were produced (n=257). Predictive validity used controlled motions representing all classes of motion sickness provocative stimuli (total n=178): cross-coupled (Coriolis); 0.2Hz frequency translational oscillation; off-vertical axis rotation (OVAR); visual-motion simulator.

Results   Predictive validity for motion was median r =0.51. The relationship between MSSQ-Short and other non-motion sources of nausea and vomiting (e.g. headaches, food, stress, viral, etc) in the last 12 months was r=0.2 (p<0.01). Reliability: Cronbach's alpha was 0.87; Test-retest reliability was around r=0.9; Part A (child) with Part B (adult) was r=0.68.

Conclusions   The MSSQ-Short provides reliability with an efficient compromise between length (reduced time cost) and validity (predicted motion susceptibility). Language variants include French, Italian, Spanish, Dutch, Flemish, German, Russian and Chinese.

References

Reason JT. (1968) Relations between motion sickness susceptibility, the spiral after-effect and loudness estimation. **Br J Psychol**; 59: 385-93.

Reason JT & Brand JJ. (1975) **Motion sickness**. London: Academic Press.

Golding JF. (1998) Motion sickness susceptibility questionnaire revised and its relationship to other forms of sickness. **Brain Research Bulletin**, 47: 507-516.

Golding JF. (2006) Predicting Individual Differences in Motion Sickness Susceptibility by Questionnaire. **Personality and Individual differences,** 41: 237-248**.**

Appendix B

Fatigue Assessment Scale (FAS)

Welcome, This is a strictly confidential questionnaire. You can skip any questions you do not feel comfortable answering.

The Fatigue Assessment Scale (FAS) is a 10-item scale evaluating symptoms of chronic fatigue developed by [48]. It treats fatigue as a uni-dimensional construct and the questions represent a combination of physical and mental health symptoms.

Fatigue Assessment Scale

The following 10 statements refer to how you usually feel. For each statement you can choose one out of five answer categories, varying from never to always. 1 = never, 2 = sometimes; 3 = regularly; 4 = often; and 5 = always.

1. I am bothered by fatigue (WHOQOL), 2. I get tired very quickly (CIS), 3. I don't do much during the day (CIS), 4. I have enough energy for everyday life (WHOQOL), 5. Physically, I feel exhausted (CIS), 6. I have problems starting things (FS), 7. I have problems thinking clearly (FS), 8. I feel no desire to do anything (CIS), 9. Mentally, I feel exhausted, 10.When I am doing something, I can concentrate quite well (CIS) [48].

	Never	Sometimes	Regularly	Often	Always
1. I am bothered by fatigue (WHOQOL)	1	2	3	4	5
2. I get tired very quickly (CIS)	1	2	3	4	5
3. I don't do much during the day (CIS)	1	2	3	4	5
4. I have enough energy for everyday life (WHOQOL)	1	2	3	4	5
5. Physically, I feel exhausted (CIS)	1	2	3	4	5
6. I have problems starting things (FS)	1	2	3	4	5
7. I have problems thinking clearly (FS)	1	2	3	4	5
8. I feel no desire to do anything (CIS)	1	2	3	4	5
9. Mentally, I feel exhausted	1	2	3	4	5
10. When I am doing something, I can concentrate quite well (CIS)	1	2	3	4	5

Appendix C

Screening Questionnaire

You are invited to participate in a research study that is analyzing eye patterns, facial expressions, driving behaviour changes, and heart rate changes to detect any symptoms of drowsiness that might help reduce the number of accidents that occur every day. Drowsy driving continues to be one of the most predominant types of road traffic accidents. At least 1 in 25 adult drivers have confirmed suffering from fatigue while driving in the last 30 days. Millions of users drive daily for various reasons increasing the chances of possible accidents. The procedure that the participants will encounter is to answer some simple questionnaires and drive on a driving simulator in one of two scenarios, a monotonous and a non-monotonous. The data that will be collected to proceed with this research includes videos of the participant's face and eyes, changes in driving patterns, and heart rate while driving.

Are you at least 18 years old? Yes No

Do you have valid full G to G2 driver's license (or equivalent)? Yes No

Was your driver's license issued at least one year before participating in this study? (Indicating one year of driving experience). Yes No

Do you have full eye vision or corrected to full eye vision? Yes No

Have you ever suffered from motion sickness while watching videos, playing video games, or any other type of scenario with electronics? Yes No

Appendix D

Activity Log

On this activity log, we are looking to know what occurred in the past 12-8 hours before your participation. We would like to know what you ate for dinner, breakfast, and lunch (If you already had one before coming to the test), how many hours did you sleep, at what time you woke up, etc.

1. What did you eat for dinner yesterday? (Please also include drinks)

2. What did you eat for breakfast today? (Please also include drinks)

3. If you already had lunch. What did you eat for lunch today? (Please also include drinks)(If your answer is no, please skip this question)

4. At what time did you go to sleep yesterday?

5. At what time did you wake up today?

6. Did you consume any caffeine related products before your test?

Appendix E

Notification of Ethics Clearance to Conduct Research with Human Participants

UNIVERSITY OF WATERLOO

Notification of Ethics Clearance to Conduct Research with Human Participants

---

Principal Investigator: Siby Samuel (Systems Design Engineering)

Student investigator: Jose Alguindigue (Systems Design Engineering)

Co-Investigator: Apurva Narayan (Systems Design Engineering)

File #: 44726

Title: Fatigue detection system using a driving simulator

The Human Research Ethics Board is pleased to inform you this study has been reviewed and given ethics clearance.

Initial Approval Date: 01/05/23 (m/d/y)

University of Waterloo Research Ethics Boards are composed in accordance with, and carry out their functions and operate in a manner consistent with, the institution's guidelines for research with human participants, the Tri-Council Policy Statement for the Ethical Conduct for Research Involving Humans (TCPS, 2nd edition), International Conference on Harmonization: Good Clinical Practice (ICH-GCP), the Ontario Personal Health Information Protection Act (PHIPA), the applicable laws and regulations of the province of Ontario. Both Boards are registered with the U.S. Department of Health and Human Services under the Federal Wide Assurance, FWA00021410, and IRB registration number IRB00002419 (HREB) and IRB00007409 (CREB).

This study is to be conducted in accordance with the submitted application and the most recently approved versions of all supporting materials.

Expiry Date: 01/06/24 (m/d/y)

Multi-year research must be renewed at least once every 12 months unless a more frequent review has otherwise been specified. Studies will only be renewed if the renewal report is received and approved before the expiry date. Failure to submit renewal reports will result in the investigators being notified ethics clearance has been suspended and Research Finance being notified the ethics clearance is no longer valid.

Level of review: Delegated Review

Signed on behalf of the Human Research Ethics Board

Heather Dekker, Ethics Advisor, hdekker@uwaterloo.ca, 519-888-4567, ext. 41506

This above named study is to be conducted in accordance with the submitted application and the most recently approved versions of all supporting materials.

Documents reviewed and received ethics clearance for use in the study and/or received for information:

file: Remuneration Form.pdf

file: Protocol_Approved.pdf

file: dosovitskiy17a.pdf

file: g27-racing-wheel-quickstart-guide.pdf

file: QS-Guide DG3_Cable and Wireless.pdf

file: emaptics.pdf

file: Driving Simulator SOP.pdf

file: RecruitmentPoster_version3_20230104.pdf

file: EmailScript_version3_20230104.pdf

file: ScreeningQuestionnaire_Version3_20230103.pdf

file: DemographicQuestionnaire_Version2_20230103.pdf

file: FatigueAssesmentScale_Version1_20221115.pdf

file: MotionSicknessSusceptibilityQuestionnaire_V2.pdf

file: ActivityLog_version1_20230104.pdf

file: GeneralConsentForm_version3_20230104.pdf

file: PhotoAndVideoConsentForm_Version3_20230104.pdf

file: InformationLetter_version3_20230104.pdf

file: OnlineConsentForms_version1_20230104.pdf

file: ParticipantAppreciationLetter_version2_20230103.pdf

Approved Protocol Version 4 in Research Ethics System

This is an official document. Retain for your files.

You are responsible for obtaining any additional institutional approvals that might be required to complete this study.