Methods and Techniques for Analyzing Human Factors Facets on Drivers

by

Botyu'u Óscar Sipele Siale

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in

Computer Science and Technology

Universidad Carlos III de Madrid

Advisor(s):

Dr. Agapito Ismael Ledezma Espino
Prof. Dr. María Araceli Sanchis de Miguel

February 2022

I would like to dedicate this Ph. D Thesis
to my parents, my sister Sandra and my girlfriend Lucía
for supporting me unconditionally
during this period of my life.

ACKNOWLEDGEMENTS

First of all, I would like to acknowledge and give my thanks to my supervisors Prof. Dr. Araceli Sanchis de Miguel and Dr. Agapito Ismael Ledezma Espino, who gave me the opportunity of getting involved in this project. Thanks a lot for your guidance and advice during all time I have belonged to the CAOS research group.

I would also like to give special thanks to CAOS research group members: Jose Antonio Iglesias, Paz Sesmero, Paula de Toledo and, Juan Manuel Alonso-Webber; because I learned a lot nearby all of you and I feel grateful to have been surrounded by great people who never refused to help me when I needed it.

My thanks and appreciation also go to my colleagues Juan Andrés Caiceido, Camilo Corrales, Jose Antonio Simmons, Angela Vargas, and Victor Manuel Zamora for sharing great laboratory moments and cooperating in this project.

Also, I would like to express my gratitude towards Prof. Dr. Cristina Olaverri-Montreal, who guided me during my internship work at Johannes Kepler University (Linz, Austria), and all ITS Research group members: Walter Moralez-Alvarez, Aso Validi, Nikita Smirnov, Georg Novotni, Yuzhou Liu and Jihed Khiari; to make me feel like part of your team.

Finally, I would like to thank the Ministry of Economy and Competitiveness for granting me the predoctoral fellowship BES-2016-078143 corresponding to the project TRA2015-63708-R, which provided me the opportunity of conducting all my Ph. D activities, including completing an international internship.

PUBLISHED AND SUBMITTED CONTENT

The published contributions fully included in this thesis are as follows:

- **O. Sipele**, V. Zamora, A. Ledezma, and A. Sanchis, "Advanced Driver's Alarms System through Multi-agent Paradigm," in *2018 3rd IEEE International Conference on Intelligent Transportation Engineering, ICITE 2018*, 2018, pp. 269–275. DOI: 10.1109/ICITE.2018.8492600
- **O. Sipele**, A. Ledezma, and A. Sanchis, "Integration Model of Multi-Agent Architectures for Data Fusion-Based Active Driving System," in Human Factors in Intelligent Vehicles, C. Olaverri-Monreal, F. García-Fernández, and R. J. F. Rossetti, Eds. River Publishers, 2020. ISBN: 9788770222037

The contributions partly included in this thesis are as follows:

- Ledezma, A.; Zamora, V.; **Sipele, Ó.**; Sesmero, M.P.; Sanchis, A. Implementing a Gaze Tracking Algorithm for Improving Advanced Driver Assistance Systems. Electronics 2021, 10, 1480, doi:10.3390/electronics10121480.
- V. Zamora, **O. Sipele**, A. Ledezma, and A. Sanchis, "Intelligent Agents for Supporting Driving Tasks: An Ontology-based Alarms System," in *Proceedings of the 3rd International Conference on Vehicle Technology and Intelligent Transport Systems - Volume 1: VEHITS,* 2017, pp. 165–172. DOI: 10.5220/0006247601650172
- P. Angelov, X. Gu, J.A. Iglesias, A. Ledezma, A. Sanchis, **O. Sipele**, R. Ramezani, "Cybernetics of the Mind: Learning Individual's Perceptions Autonomously," IEEE Syst. Man, Cybern. Mag., vol. 3, no. 2, pp. 6–17, Apr. 2017, DOI: 10.1109/MSMC.2017.2664478
- I. Škrjanc, G. Andonovski, A. Ledezma, **O. Sipele**, J. A. Iglesias, and A. Sanchis, "Evolving cloud-based system for the recognition of drivers' actions," *Expert Syst. Appl.*, vol. 99, 2018. DOI: 10.1016/j.eswa.2017.11.008
- G. Andonovski, **O. Sipele**, J. A. Iglesias, A. Sanchis, E. Lughofer, and I. Skrjanc, "Detection of driver maneuvers using evolving fuzzy cloud-based system," in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2020, pp. 700–706. DOI: 10.1109/SSCI47803.2020.9308520

OTHER RESEARCH MERITS

The work V. Zamora, **O. Sipele**, A. Ledezma, and A. Sanchis, "Intelligent Agents for Supporting Driving Tasks: An Ontology-based Alarms System," in *Proceedings of the 3rd International Conference on Vehicle Technology and Intelligent Transport Systems - Volume 1: VEHITS,* 2017, pp. 165–172. DOI: 10.5220/0006247601650172, whose oral presentation was conducted by **O. Sipele** was selected for the Best Poster Award of this conference.

The internship work was carried out in the Intelligence Transportation Systems research group of Johannes Kepler University (Linz, Austria) under the supervision of Professor Ph.D. Cristina Olaverri-Monreal propitiates the contribution on the several publications as follows:

- **O. Sipele**, A. Ledezma, and A. Sanchis, "Integration Model of Multi-Agent Architectures for Data Fusion-Based Active Driving System," in Human Factors in Intelligent Vehicles, C. Olaverri-Monreal, F. García-Fernández, and R. J. F. Rossetti, Eds. River Publishers, 2020. ISBN: 9788770222037
- W. Morales-Alvarez, **O. Sipele**, R. Léberon, H. H. Tadjine, and C. Olaverri-Monreal, "Automated Driving: A Literature Review of the Take-over Request in Conditional Automation," Electronics, vol. 9, no. 12, p. 2087, 2020. DOI: 10.3390/electronics9122087
- W. Morales-Alvarez, F. M. Moreno, **O. Sipele**, N. Smirnov, and C. Olaverri-Monreal, "Autonomous Driving: Framework for Pedestrian Intention Estimation in a Real-World Scenario," in 2020 IEEE Intelligent Vehicles Symposium (IV), 2020, pp. 39–44. DOI: 10.1109/IV47402.2020.9304624

Abstract

With millions of cars moving daily, driving is the most performed activity worldwide. Unfortunately, according to the World Health Organization (WHO), every year, around 1.35 million people worldwide die from road traffic accidents and, in addition, between 20 and 50 million people are injured, placing road traffic accidents as the second leading cause of death among people between the ages of 5 and 29. According to WHO, human errors, such as speeding, driving under the influence of drugs, fatigue, or distractions at the wheel, are the underlying cause of most road accidents. Global reports on road safety such as "Road safety in the European Union. Trends, statistics, and main challenges" prepared by the European Commission in 2018 presented a statistical analysis that related road accident mortality rates and periods segmented by hours and days of the week. This report revealed that the highest incidence of mortality occurs regularly in the afternoons during working days, coinciding with the period when the volume of traffic increases and when any human error is much more likely to cause a traffic accident.

Accordingly, mitigating human errors in driving is a challenge, and there is currently a growing trend in the proposal for technological solutions intended to integrate driver information into advanced driving systems to improve driver performance and ergonomics. The study of human factors in the field of driving is a multidisciplinary field in which several areas of knowledge converge, among which stand out psychology, physiology, instrumentation, signal treatment, machine learning, the integration of information and communication technologies (ICTs), and the design of human-machine communication interfaces.

The main objective of this thesis is to exploit knowledge related to the different facets of human factors in the field of driving. Specific objectives include identifying tasks related to driving, the detection of unfavorable cognitive states in the driver, such as stress, and, transversely, the proposal for an architecture for the integration and coordination of driver monitoring systems with other active safety systems. It should be noted that the specific objectives address the critical aspects in each of the issues to be addressed.

Identifying driving-related tasks is one of the primary aspects of the conceptual framework of driver modeling. Identifying maneuvers that a driver performs requires training beforehand a model with examples of each maneuver to be identified. To this end, a methodology was established to form a data set in which a relationship is established between the handling of the driving controls (steering wheel, pedals, gear lever, and turn indicators) and a series of adequately identified maneuvers. This methodology consisted of designing different driving scenarios in a realistic driving simulator for each type of maneuver, including stop, overtaking, turns, and specific maneuvers such as U-turn and three-point turn.

From the perspective of detecting unfavorable cognitive states in the driver, stress can damage cognitive faculties, causing failures in the decision-making process. Physiological signals such as measurements derived from the heart rhythm or the change of electrical properties of the skin are reliable indicators when assessing whether a person is going

through an episode of acute stress. However, the detection of stress patterns is still an open problem. Despite advances in sensor design for the non-invasive collection of physiological signals, certain factors prevent reaching models capable of detecting stress patterns in any subject. This thesis addresses two aspects of stress detection: the collection of physiological values during stress elicitation through laboratory techniques such as the Stroop effect and driving tests; and the detection of stress by designing a process flow based on unsupervised learning techniques, delving into the problems associated with the variability of intra- and inter-individual physiological measures that prevent the achievement of generalist models.

Finally, in addition to developing models that address the different aspects of monitoring, the orchestration of monitoring systems and active safety systems is a transversal and essential aspect in improving safety, ergonomics, and driving experience. Both from the perspective of integration into test platforms and integration into final systems, the problem of deploying multiple active safety systems lies in the adoption of monolithic models where the system-specific functionality is run in isolation, without considering aspects such as cooperation and interoperability with other safety systems. This thesis addresses the problem of the development of more complex systems where monitoring systems condition the operability of multiple active safety systems. To this end, a mediation architecture is proposed to coordinate the reception and delivery of data flows generated by the various systems involved, including external sensors (lasers, external cameras), cabin sensors (cameras, smartwatches), detection models, deliberative models, delivery systems and machine-human communication interfaces. Ontology-based data modeling plays a crucial role in structuring all this information and consolidating the semantic representation of the driving scene, thus allowing the development of models based on data fusion.

Table of contents

List of figures

List of tables

Acronyms and abbreviations

ADAS	Advanced Driving Assistance System
ADS	Automated Driving System
ANS	Autonomous nervous system
BIC	Bayesian Information Criterion
BVP	Blood volume pulse
CWT	Continuous Wavelet Transform
DDT	Dynamic Driving Task.
DTW	Dynamic Time Warping
EC	European Commission
EDA	Electrodermal activity
EDL	Electrodermal level
EDR	Electrodermal response
GMM	Gaussian Mixture Model
HFE	Human Factors/Ergonomics
HRV	Heart rate variability
ITS	Intelligent Transport System
KDE	Kernel Density Estimation
LOOCV	Leave-one-person-out cross-validation
OEM	Original Equipment Manufacturer
PCA	Principal Component Analysis
PNS	Parasympathetic nervous system
PPG	Photoplethysmography
PRV	Pulse rate variability
SAE	Society of Automotive Engineers
SCL	Skin conductance level
SCR	Skin conductance response
SNS	Sympathetic nervous system
TS	Timestamp
TF-IDF	Term Frequency-Inverse document frequency
WHO	World Health Organization
WVT	Wavelet

1. Introduction

With millions of cars traveling daily, driving is the most performed activity worldwide. Except for certain cities such as Singapore, wherein public transportation is the preferred commuting choice, statistical studies about commuting mode and car-sharing indicate that private car is the most extended mobility option for regular conveyance across all countries[1]. Furthermore, as far as travel mobility, private and rental cars were the predominant means of transport in the country members of the European Union. Cars cumulated almost 7 out of 10 domestic trips, and it was the second choice after the airplane as regard outbound trips such as was reported by Eurostat in 2017 [2].

Unfortunately, according to the World Health Organization (WHO) report regarding road safety elaborated in 2018 [3], every year around 1.35 million people die across the world due to road traffic crashes and, between 20 and 50 million people were injured besides, with numerous suffering an incapacity because of the severity of their injuries [4]. Moreover, as this report describes, traffic crashes are the second fatality causes of persons within 5 and 29 years old.

Due to the relevance of road transport systems, road safety issues are in the spotlight of Community Governments. Policy orientations are constantly readapting to guarantee the fulfilling the road safety program objectives that deal with reducing the fatality rates caused by traffic crashes. Road safety challenges of the Horizon 2020 project (H2020) defined the strategic plan to cut the fatalities by 50% in European roadways as the primary objective. Within the strategic objectives defined in the *Policy orientations on Road Safety 2011-2020* [5], the improvement and integration of new technologies, including vehicle network connectivity and Intelligent Transport Systems (ITS), are taken an essential role in the achievement of safer roads, focusing on the mitigation of risk factors concerning human errors due to their remarkable predominance in the traffic crash rates. Thus, the H2020 framework has marked the guidelines for the member countries' new transportation and road safety initiatives. Currently, in 2021, the funding plan Horizon Europe has budgeted 95.5 billion euros within years 2021 and 2027 assigned to promote the research and innovation, industrial competitiveness, and optimize the investment impact within a European Research Area [6]. In the scope of climate, energy, and mobility, challenges related to the safety of connected and highly automated vehicles are taking much relevance due to the incipient arrival of new technologies making the mobility paradigms evolve [7].

Crashes directly related to driver's distraction, e.g., head-on collisions, run-off-road, and intersection crashes, involve most fatalities and severe injuries in high-income countries [8].

Studies such as those conducted by Rolison et al. [9] argued that the leading causes of collisions are directly related to driver impairment and distractions despite not being included in official accident records due to the lack of completeness of traffic accidents reports.

However, the global report entitled "*Road safety in the European Union. Trends, statistics and main challenges*", elaborated by the European Commission in 2018 [10], exposed how the days and hours of the week influence the incidence of traffic deaths. This report reveals that death peaks occur in the afternoon during workdays, coinciding with the period when traffic volume increases and any human error cited before is much more susceptible to causing a traffic crash.

Therefore, studying the underlying factors involved in driver's distractions is crucial to reducing the rate of traffic crashes due to this cause.

1.1. Motivation

Human Factors and Ergonomics (HFE) is a multidisciplinary research field that involves diverse knowledge areas. Chapanis [11] defined the concept of Human Factors as:

> *"A discipline that discovers and applies information about human behavior, abilities, limitations and other characteristics to the design of tools, machines, systems, tasks, jobs, and environments for productive, safe, comfortable and effective human use."*

Human factors have an essential role in the distinct facets of industrial processes, such as designing and assessing new products. The standards published by Ahlstrom [12] provide framework guidelines for supporting the selection, analysis, design, development, and evaluation of vehicle systems and equipment.

In the last twenty years, the widely introduced Advanced Driving Assistance System (ADAS) has been the technological solution for reducing human errors, having as main aim to support the driver during their primary task. According to its strict definition, an ADAS is characterized as a system that explores the vehicle environment, applies complex signal processing, and actively supports about vehicle's heading and lateral inspection by direct interaction with the driver [13].

Currently, well-established ADASs base their functioning on exteroceptive sensors principally, such as was reviewed by Bengler et al. [14]. In the ADAS timeline, systems that support informed driving evolved palpably between 2000 and 2007. From purely informative systems, e.g., navigation systems, towards more advanced systems such as Forward Collision Warning, Lane Departure Warning, and Adaptive Cruise Control, which provide warning information employing ultra-sonic, radar, and infrared sensors and cameras for the inspection of car's surrounding, the in-vehicle information raised notably.

However, in 2007, the driver drowsiness detection system was introduced as the first system based on driver state analysis [15]. The Volvo approach relied on detecting the gradual detriment of drivers determining impairments through the irregularities between vehicle trajectory and lane markings.

Parallelly, Toyota incorporated CCD cameras in its luxury model cars to detect driver attentiveness first by gazing monitoring and by eye blinking analysis posteriorly. These distinct ways of assessing the driver's engagement in driving tasks marked a notable change in the research and development paradigms of ADAS.

Back in the day, with the arrival of electronic systems such as power steering and anti-lock braking systems, the driver modeling paradigm emerged, introducing driver's characteristics to enhance the active safety systems performance and improve the driving experience. Early studies focused on delineating the abstraction of driving tasks. Michon (1985) distinguished the diverse aspects of a driving task hierarchically [16]. Thereby, operational, tactical, and strategic levels were conceptualized regarding the timescale of the actions performed to drive and control, and tactical actions were considered the most relevant actions linked with driving safety. Kirwan and Ainsworth (1992) proposed general methods such as Hiera rchical Task Analysis (THA) or Time Line Analysis (TLA) to represent human tasks in the context of industrial procedures. Posteriorly, driver models incorporated aspects about driver's mental processes ceasing to consider only driving-related actions. Endsley (1995) introduced the situation awareness model as the cognitive model representing the driver's decision-making process [17].

More recently, studies focused on describing the driver's state, delving into the conditions that propitiate distractions and impairments. According to dynamically changing conditions of vehicle and environment previously stated by Kirwan and Ainsworth [18], Cacciabue introduced experience, personality, task demand, driver state, and situation awareness as the five measurable parameters that directly affects the driver's behavior and conducted a study to estimate how susceptible is a driver to conduct mistakes [19].

Stanton et al. [20] exposed that the quality of interaction process with automated systems highly depends on six relevant factors, including the state of the locus of control, trust, stress, situation awareness, mental model, mental workload, and task demands, thereby describing the driver's state at the moment when they receive the system feedback. Recarte and Nunes described the sources of distraction in more depth and delved into visual and cognitive issues according to their categorization of distractions based on four main causative factors of poor perception and misunderstanding of the driving scene [21].

According to exposed above, integrating driver monitoring models that introduce driver's characteristics is a fundamental step towards improving active safety systems' functioning and driving experience.

Currently, monitoring aspects open up several research directions. Horizon Europe challenges in mobility ask for new monitoring shapes in which inspection involves drivers and passengers. These monitoring topics include health status, level, and point of attention, and they should coordinate with intelligent safety systems and adequately linked with a human-machine interface.

Another relevant driver monitoring topic concerns partial and conditional driving automation (SAE levels 2 and 3). In this scope, driver monitoring is crucial to determine whether the human driver is supervising the automation feature functioning or not because

they should be engaged in the driving task when the limit of the operational driving domain is approaching.

1.2. Research questions

Concerning driver states, two tightly interrelated driver modeling approaches have significantly influenced applications on driver monitoring. First, driver models determine the anatomical signs associated with impairing driver conditions. Second, driver models regarding the analysis of physiological signs assume that the impaired condition is the underlying effect of altered consciousness states. The assessment of how the driving-related tasks are performed can infer the nearness between current driving behavior and the driver's regular driving patterns. Transversally, the driving activity on its own can also provide information about the driver's state.

Furthermore, the employment of driver monitoring methods as part of the input set taken by ADAS levels 1 to 4 [22] suggests as principal requirement managing hierarchical information flows under a common data infrastructure. About how novel approaches are tested, it is crucial to replicate data integration infrastructure in a flexible and scalable platform that can reproduce driving situations without supposing real jeopardy for test drivers. Accordingly, a microscopic driving simulator enables the primary functionality test to examine the operational design domain conditions of ADAS in development without neglecting safety constraints. The statements exposed above arose the four research questions for this Ph.D. thesis, as follows:

- Which technological and technical resources are demanded in integrating driver behavior modeling on ADAS?
- Which factors should be considered for collecting data regarding driving-related tasks?
- Which conditions should fulfill a driving test for assessing a new driver-centered safety application during its development phase?
- How reliable are the non-intrusive physiological sensors for real-time determining the unfit mental state?

1.3. Objectives

This thesis addresses knowledge representation, data collection, and patterns identification.

This thesis aims to address the integration of different human factor facets in the scope of driving, employing gathered information collected from the cockpit employing nonobtrusive sensors. The main application case explored in this thesis is the obtention and validation of a stress detection model using wearable sensors and its integration into driving-based knowledge models.

Accordingly, the derived sub-objectives of this thesis are described as follows:

O1 Proposing an integration architecture model able to hold information and transport multiple abstraction levels concerning diverse aspects of the driving domain.

O2 Elaborating a study for generating driving-related activities datasets using driving controls exclusively.

O3 Exploring the reliability of physiological signals for identifying stress patterns in real-time.

O4 Validating the usability of the proposed approach of physiological signals for identifying stress patterns in real-time.

1.4. Structure of the document

After this introduction, the remainder document is divided into the following chapters:

- Chapter 2 reviews the state-of-art of the related topics covered in this thesis.
- Chapter 3 describes the designed protocol for collecting related driving tasks.
- Chapter 4 presents the methodology applied to collect physiological signals during stressor exposure elicitation.
- Chapter 5 delves into the different methods based on unsupervised learning to identify stress patterns using electrodermal activity as the leading physiological indicator.
- Chapter 6 details the proposed architecture model for integrating multiple facets of the driving scene and presents the validation experiment conducted in a driving simulator.
- Chapter 7 overviews the findings of this Ph.D. thesis and presents the future research guidelines.

2. State-of-Art

To better understand how human factors can improve road safety, first, it is necessary to analyze in general terms which driver's aspects are involved in the current perspective of active safety systems. Figure 2.1 provides a global view about the driving domain aspects and the role that human factors play in the constant information exchange between the driver, the car equipped with active safety systems, and the environment, thereby defining the dynamic driving task (DDT). The Standard SAE 3016 [22] defines DDT as:

"*All of the real-time operational and tactical functions required to operate a vehicle in on-road traffic, excluding the strategic functions such as trip scheduling and selection of destinations and waypoints.*"

However, operational (lateral and longitudinal vehicle control) and tactical (maneuvering correctly signalized) functions must be coupled adequately with driving conditions that the environment dictates. For instance, exceeding the speed limit on a foggy day or not adapting the speed on the arrival of dense traffic areas constitute driving misconducts that potentially can provoke a traffic crash.

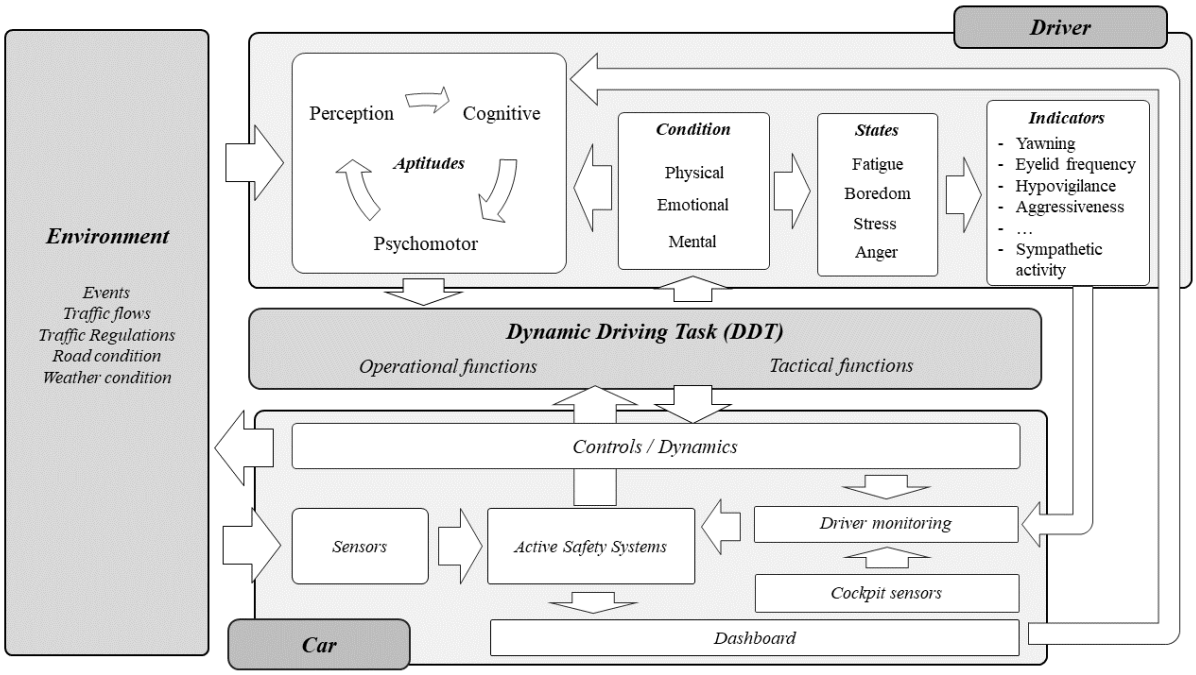


Figure 2.1. Conceptual framework of driving task and the role of human factors from an active safety systems perspective.

Consequently, a well-performing driving task mostly depends on the driver's state because the driver's condition leads to the aptitudes of perceiving the environment correctly, conducting effective decision-making and, performing the appropriate maneuver efficiently.

As a starting point of this state-of-the-art, it will introduce the underlying causes of human errors during the driving performance.

The major causes of accidents come from fatigue, and it can manifest both physically and psychologically. The detriment of muscular power and coordination directly results from physical fatigue, and these capabilities are crucial for well-performing operational functions [23]. Mental fatigue can be originated by different causes, both related and non-related with the driving task. Many factors, including sleep, nutrition and, physical health, can contribute to an early appearance of fatigue, such as was reviewed by Lal and Craig [24]. However, fatigue emerges when the driver performs DDT for a long time, especially in monotonous driving environments. The low rate of environmental stimuli provokes hypovigilance on the driver that is directly associated with boredom. Consequently, the driver decreases their engagement level with the DDT and their alertness [25].

When impaired driving states are analyzed, stress is also a key factor to keep in mind. Diverse factors can induce stress on a person, including ambient factors (noise, extreme temperatures) and physical state (fatigue) [26], the exposition to psychological stressors (fear, anxiety) [27], and economic and socio-cultural aspects (life events, chronic straits, daily hassles) [28]. Endsley [29] identified and interpreted how stress can harm cognitive faculties, and specifically, which factors can lead to failures in the decision-making process as part of situational awareness (SA). This research substantiates the belief that the different shapes of stress reduce the people's field of attention, thus disregarding some aspects in favor of others during the decision-making, this being a fact with fatal consequences in emergency situations. Moreover, the author exposed the underlying SA issues related to task performing in dynamic systems. For instance, regarding the task nature, the performance of tasks with elevated levels of workload and complexity during extended periods supposes a stressor factor that gradually deteriorates the SA [11].

Anger is also a relevant factor of the driver's state, whose origin can be associated with the driver's personality trait and a transitory state resulting from stress. It used to be expressed by facial and body expressions, verbal communication, and aggressive driving patterns. Accordingly, specific psychological tools were presented to quantify the personality traits related to driving as an index of health. Driving Anger Scale (DAS) [30] describes the driving trait anger by six clusters that include the adopted behaviors when some specific situations occur, such as (1) hostile gestures, (2) illegal driving, (3) police presence, (4) slow driving, (5) discourtesy and (6) traffic obstructions. Driving Anger Expression Inventory (DAX) [31] identified how drivers react when angry feelings emerge. Accordingly, there are three ways of coping with anger: with constructive expressions, in which the driver tries to stay relaxed, and with anger expressions manifested by verbal abuses, personal physical aggressions, or driving aggressive misconducts.

The explained factors above manifest that human errors can arise from diverse causes, and the primary purpose of new car safety features should focus on mitigating the effects that impaired states on driving. Thus, driver monitors should target the visible or measurable symptoms of the impaired states described above.

One essential requirement that driving monitors must fulfill is using non-obtrusive sensors, such as cameras and wearable health devices, to gather data from the car's cockpit without affecting the driving performance.

Another covered aspect for driving monitors is recognizing and assessing the performed driver's activities related to the driving task. Maneuvers and the measure of the vehicle's position on the road are also critical aspects for determining the correct interaction with the dynamic evolving of the driving scene with indicators such as variation of lane deviation [32].

New prospect of active safety systems should manage heterogeneous information sources, gathering environment and driver's condition, thereby composing an abstraction of the driving scene. The safety application performance improvement and the automation level increment must tackle different issues. For example, the concept of fallback emerged due to the need of determining the limits both of a safety application (automation level 3) and driver's aptitudes (automation level 4), being this the foundation of taking over request whose operation constitutes the transition from manual to automation driving, and vice versa [22]. Control actions depend on the purpose of driving safety application where active safety systems can only warn to the driver (level 0) or perform operational functions (within levels 1 and 2) and also tactical functions (within levels 3 and 4).

Driving monitoring is crucial for improving ergonomic aspects about the interoperability of coordinate ADAS and determining driver fallbacks in the scope of automated driving. This general view reveals the need to introduce the human factors facets as part of the abstraction of the driving scene to enrich the reasoning of the active driving system. However, the abstraction of the driver's state covers broad diversity aspects, both physical and psychological, whose integration can result pretty tedious.

After this introduction of general definitions within the scope of HFE in driving, this review of state-of-art is framed on three main aspects:

- **The usage of visual information of the driver's physical features to analyze and detect impaired driving states**.
- **The methods intended to determine stress from physiological data obtained with non-obtrusive devices.**
- **The methods applied to integrate driver's aspects in the driving scene analysis.**

2.1. Physical factors based on visual features

This section focuses on reviewing the methods which analyze the driver visually for detecting the characteristic mannerisms of their face and body associated with the driver impairment states. Accordingly, image processing and computer vision methods are used to detect the visible aspects of unfit driving conditions.

2.1.1. Face detection and face model representation

The analysis of physical factors highly depends on the accuracy of algorithms for locating the face and its distinct parts. The widely accepted Viola-Jones algorithm [33] was the first method that tackled the face location effectively. However, the accuracy of the Haar cascade depends on the face pose, thus requiring the deployment of an ensemble of several Haar cascade models to locate the face in the image regarding the face orientation. This fact does not avoid the detection blinks and the scale changes of the location area. Newer approaches based on deep learning, such as BlazeFace [34], provide high stability on face detection, independently of the face orientation variations, using limited computational resources.

Another approach of driver state analysis relies on the geometrical analysis of the facial structure. Deformable models are the foundation of the face representation model. The face shape is represented as a flexible 2D curve or a 3D surface that can be deformed to fit a specific expression of that face [35].

Deformable model-based algorithms such as active shape model (ASM), active appearance model (AAM), or Face Alignment require a landmark database to adjust their topology and fit the model according to the image feature at each point. Figure 2.2 shows the results of the Face alignment algorithm across different subjects and facial expressions, such as were presented by Gu and Kanade [36]. Nowadays, RGB-D cameras have improved the accuracy of deformable models because the depth channel allows gathering a 3D scene, thus aggregating another dimension to the plane color image. The depth information is obtained by infrared light. However, using depth information for the appearance model presents weakness when someone wears accessories such as glasses. Moreover, Machine learning object detection based on deformable models such as MediaPipe [37] has exhibited great accuracy and detection stability in real-time using RGB camera as input in good illuminated scenes. This fact reduces the access cost to this technology, thus facilitating the integration of more complex systems.

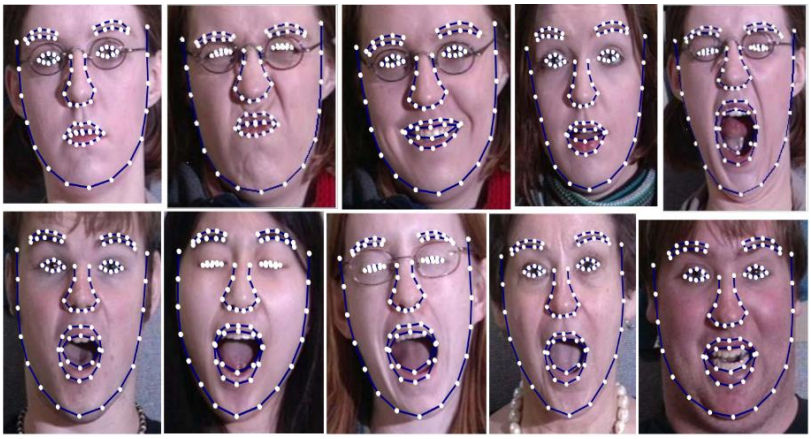


Figure 2.2. Facial structure model obtained by face alignment algorithm [36].

Face detection is the foundation of locating and analyzing distinct facial features. Following subsections are organized according to the facial features that present to mannerisms inextricably linked with unfit driving states.

2.1.2. Eye-related measures

Eye region analysis can provide a robust assessment of the mannerisms related to the detriment of driver's alertness. Eye-related measures, including the percentage of eyelid closure over the pupil (PERCLOS), eyelid distance, blink latency, blink speed pupil dilation, and gaze direction, were indistinctly used for the assessment of both fatigue [38], [39] and mental workload [40].

Widely accepted, PERCLOS was the pioneer measure that quantifies the driver's alertness from eyelid behavior analysis [41], and it relied on analyzing how slow the eye blinks are rather than only getting focused on the blink rate. Some examples of several proposed studies show the diversity of methods to estimate PERCLOS accurately using non-obtrusive visual-based systems. For instance, Mandal et al. [42] presented a framework to estimate the driver's fatigue score by continuously monitoring the eye openness level based on spectral Regression (SR), and machine learning methodology to estimate was applied to determine the eyelid closure. Instead, Darshana et al. [43] used pupil detection as an eye closure estimator. Lin et al. [44] trained a classifier based on Haar-like features by AdaBoost algorithm to determine the eye closure state. A recent study conducted by Zhang et al. [45] applied a convolutional neural network (CNN) to estimate the eye's aperture state.

Eye-Tracking signal (ETS) and pupil features can also provide assessment information about the driver's state. Saccadic eye movements are strongly related to visual fatigue. Catalbas et al. [46] estimate the speed and acceleration of the eye pupil using an IR-Camera. The peak detection of eye pupil acceleration in different driving conditions of low and high traffic density demonstrated that when a driver is fatigued, the rate of eye movements is significantly lower than in the normal state. Flores et al. [47] addressed the drowsiness detection independently of illumination conditions from the RGB-D sensor. The proposed approach applies the reconstructed 3D scene in color space to improve the accuracy of the driver face detection, and the extraction of the eyes regions of interest (ROI) was used for assessing the drowsiness index.

Despite having employed obtrusive devices, Xu et al. [48] about pupil dilation and gaze fixation in fatigue conditions was used for monitoring the eye. This study demonstrated how the pupil area decreases and the fixation time increases significatively when the fatigue state emerges.

2.1.3. Mouth behavior

Yawning is also a common visual symptom of fatigue and drowsiness. Saradadevi and Bajaj [49] used a Haar cascade model to locate the mouth and, subsequently, a support vector machine (SVM) to conduct a binary classification of yawning. However, this method is extremely sensitive to image rotation and perspective. Abtahi et al. [50] stated a method for identifying the mouth on the face image using a map transformation in the *YCbCr* color space and proposed a geometrical analysis of the detected area to determine the yawning. However, even this method shows results apparently, the method was not validated with a dataset.

Fan et al. [51] applied a Gabor filter to extract the texture features of the mouth corners and a Lineal Discriminant Analysis (LDA) to classify yawning and non-yawning examples. This technique exhibits more accurate classification results than geometric features-based classifiers.

From the perspective of machine learning, the existence of accessible databases is crucial. Abtahi et al. [52] gathered an extensive video collection of yawing examples, thus providing the research community resources to train and validate yawning models.

2.1.4. Face expressions detection based on facial features extraction

Another approach to identifying unpaired conditions on the driver relies on analyzing facial expressions using landmarks-based face representation models. Accordingly, Facial Action Coding System (FACS) [53] widely describes the facial behavior related to diverse emotions, and it is the foundation of some studies to identify the driver's fatigue [54]. Following this research line, Nakamura et al. [55] extracted the facial feature points using a face alignment algorithm and applied unsupervised learning to identify the commonest drowsiness face expressions and the transition period towards the fatigue state.

Face expressions can also provide information about the driver's emotional state. In this line, From driver's RGB images collected in real-time, Sesmero et al. [56] proposed an explanatory deep-learning-based model for detecting among eight possible driver's emotional states and the performing of ten activities both driving-related such as using the GPS or performing a safe driving and non-related driving activities. The presented results exhibit an accurate prediction, mainly when opposed emotions are elicited. Theagarajam et al. [57] proposed image processing based on applying the Gabor filter to remove image background removal to improve the classification performance of emotional facial expressions.

Face representation models can sustain the hypovigilance analysis as well. Benoit and Caplier [58] focused their study on the blink and yawning frequency, detecting the state of eye and mouth from the energy estimation of the image, resulting point-based face model obtained with the face alignment algorithm.

2.1.5. Driver Monitor Systems

Drivers monitor systems (DMS) were firstly introduced in industrial scopes such as mining, in which working conditions fostered by 24/7 shift roasters increase the risk of committing heavy machinery operators errors [59]. Gradually, DMS became a commercial solution with its application in vehicle fleets that cover a transportation service, introducing GPS and communication innovations that allow warning service supervisors about fleet drivers' state [60]. Currently, the DMSs are being introduced into the Tier 1 supply parts of automotive original equipment manufacturers (OEM) due to the need to improve the performance of ADAS.

OEMs currently offer solutions focused on detecting driver's impaired states and distracting actions through visual features analysis. For instance, Figure 2.3 shows the

interface of *Driver Sense* software-based system development by Cipia Vision Ltd. (Israel) [61]. *Driver Sense* offers a complete solution able to recognize among 13 features that describe the driver's state, such as head pose, eyelids, pupil dilation, mouth openness, and non-related driving activities such as smoking or phone usage.

Figure 2.3. Driver Sense software-based DMS of Cipia Vision Ltd [1].

Quectel Wireless solutions (China) provide similar solutions for determining the driver status in the Asiatic OEM market, measuring the driver's attentiveness through gaze tracking, eye blinking, and head movements analysis [62].

In the scope of academic research, studies focused on tackling the weakness of face detection employing tight budget vision systems. For example, Peláez et al. [63] proposed the reconstruction of a 3D color mesh of the face from an RGB-D sensor to address the illumination issues that strongly affect the stability of the face detection model deployed in the driving domain.

2.1.6. Take-home messages

The take-home messages of the literature review concerning the visual analysis of driver's features suggest that:

- The redundant based on different vision systems technologies can improve the global stability of detection systems.
- The accuracy of face detection and face feature location methods are crucial to conducting a precise evaluation of the driver's physical state.
- The analysis of visual features can approach the assessment of emotions and cognitive states that potentially reduce the driver's attentiveness, even though the origin of the driving skills impairment is not related to physical unfit states.

[1] Retrieved from https://www.automotiveinteriorsworld.com/news/sensors/eyesight-technologies-teams-up-with-globoconnect-for-driver-monitoring-system.html

2.2. Psychological factors based on health indicators

Stanton, Young, and Walker argued the concern of analyzing the psychological aspects of the driver. Among the psychological factors, the different shapes of stress can alter the attention capacity, having the hypovigilance as the most direct consequence and colloquially resume with the sentence "*I was seeing, but I was not watching*" [64]. Lazarus and Folkman (1984) defined the concept of stress as follows:

> "*a pattern of negative physiological responses occurring in situations where people perceive threats to their well-being which may be unable to meet.*"

According to American Psychological Association (APA), stress can classify into three types (acute stress, episodic acute stress, and chronic stress) regarding their characteristics, symptoms, duration, and treatment approaches [65].

Acute stress caused by sudden events potentially hazardous during the driving task can harm the driver's aptitudes. Short-term symptoms, including trembling hands, muscle spasms, palpitations, and rapid pulse, can adversely affect decision-making.

It is necessary to understand how the human body reacts against environmental stimuli. The autonomic nervous system (ANS) handles the involuntary control of glands, cardiac muscles, and smooth muscles located on the walls of hollow viscera, the airways, the blood vessels, the iris, the ciliary body of the eyes, and arrector pili muscles of the hair follicles. The ANS consists of the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS), whose activity is balanced and dynamically regulated by the homeostasis process. Against a hazard, the human body reacts with a "*flight-or-fight*" reaction where the sympathetic activity dominates. This alteration in the body triggers the "*rest-and-digest*" functions autoregulation mechanism in which the body actuates, releasing biological processes to recover the normal indices [66], [67], [68]. However, the body response against a stimulus is a subjective factor because stressor does not affect all persons in the same way. Prior research substantiates the belief that certain psychosocial factors promote developing stress resilience, thus enhancing the adaptation ability to cope with the different shapes of stress [69].

The understanding and knowledge about how stress resilience is constructed can describe which traits are more likely vulnerable to stress, and these profiling methods are specially used when the task deployment involves prologued exposures to hostile environments [70].

Stress detection in real-time domains is still an open question addressed by the multidisciplinary field of Affective Computing [71], whose primary purpose can resume as the understanding and recognition of the emotional states that a person goes through. This research field involves different disciplines, including signal processing, feature extraction, machine learning, statistical analysis, and, in more general terms, cognitive science.

The collection of inputs for determining emotions covers a broad and diverse range of information sources such as psychological self-assessment questionnaires, information purely physiological (e.g., heart rate, electrodermal activity, electroencephalogram), the

different expressions of interaction and communication (e.g., facial expressions, postures, body gestures, voice tone, mouse events, text and sentiment analysis) [72].

This section explores stress detection through physiological data analysis. This state-of-art review only contemplates the data gathered with sensors of non-intrusive and wearable devices. Accordingly, heart measurement and electrodermal activity foundations are exposed, delving into their most relevant aspects and technology limitations. Finally, the influence of physiological signals is evaluated in the driving context from the perspective of health monitoring, exploring its analytical, technical, and technological aspects.

2.2.1. Feature extraction from physiological signals

As was explained in the introduction, heart rate and electrodermal activity are the most used indicators for determining emotional changes. However, these digital signals produced from analogical sensors require a pipeline of transformations to extract useful information to conduct the posterior analysis.

It is necessary to distinguish between the physiological measurement regarding resources and technology employed its collecting. Accordingly, two main groups are distinguished:

- **Laboratory equipment**, more accurate technology but as well very intrusive. Electrocardiograph (ECG), Electroencephalograph (EEG), and Electromyograph (EMG) are some examples of medical equipment that converts electrical activity by electrodes into visual records.
- **Wearable technology** is opposed to the previous one, i.e.; it can be used in daily life; however, their resolution is lower. Currently, the electronic market offers a wide diversity of health devices, including smartbands and smartwatches, having a high acceptance by consumers.

Once differentiated between both kinds of technologies, thereafter, only physiological parameters collected with wearable systems will be covered, especially wrist-worn wearables. Moreover, physiological studies can be distinguished according to the signal recording period taken for the analysis. Accordingly, Shaffer et al. [73] had distinguished ultra-short-term (recordings lesser than 5 mins), short-term (recording of 5 mins approx.), and long-term (recordings equal or larger than 24 hours).

Blood volume pulse (BVP) and pulse rate variability (PRV)

Photoplethysmograph (PPG) is an optical technology instrument for measuring the changes of blood volume in the microvascular bed of tissues [74]. PPG waveform consists of two superimposed components: a non-pulsatile "direct current" (DC) component and a pulsatile "alternating current" (AC) component, the latter being the carrier of blood volume information within the sample [75].

Blood volume pulse (BVP) is the waveform obtained by filtering PPG, the high pass filter arbitrary set between 0.05 Hz and 0.5 Hz, the most popular choice. Nonetheless, the

election of filtering bounds depends on the environment where the signal is gathered and the possible heart rate limits.

BVP signal allows determining information about the heart rate, extracting the time interval between two consecutive heartbeats. Heart rate (HR) is the speed measurement of the number of heart contractions (beats) per time unit expressed in beats per second (BPM).

Also, Inter-beat interval (IBI), also known as beat-to-beat interval, is the elapsed time between two consecutive beats expressed in msec. Both measures are alternatives and used indistinctly as the formula converts IBI in BPM (2.1):

$$BPM = \frac{60{,}000}{IBI} \tag{2.1}$$

Accordingly, P-P intervals are the time series composed of the elapsed time between pulse within a specific period. This terminology is derived from the terms R-R intervals and N-N interval (N indicates normal cardiac time free from artifacts) employed in ECG analysis. Their close relation can be confused and can induce to use of these terms indistinctly; however, their interpretation is different due to physiological data being collected with absolutely different technologies. After clarifying this aspect, for a period of *t* seconds, the P-P intervals are defined as (2.2):

$$PPI = (pp_1, pp_2, \dots, pp_n) \tag{2.2}$$

Figure 2.4 depicts obtaining the PP series assuming that BVP is a free-artifacts signal. As can be observed, the location of the different heartbeats on the BVP signal allows extracting all IBIs across the heartbeats, thus producing transformation of an electric magnitude into a temporal magnitude. As a result, a deeper study about how the NN intervals behave conducted in a specific period can provide a more accurate physiological indicator of a person's state.

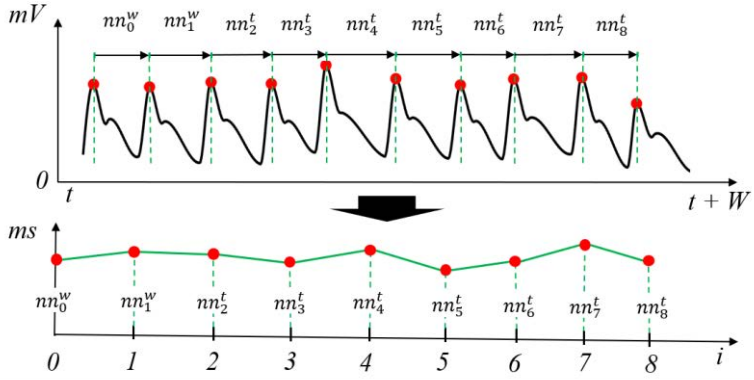


Figure 2.4. Obtaining process of PP Interval series (bottom plot) from a BVP waveform (top plot) in which arterial palpitations (NN) are located and elapsed time between them are extracted.

The study of PP intervals is based on the extraction of the pulse rate variability metrics (PRV) that describe pulse fluctuations around the mean HR in a specific period. The PRV metrics are identically computed as heart rate variability (HRV) metrics that proceed from ECG. Shaffer and Ginsberg [76] broadly reviewed all well-known HRV, classifying the

distinct features according to the metric nature, including the function domain (time, frequency, lineal, non-lineal, and geometrical) and the term of the analysis (ultra-short, short, and long terms). Besides, this review analyzes which HRV was applied in different scopes, such as arousal eliciting experiments and high-performance sports training.

Currently, there is an open discussion concerning whether the findings related to specific patterns of HRV with SNS activity are transferable to PRV metrics. The controversy about PRV misuse is still ongoing because both smartbands and smartwatches have scaled in the electronic devices market in the last years. Accordingly, multiple studies questioned whether PRV could surrogate HRV or not, stating a clear distinction between both biomarkers. For instance, Shäffer and Vagedes [77] inspected studies that evaluated the accuracy of PRV as an estimate of HRV, independent of the technology used. The findings of this review highlighted the acceptable accuracy of PRV in calm conditions such as sleeping and resting, underlined the effects of cardio-ventilatory coupling in short-term analysis and, disproved the PRV usage for assessment conditions such as physical activities and some mental stressors because its concordance with HRV is regularly and seriously compromised. Yuda et al. [78] corroborated some aspects pointed out by Shäffer and Vagedes; however, they concluded that PRV and HRV should be considered as separated biomarkers, suggesting that physiological properties of PRV should be investigated. Moreover, some studies evaluate PRV in different scenarios to constitute the scientific evidence to accept or refuse PRV as a reliable index of ANS activity. For instance, Mejia-Mejia et al. [79] investigated which differences saw PRV and HRV during a fully-body cold exposure stress test and, as a result, most PRV metrics increased their values during the test.

Despite the discussion exposed above, it is relevant to mention the findings that allow interpreting the relation between HRV and mental stress. Prior research generally confirms that reliable stressors decrease the value of HRV metrics, except for some metrics that react inversely [80][81] and, the SNS and PNS activity can distinguish in the spectral plane, thereby being able to identify the contribution of both in the regulation process.

However, PRV analysis has associated some issues that can negatively impact the interpretation of these metrics. First of all, the dislodging of the sensor due to motion artifacts (commonly known as artifacts) can distort the entire or partially the PPG signal and consequently impede the correct obtention of NN intervals. Some studies propose approaches based on reconstructing the contaminated epochs of the signal [82]. Moving average, low-pass filter, and Kalman filters are the addressed methods for solving this issue, such as was reviewed by Tamura et al. [83].

Another major issue concerns the common presence of ectopic beats on the signal, which considerably influences the distinct kinds of PRV metrics [53][54]. Ectopic beats are caused by excessive activity of the sinoatrial node and are identifiable as two successive irregular beats (a premature beat followed by a longer than normal beat). Consequently, handling the detection and correction of this kind of discrepancies on the PP intervals is still an open question with different proposes for solving this issue [86], [87], [88], [89].

Electrodermal activity (EDA)

Electrodermal activity (EDA) refers to the alteration of the electrical properties of the skin due to the secretion of eccrine sweat glands [90] as a result of SNS activity, being one of the most broadly used biosignals in psychophysiology [91]. Dawson et al. [92] explained the story behind this phenomenon discovered by Fèrè (AD 1888), who stated that skin becomes a better electricity conductor in the presence of external stimuli. After that, Tarchanoff (AD 1890) demonstrated that electric potential changes were measurable lodging two electrodes over the skin, thus establishing the methodology applied to quantify and record this physiological signal at present.

Nowadays, one of the discussions resides in determining the better location of the electrodes for measuring emotional arousal. Palms and soles contain the highest concentration of eccrine sweat glands (between 600 and 700 glands/cm2) [93]. However, some studies such as was conducted by van Dooren et al. [94] have demonstrated the lack of correlation between the density and the activity of sweat glands and the emotional responses, identifying forehead, thighbone, shoulders, and wrists as skin areas highly correlated with the electrodermal activity manifested on fingers.

EDA signal carries two superimposed components: the tonic activity called electrodermal level (EDL) is a slow-changing associated with activation state, and the phasic activity called electrodermal response (EDR) is fast-changing and sensitive to stimulus newness, strength, and significance [92].

Several studies address separating the EDA components explicitly. Greco et al. [95] proposed an algorithm based on a convex optimization model to decompose the EDA in both its components. Prabhakaran and Ayyamperumal [96] exposed the computational complexity issues of methods previously reported, including convex optimization model, sparse decomposition model, and compressed sensing-based decomposition (CSD). They transformed the EDA component separation into a convex-cone optimization problem to improve the accuracy and performance of previously reported methods. Anusha et al. [97] defined a method to process small time windows based on the least-square regression line to remove the trend produced by the superposition of EDL, thus isolating the EDR signal.

Among the techniques employed to detect EDA signal stress patterns, feature extraction based on descriptive statistics is the most applied methodology. A systematic review was conducted in Web of Science (WoS) to identify which EDA features can be represented in the scope of stress detection with wearable devices.

The review encompassed the papers within indexing years 2010 and 2019 that matched with searching string: *(("EDA" or "electrodermal activity" or "galvanic skin response")) and ("stress " and ("detection" or "recognition")) and ("feature extraction" or "features") and "wearable"*.

As a result, 29 publications were identified after primary research by WoS. 2 papers were refused because the experimental equipment employed was not wrist-worn devices, and it was too invasive for real integration in a vehicle.

Table 2.1 outlines the most representative descriptive statistical features extracted from the EDA signal. They were found by a literature review regarding the studies that conducted

an analysis based on feature selection for determining arousal and stress. Within the 27 resulting papers, the features were selected and categorized according to EDA components.

TABLE 2.1. EDA DESCRIPTIVE STATISTICAL FEATURES.

Signal	Feature	Unit	Description	Ref
EDA	*EPC*		The rate of positive changes of SC per time interval.	[98]
	avgEDA	μS	Mean skin conductance	[17]
	quant10 quant25 quant50 quant75 quant90	μS	Quantile thresholds values (10%, 25%, 50%, 75%, 90%) inferred from SC signal.	[99], [100], [101]
	IQR	μS	Interquartile range of signal	[99]
	Crest		Crest factor of the EDA	[18]
	varEDA	μS	Variance of SC (equivalent to Hjorth activity)	[102]
	stdEDA	μS^2	Standard Deviation of SC	[102], [99]
	SumNegDiffSC		Summation of positive values of 1st derivative of the signal	[102]
	SumPosDiffSC		Summation of positive values of 1st derivative of the signal	[102]
EDL	*Slope*		Linear regression slope	[99], [100], [102]
	Avg, min, max	μS	Average, minimum, and maximum EDL	[100], [18]
	rangeEDL	μS	Difference between maximum and minimum EDL	[103]
	ctEDL		Ratio of range to absolute mean	[103]
	cvEDL		Ratio of standard deviation to mean	[103]
	Intercept		Linear regression value of y given x=0	[99]
EDR	*EF*	Peaks/s	Frequency of SCRs	[98], [17]
	PowSCR		Power of responses	[17]
	MaxPeak	μS	Magnitude value of maximum peak	[19]
	SkewnessAmpPeak		Skewness calculated for amplitudes of local extremes	[19]
	KurtosisAmpPeak		Kurtosis calculated for amplitudes of local extremes	[19]
	MaxDeltaBackward MaxDeltaForward	μS	The maximum difference between the current local extrema and the amplitudes of the earlier and later signal time window, respectively.	[102]
	RangePeak	μS	Difference between maximum and minimum EDR	[19]
	StdAmpPeak	μS	Standard deviation of local extremes	[19]
	VarAmpPeak	μS	Variance of amplitude local extreme values	[19]

Nevertheless, the analysis of EDA presents a principal issue concerning arousal generalization. Brathwaite and Watson [104] highlighted the subjective character of the arousal, showing how the exhibiting of the same set of stimuli across several individuals produces a wide variety of scales in the electric conductance measurement. This fact supposes a challenge when the study compares arousal elicited by the same stimulus across individuals and the model convergency based on inductive learning techniques from these

data. In this case, classic data normalization methods are not applicable, mainly when the dataset contains outliers. Accordingly, different transformations were proposed to become the EDA recordings comparable across several individuals.

In this paper, Brathwaite and Watson concluded that standardization methods such as z-score transformation and ratio transformation equate the sample variability, thus maximizing the correction of responses magnitude across individuals and refusing to apply a logarithmic transformation over SCR as a valid normalization method.

Another related issue is the noise caused by motion artifacts. Artifacts could corrupt the signal, thereby introducing false SCRs interpretable as event-related SCRs. Among the technique intended to solve this issue, frequency filtering is the most applied approach.

2.2.2. Methods for the stress assessment

Widely accepted psychometric instruments are extensively employed in stress detection studies. In the psychophysiological domain, stress perceived can suppose the only way to establish a correlation between events and phenomena reflected in physiological data. In this way, self-assessment questionnaires provide objective information from personal opinions.

The stress state caused by people's life events can be quantified. The American psychologist Sheldon Cohen devised Perceived Stress Scale (PSS)[105]. This test allows rating how stressful are someone's life situations in a mid-term. Other psychometric Social Readjustment Rating Scale (SRRS) or Holmes-Rahe Stress Scale [106] is aimed to quantify the long-term stress caused by 43 life events called "Life changes Unit" (LCU) experimented in the last two years. These self-assessment questionnaires state a score that quantifies a person's stress with a numerical value.

Event-related questionnaires allow obtaining a self-assessment about how stress was perceived during the stressor exposure. Healy and Picard employed a Likert scale to rate the driving events. Other studies get focused on assessing the stressor exposition by widely accepted stress indicators. For instance, Castaldo et al. [107] evaluated the ultra-short-term HRV metrics with HRV metrics extracted from 5 mins ECG signal excerpts, widely demonstrated stress indicators.

More advanced studies conducted in clinical environments employ biochemical analysis to determine stress elicitation. Anusha et al. [101] employed salivary cortisol concentration samples taken between the experimental process to determine changes in the sympathovagal activity.

2.2.3. Driver health state monitoring

Stress acute episodes and strokes caused by coronary failures and diabetes are the main challenges of impairing health aspects in driving. Detecting drivers' states from a physiological perspective encompass analytical, technical, and technological aspects.

Interpretation of physiological data in driving context

Studying which physiological measures can provide the most reliable information about the driver's health state is fundamental. Lohahi et al. [108] reviewed how the emerging cognitive states affect the driver's ability to explore and analyze the driver's collectible psychophysiological measures (intrusive and non-intrusive). Studies of this kind facilitate the understanding of the relationship between arousal states and how indicators behave, thus resulting crucial to assess the reliability of using specific sensors and the cost-benefit of their integration in the car cockpit.

Stress detection models in the driving context

From a technical point of view, detecting poor health conditions in the driving context is still an open issue. In 2000, Healy conducted the pioneer study on measuring drivers' stress levels by various sensors to record physiological signals during driving tasks [109]. Despite using laboratory equipment to collect the data, this study stated stress detection in a driving environment as a research field. Further, Healy and Picard [110] provided a public access database of stress elicited in different driving conditions, and multiple stress detection studies employed this database to validate their proposed approach.

Rastgoo et al. [111] reviewed the classification techniques addressed to predict driver's stress levels employing multiple data sources, including physiological sensors, vehicle dynamics, and individual information. EDA, EMG, ECG, and respiration belts employed physiological sensors. However, employing obtrusive sensors is not realistic because this required instrumentation does not fulfill the safety and comfort standards despite recent technological advances. Although the spectrum of machine learning methods was widely covered, all reviewed papers in 2000 and 2018 share feature extraction based on descriptive statistics and features selection as data mining methods to identify stress patterns. This survey presents an uncovered aspect concerning how studies approached the treatment of individual differences or whether otherwise models were trained individually, thus lacking generalization ability.

Technological advances in cockpit instrumentation

Technological advances in cockpit instrumentation sustain the possibility of assessing the driver's health condition. Subramaniyam et al. [112] reviewed the studies that address the integration of IoT sensors in the car cockpit intended to monitor driver's healthiness and enhance driving comfort. The sensorization of elements that are constantly touched or in contact with the driver, including the steering wheel, seat belt, and seats, is proposed to integrate non-obtrusive sensors such as EDA, ECG, respiration frequency, and body temperature.

Nowadays, main car trademarks are starting to embrace new solutions for monitoring the driver's health status. However, trade secret impedes obtaining specific information about innovative designs that tackle health sensor instrumentation in cars, limiting the communication of these advances to short press releases in specialized publications. Ford is investigating the incorporation of capacitive plates in the driver's seat to capture the heart

electrical signals, actuating as an ECG sensor [113]. Toyota is developing an integrated system for detecting heart anomalies such as arrhythmias and atrial fibrillation [114].

Figure 2.5. Driver's seat prototype that integrates an ECG sensor designed by Ford [2]

2.2.4. Take-home messages

- Non-obtrusive health devices can provide a vast amount of information about a person's emotional state.
- The open discussion about whether PRV can surrogate HRV impedes the reliable usage of PRV as a marker of sympathetic activity until it is scientifically evidenced.
- EDA is a reliable indicator of sympathetic arousal; however, individual differences and motion artifacts suppose a handicap for approaching general models.

2.3. Driver ability assessment in driving safety models

In the previous subsections, the analysis of driver states was treated regardless of how it can be integrated with other safety models. This subsection tries to summarize the employment of data fusion, including driver's data and driving scene details, as the basis of the development of new safety applications.

2.3.1. Data fusion models

Data-fusion based safety systems are studied in many contexts, which can be syncretized as follows:

- **Detection of driver's impaired states regarding environmental events.** Concerning data fusion of driver's physical aspects and environment information, Following this research line, Jabon et al. [115] proposed a facial analysis approach based on time and frequency feature extraction from face landmarks combined with vehicle surrounding information to predict severe traffic crashes tested in a driving simulator system.

[2] Retrieved from https://spectrum.ieee.org/3-ways-ford-cars-could-monitor-your-health)

Fletcher and Zelinsky [116], and posteriorly Allamehzadeh and Olaverri [117] built their reasoning model by correlating the driver's eye gaze with environment events such as the pedestrian presence to determine the driver's attentiveness.

Leonhardt et al. [118] expose a probabilistic model which includes driver's gaze, vehicle dynamics, and controls to predict the driver's behavior directly related to the lane-changing maneuver.

Moreover, several studies combine the driver's physical and physiological aspects to evaluate the driver's wellness. For instance, Lee and Chung [119] determined by a fuzzy Bayesian network the drowsiness state is that evaluates the variations of driver's eye state, electrocardiogram (ECG), photoplethysmography (PPG), temperature, and the vehicle speed. Rigas [120] proposed a Bayesian network to establish a relation between physiological measures collected using non-obtrusive wearable devices and the driving events.

- **Assessment of driver's inspection of automated driving functions.** The coming of conditional automation (3rd SAE level) currently arouses different perspectives to assess the driver's state during the take-over-request (TOR). Gold et al. [121] described the driver's take-over process as a sequence of several stages which transit from recovering the cognitive information process until acting. The last research field studies, eye gaze, and movement are used for the driver's attention assessment to lead the take-over acceptance after conducting a non-driving related task [122][123].

2.3.2. Knowledge representation

As contemplated in previous sections, driver-related information can integrate very diverse aspects. In the context of data-fusion-based approaches, the representation of heterogeneous data, including environment, driver, and car aspects, is essential to design new and more robust safety applications.

In this way, employing knowledge-based systems (KBS) in driving safety allows dealing with the complexity of this domain, unifying all aspects of a driving scene according to the abstraction of a semantic network of concepts. In addition, KBS enables the usage of rule-based systems and an inference engine in the functioning of safety, thus allowing the implementation of complex behaviors. KBS has been employed to tackle issues such as inferring the driving context [124][125], coordinating HMI information with monitoring aspects [126] and, assessing the operational driving domain requirements of automated driving functions[127].

In the scope of driver's state representation, the conceptualization of physical and cognitive states are part of the driver entity. For instance, the *ISI-PADAS* project defined an ontology-based representation to contemplate all aspects of driver behavior and their influence on the driving task [128]. In this case, driver conceptualization encompasses driver profiling properties such as gender, age, and driving distance, among others; driver perception aspects according to their visual perception, the driver actions (steers, gear shifts,

and braking, among others), and cognitive driver states such as workload, distraction level, situation awareness, and fatigue.

2.4. Legal and regulatory framework

It is important to highlight that driver monitor systems must fulfill a legislative framework that regulates their design aspects.

In European Union, Regulation (EU) 2019/2144 of 27 November 2019 on "*type-approval requirements for motor vehicles and their trailers, and systems, components and separate technical units intended for such vehicles, as regards their general safety and the protection of vehicle occupants and vulnerable road users*" established the requirements of systems intended to assess the driver availability [129]. Article 6 states the requirements of advanced vehicle systems, including driver drowsiness, attention warning, and advanced driver distraction warning systems. For these cases, the time limitation for storing and accessing the recorded data is the primary requirement that the driver monitor system must fulfill.

2.5. Summary

This chapter exposed the distinct aspects in which human factors are involved in driving. The reviewed state of art supplies a general view of the challenges and the research gaps of integrating the human factors facets in the driving context, its acceptance degree by car trademarks and OEMs and, its nearness to becoming part of private cars standard equipment.

The advances in research about driver monitoring based on visual features analysis have been substantial and, the incorporation of cars is a palpable reality. However, the driver's physiological analysis is still an open question. The multimodal approach for detecting cognitive states could not be the most appropriate when the car cockpit integration is the primary purpose. Future research directions should be oriented to employ non-intrusive devices to monitor the driver's health, proposing reliable techniques involving fewer sensors. Further, stress detection suggests exploring processing methods not related to extracting statistical features.

3. Methodology for collecting and labeling maneuver data

The study of Driving Behavior Modeling (DBM) involves multiple aspects concerning driving activities recognition. Among driving-related activities, the incorporation of predictive models of maneuvers and driver intent in advanced driving assistance systems allow their performance improvement in terms of safety, ergonomics, and driving experience.

Several cognitive models stated the drivers' behavior according to a structure based on various levels back in the day. Michon (1985) proposed a general model in which the road user task is hierarchically structured. In that structure, the driving task is composed of three coupled levels of skills and control: strategic (planning), maneuvering (tactical), and control (operational) [16]. The external outputs of these three levels are respectively: general plans, controlled action patterns, and automatic action patterns. Fifteen years later, Christ et al. [130] introduced the GADGET-Matrix wherein the behavioral level was added at the top of that hierarchical model previously stated by Michon. This level refers to personal preconditions and characteristics, and it has the highest priority because such dispositions heavily influence driving decisions at lower levels. Figure 3.1 depicts the hierarchy of these three levels of actions and how the driver's personality traits and current state govern them.

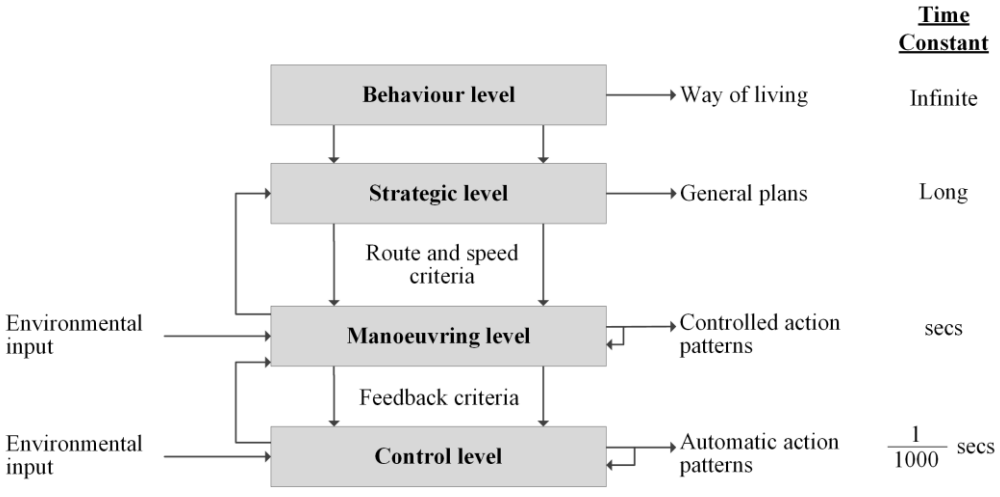


Figure 3.1. The hierarchical structure of the driving task [131].

Panou et al. [131] manifested the influence of these theoretical foundations of driver behavior in designing new safety functions. Abu Ali and Abou-zeid reviewed the driving applications at each level of driving task, providing some insights about how the car's safety

and intelligent transport systems are evolving due to increasingly complex driver models [132].

Currently, driver models are intended to handle vehicle control aspects, and their diverse applications cover all levels of driving-related tasks. Concerning the operational level, the modeling of immediate actions, such as stopping and emergency braking, intends to comply with traffic regulations or ensure driver and passengers' safety [133]. At the tactical level, models are aimed to predict the set of maneuvers that accomplish short-term goals in real-time. This improves ADAS functions, compensating vehicle trajectory to avoid possible hazardous and conflicting situations [134]. Driver's intent prediction employs the driving controls, exteroceptive sensors that provide the vehicle's surrounding information, and driver monitoring data such as driver gaze and hands positions as a part of models inputs. At the strategic level, driver models are focused on satisfying the driver's preferences to complete the route plan. Dai et al. [135] explored techniques for selecting the travel trajectories that minimize the travel cost according to several criteria aligned with driver's preferences. In addition, the arrival of communication technologies with V2I protocols facilitates addressing the driver model at the strategic level to improve safety and mobility in high-density traffic flows [136]. At the behavioral level, driving models address to predict misconducts that affect driving safety and fuel consumption, thus minimizing human error accidents and improving environmental health. For instance, Hong et al. [137] employed data extracted from smartphones, including maximum, average, and standard deviation of speed, speed change, longitudinal/lateral acceleration to characterize the driving style and, thus, detect aggressive driving behaviors. Zheng et al. [138] studied how driving style affects pollution emissions through acceleration and deceleration behaviors. Moreover, in the scope of efficiency of electric vehicles, Varga et al. [139] propose an estimation of the state of charge according to the driver's aggressiveness regards how the acceleration pedal is operated.

About maneuver detection, stopping, braking, left and right turnings, lane changing, and lane-keeping is set the maneuvers that conform as the detection targets and, two detection perspectives can be distinguished as was reviewed by Doshi [134]: (1) Single target prediction (2) multi-target prediction. Maneuver detection is mainly addressed through inductive learning techniques, being Hidden Markov Models and clustering the most applied techniques.

All aspects mentioned above highlight the necessity of a methodology for collecting driving maneuvers that allow training a set of predictive models, thus leading any inductive learning process.

This chapter describes the experimental protocol designed for conforming maneuver datasets using a driving simulator environment. This first approach of recognizing driving activities requires inspecting the different elements involved in the vehicle control, including the steering wheel, pedals, direction indicators, and the engaged gear.

Studies published by Škrjanc et al. [140] and, after by Andonovski et al. [141], employed the data collected through the proposed methodology for addressing the recognition of specific maneuvers based on clustering techniques. Both studies focused on detailing which nested structures compose driving-related activities.

After this introduction, remain chapter is organized as follows. Section 3.1 describes how scenarios for collecting maneuver data were designed. Section 3.2 specifies how the dataset is structured, which driving controls signals are used, and the vehicle dynamics parameters. Next, Section 3.3 overviews the experimentation conducted from this dataset. Section 3.4 displays the resulting dataset graphically and exposes the reported experimental results of detecting maneuvers. Section 3.5 summarizes the findings regarding the collected data.

3.1. Maneuvers catalog and ad-hoc scenarios modeling

Different driving-related actions concerning operational functions can be distinguished according to the performing regularity along the route. For instance, the number of stops during the route is a common task that can be repeated dozens of times until achieving the destination. However, specific maneuvers such as three-point turning are necessary when committing routing mistakes. Another classification relies on the complexity of the maneuver concerning the number of operators on the driving controls that the maneuver involves. Table 3.1 describes the maneuvers contemplated in this study as follows.

TABLE 3.1. DESCRIPTION OF THE COLLECTED MANEUVERS.

Maneuver	Performing circumstance
Stopping	Arrival to a red traffic light, stop signal or traffic jam with stopped cars.
Overtaking	Left lane is empty and leading vehicle running at lower speed.
Left/Right corner turns	Following the route plan to reach the destination
U-Turn	Change direction in a wide roadway due to a wrong decision on the route plan.
Three-point turn	Change direction in a narrow roadway due to a wrong decision on the route plan.

According to the maneuvers previously described, several driving scenarios were designed to reproduce the circumstances wherein each maneuver must be performed. Thus, the design of each specific scenario is focused on the repetition of its coupled maneuver exclusively.

3.1.1. Stopping

The scenario was designed to accomplish stopping requirements strictly. For that, a scenario with an intersection regulated by a stop vertical sign and a stop marking conforms to the most suitable circumstance. Other situations such as presenting either a pedestrian crossing or a traffic light on red can arouse other drivers' actions to avoid stopping the car, especially in experimented drivers that include fuel consumption optimization in their behavior patterns. Figure 3.2 illustrates how the drivers must stop the vehicle when approaching an intersection with a stop sign. Moreover, including cross traffic introduces the need to accomplish this short goal safely and increases the scene credibility to obey the stop signs.

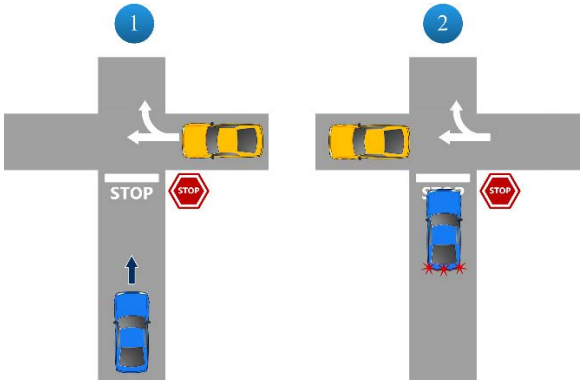


Figure 3.2. Stopping scenario based on intersection approach (blue car: ego-car; yellow car: cross-traffic).

3.1.2. Overtaking

The preconditions to perform an overtaking maneuver introduce the interaction with the following vehicle. In urban area traffic regulations and two-lane roads, the overtaking decision is driven by the reduced speed at which the leading car circulates. Figure 3.3 illustrates how the driving scenario was designed to fulfill the preconditions of overtaking maneuvers. As can be observed, multiple actions, including rear mirror inspections, speed increasing, and lane changes, compose this tactical maneuver. In this design of preconditions, vehicles circulating on the left lane were included to avoid the drivers not inspecting the left rear mirror at the beginning of the maneuver.

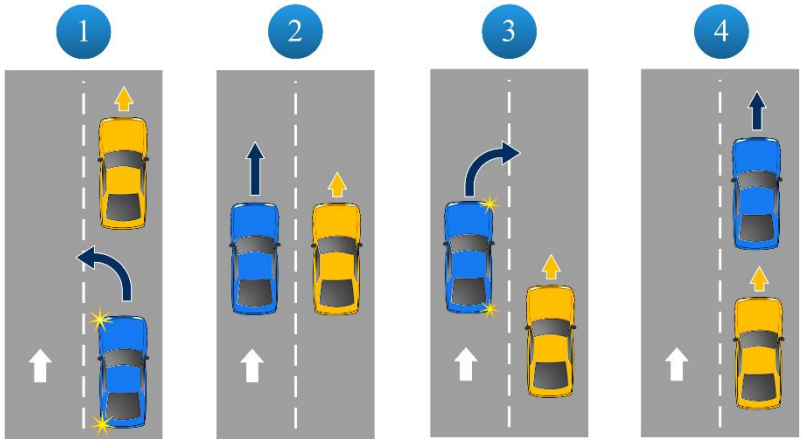


Figure 3.3. Overtaking scenario modeling (blue car: ego-car; yellow car: leading car).

3.1.3. Corner turns

Designing scenarios for collecting turning maneuvers involves strategic decisions to accomplish the route. In this case, route planner instructions lead to the performance of turning maneuvers. Figure 3.4 shows how the route planner triggers turning at the next road junction. The scene is always presented in a green traffic light circumstance where cross-traffic is stopped. Thereby, turning maneuvers are completely isolated from other actions. Otherwise, yielding actions would be included, increasing the complexity of collected maneuver data. Therefore, both left and right turnings were designed in the same way.

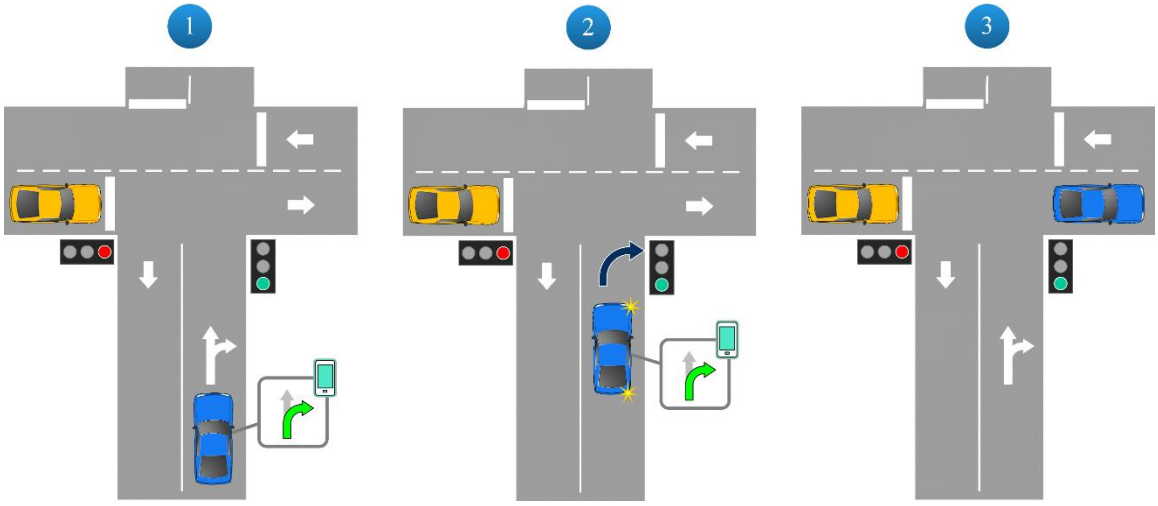


Figure 3.4. Right cornering after route planner instructions (blue car: ego-car; yellow car: cross-traffic).

3.1.4. U-Turn

U-Turns are specific maneuvers of urban and suburban areas. Figure 3.5 shows a U-turn being performed on a one-lane road. It relies on performing 180 degrees turning to reverse the direction of travel. The precondition of performing a U-turn involves the total width of the road platform. Notice that the roadway must be wide enough to complete this maneuver without stopping the car.

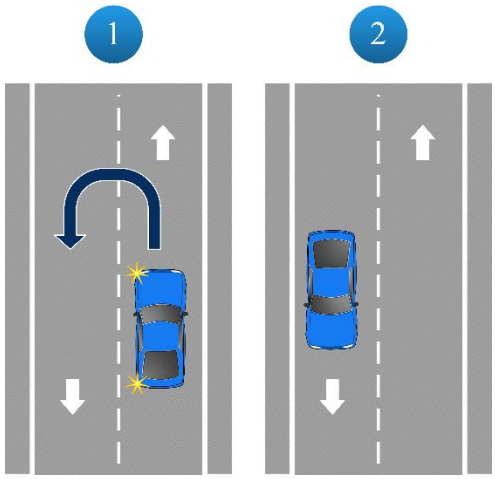


Figure 3.5. Performing of U-turn maneuver in a one-lane road.

3.1.5. Three-point turn

A three-point turn is a maneuver required when the road platform is too narrow for performing a U-turn maneuver. Like the U-turn maneuver, three-point turnings are usual in urban and suburban areas, aiming to reverse the direction of travel. Figure 3.6 depicts the actions which compose this maneuver, including steer to the right approaching the right shoulder, steering to the left in reverse towards the left shoulder, and finally, steering to the right to continue the route.

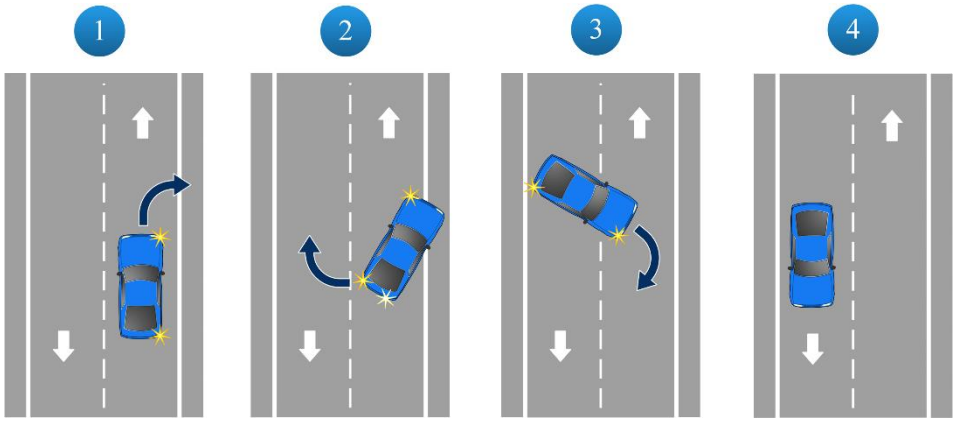


Figure 3.6. Performing a three-point turn maneuver in a one-lane road.

3.2. Experimental protocol

The experiment was deployed on the driving STISIM Drive simulator [142]. An experiment session consisted of around 30 minutes of driving in each of five designed scenarios for collecting the maneuver. Before starting each scenario, drivers were informed about the driving task that they had to perform to complete the challenge successfully. During the experiment running, a laboratory technician must supervise the driving to identify when the target maneuver starts and finishes each time it is performed. For that, keystrokes on the keyboard were used to create marks surrounding each maneuver instance. In this case, the mark is an integer value that starts with the value 0 and, it increases its value each time the designated key is pressed down.

For this first approach, six male drivers of 34.33 ± 7.28 years (mean ± SD), with at least one year of driving experience and different nationalities (Spain (3), Slovenia, Panama, and Colombia), agreed to participate in the experiment sessions. Dated drivers were informed about collecting driving controls data and signed a consent form for the storage and future data usage (see Annex II).

Figure 3.7 shows how the dataset is structured after the data collecting process. Once all drivers had conducted the driving test experiment, the dataset was preprocessed preliminary to adequate the data to the real-world domain.

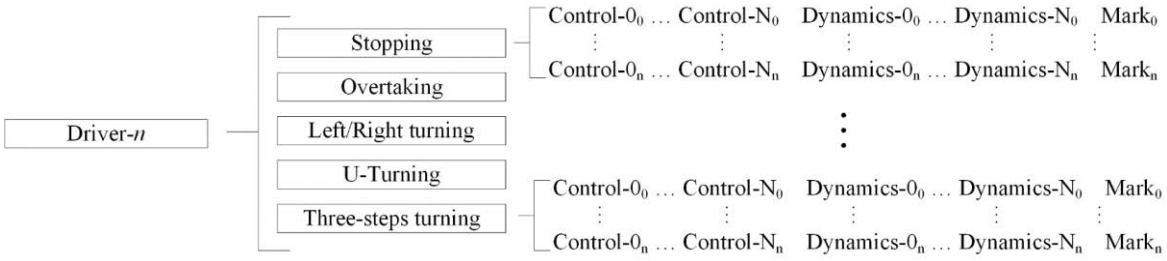


Figure 3.7. Structure of the collected dataset concerning maneuver.

Accordingly, information relative to vehicle controls was normalized to decouple the data from the own values of the used electronic systems.

3.3. Experimentation

As mentioned in the introduction, studies conducted by Škrjanc et al. [140] and, after by Andonovski et al. [141], utilized this dataset to produce a hybrid detection model based on evolving cloud-based algorithm and knowledge base for detecting the maneuvers in a multi-target way.

The study conducted by Škrjanc et al. [140] stated the foundations of the research line after followed by Andonovski et al. In this study, it was defined the process of acquisition that includes fundamental parts such as data preprocessing and the hierarchical taxonomy that includes atomic actions, tasks, maneuvers as the distinct complexity levels that a specific maneuver can be decomposed. Figure 3.8 depicts the workflow in which several transformations are applied on the raw car data suffer until detecting a maneuver.

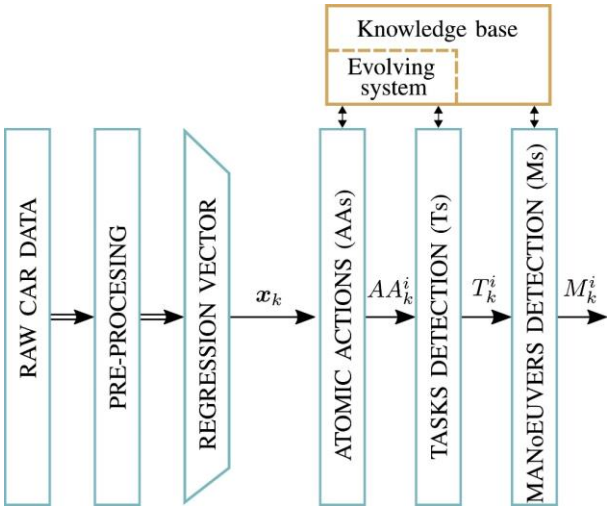


Figure 3.8. Data acquisition system and maneuvers detection [140].

In brief, pre-processing relied on discretizing all model inputs regarding how the trend of signals evolve: for speed (v), revolutions per minute (r), steering wheel rotation (w), gas pedal (g) and brake pedal (b), the values {-1,0,1} indicate negative/neutral/positive changes; in case of clutch pedal (c), the values {0,1} indicate deactivate/activate states; and regarding gear switch(s), the values {-1, 1} represent a downshift/upshift gear, respectively. After, discretized signals were row-wise formatted to conform regression vector. Both atomic actions and tasks were recognized through a Fuzzy rule-based (FRB) system adapted with the *AnYa* method [143]. According to this hierarchical taxonomy, the space defined by the regression vector was used to distinguish the point clouds representing all atomic actions. Subsequently, these atomic actions were used to conform point clouds, thus defining the tasks. Finally, a knowledge base was employed to determine the sequence of tasks that correspond to each target maneuver.

The study conducted by Škrjanc et al. addressed overtaking and stopping maneuvers detection. As their part, Andonovski et al. focused on recognizing U-turn and three-point turn maneuvers. Experimentation consists of two phases: the learning phase and the validation phase.

The entire process is conducted at the learning phase, fitting the FRB models corresponding to the atomic actions and tasks and establishing the knowledge base of

maneuvers. At the validation phase, collected data were employed to determine the reliability of the proposed approach.

3.4. Results

The stored information related to a maneuver involves the driving control signals and the vehicle dynamics parameters as well and, it was stored at 30 Hz. The attributes that compose the dataset are listed as follows:

- **Steering wheel angle** [-1, 1]. Vehicle control value that indicates how much the steering wheel is turned (value 0 is that the steering wheel is centered.
- **Gas pedal.** [0, 1]. The vehicle control value indicates how deep the brake pedal is pressed.
- **Brake pedal** [0, 1]. It indicates how deep the brake pedal is pressed.
- **Clutch pedal** [0, 1]. It indicates how deep the clutch pedal is pressed.
- **Left turn signal** {0, 1}. It indicates whether the right turn signal is on (1) or off (0).
- **Right turn signal** {0,1}. It indicates whether the left turn signal is on (1) or off (0).
- **Gear** [-1, 6] Current engine gear.
- **Handwheel torque (N/m).** Vehicle dynamics measure the applied force on the steering wheel for holding the vehicle's turning angle.
- **Speed (Km/h).** Current vehicle speed.
- **Engine speed (RPM).**

Figure 3.9 shows how signals present a repeated pattern when the same maneuver is repeated along the time. In this case, low speed, first and reverse gears, and abrupt changes in the steering wheel angle are the most remarkable features of the three-step turning maneuver.

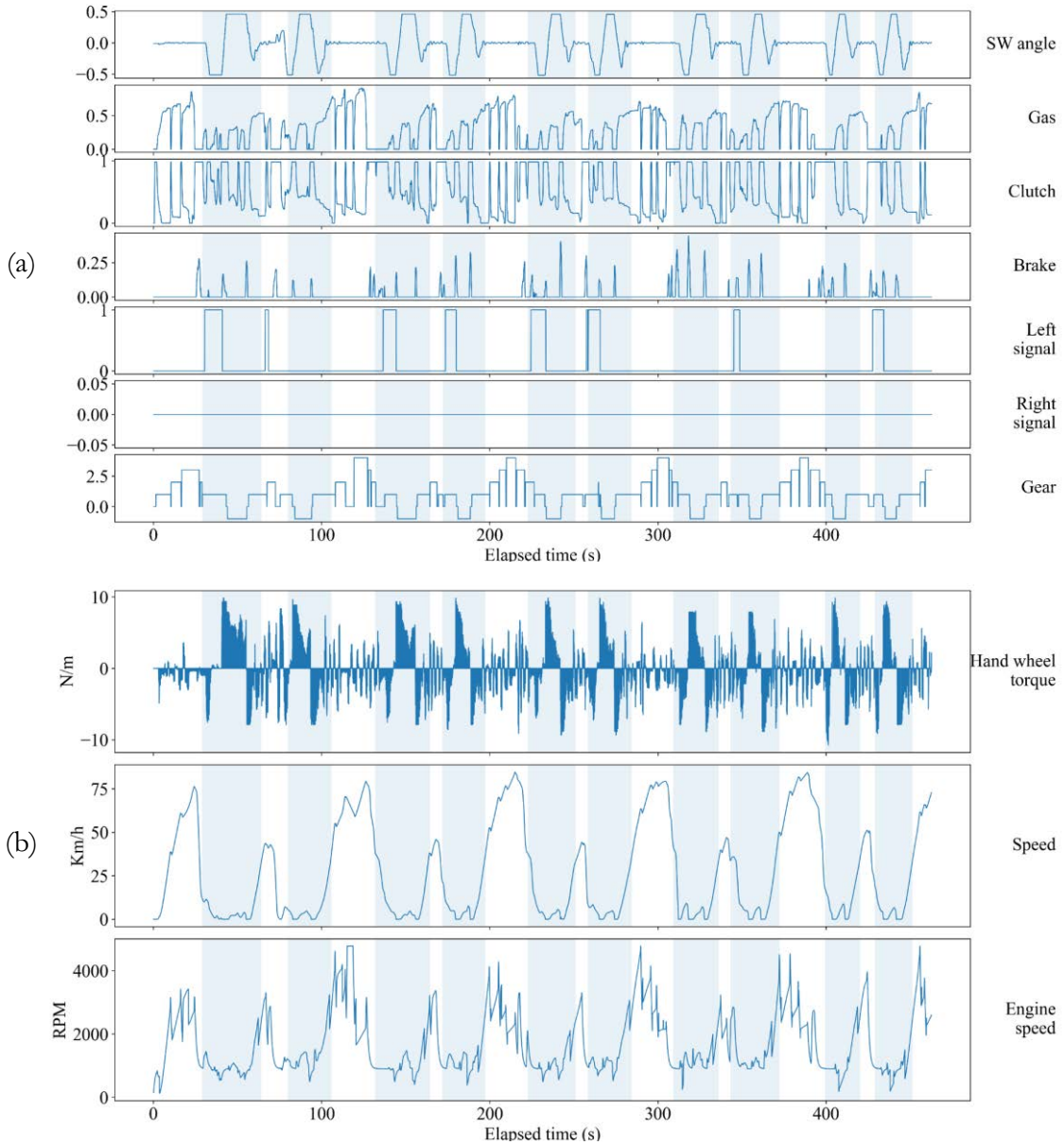


Figure 3.9. Collected data from one driver running the three-point turnings scenario (blue shaded areas: maneuver performing). (a) Driving controls (b) Vehicle dynamics.

In the same way, Figure 3.10 illustrates the driving patterns adopted when overtaking is performed. In this case, it can see how the activation of the left turn signal precedes a speed increase joined with a slight rotation of the steering wheel to the left, and the right turn signal concludes the maneuver with the same opposing amount of steering wheel rotation.

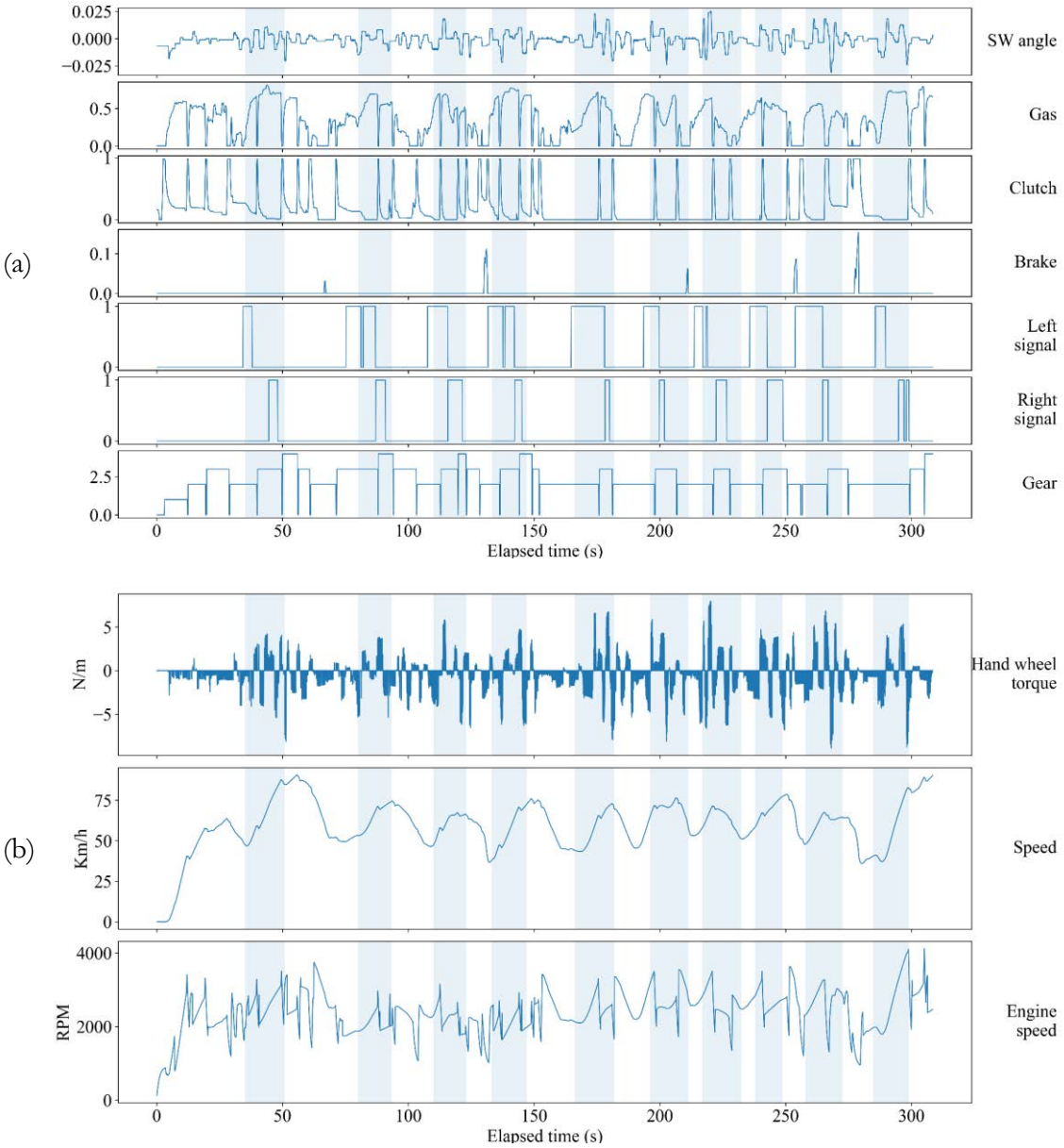


Figure 3.10. Collected data from one driver running the overtaking scenario (blue shaded areas: maneuver performing). (a) Driving controls (b) Vehicle dynamics.

In total, 342 labeled maneuvers compose the dataset. Moreover, four drivers repeated the experiment, but the data were not labeled during data collection on this occasion. Thus, this dataset contains a subset of non-ground-truth maneuvers that can be used for testing.

Table 3.2 summarizes the distribution of instances according to the kind of maneuver.

TABLE 3.2. SUMMARY OF MANEUVER DATASET.

Maneuver	Labeled instances	Non-labeled instances
Stopping	90	60
Overtake	60	40
Left/Right turn	72	48
U-Turn	60	40
Three-point turn	60	40

Maneuver	Labeled instances	Non-labeled instances
Total	**342**	**228**

Regarding experimentation results, Škrjanc et al. detected 51 atomic actions and ten tasks. The validation was conducted by graphical representation of how *Overtake* and *Stopping* maneuvers were detected. Figure 3.11 depicts the validation process for detecting tasks and their corresponding maneuvers. Notice that a maneuver is not detected unless the sequence of tasks has been recognized.

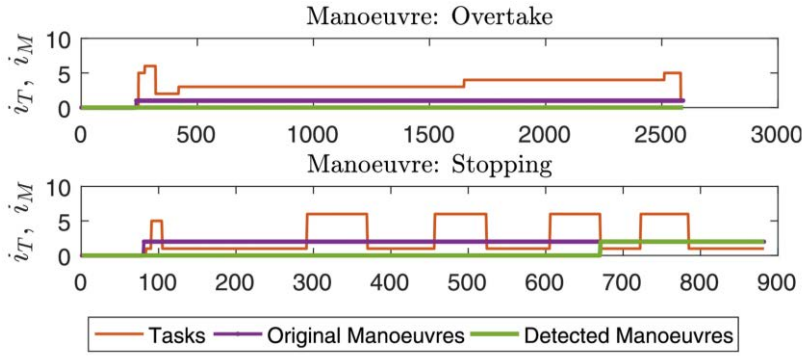


Figure 3.11. Validation of maneuver detection conducted by Škrjanc et al. [140]. *Tasks detection (red lines) and maneuver detection (green lines), maneuver ground-truth (purple lines).*

Andonowski et al. characterized 153 atomic actions and ten tasks for the *U-turn* and *Three-point* turn maneuvers. As a validation method, score metrics including accuracy, specificity, and sensibility were computed to quantify the quality of the maneuver detection model. Table 3.3 shows the metric results of detection for the *U-Turn* and *Three-point turn* maneuvers. It is important to mention that the ground truth considers a maneuver all the time it is active. Accordingly, the *U-turn* maneuver obtained a high rate of false negatives (sensibility) due to its long temporal extension. Oppositely, the *Three-point turn* maneuver had a high false positives (specificity) rate and likely can be attributed to the high number of tasks that the three-point turn maneuver involves.

TABLE 3.3. SCORING METRICS OF U-TURN AND THREE-POINT TURN MANEUVERS VALIDATION CONDUCTED BY ANDONOWSKI ET AL. [141].

Maneuver	Sensibility	Specificity	Accuracy
U-Turn	8.70%	95.3%	40.9%
Three-point turn	99.7%	25.8%	60.3%

3.5. Summary

This chapter presented the entire process to conform a maneuvers database, detailing the design aspects regarding the scenario specification and how maneuvers were collected and labeled. This detailed process directly contributes to the work published in [140] [117].

The proposed protocol for collecting maneuvers has revealed some findings. The collected dataset reflects reliable results regarding instances homogeneity, even when drivers have grown in different socioeconomic environments and driving cultures. However,

repeating the same driving task in the driving simulator seems that provoke fatigue in the drivers. Driving environments lacking in road changes, urban elements, and other dynamic elements are likely to negatively affect the drivers, increasing their fatigue after running several scenarios.

Conducted experimentation based on evolving systems for identifying maneuvers and determining which subtasks compose them had promising results. However, turn lights signals were not included and, this could have contributed to achieving better accuracy results, especially regarding the false positive rate.

The dataset will be extended for future works, including other situations that enrich the catalog of maneuvers, both specialization and diversity aspects. Moreover, future collecting protocols could include higher semantic levels concerning maneuver, including concepts such as aggressive patterns characterized by fast and unexpected responses against events. Concerning signal collected, driver monitoring attributes such as driver's head rotation and gaze direction could be included as long as the used models are accurate enough to not aggregate noise to the sample. Driving monitoring could provide more predictive models that suggest earlier which maneuver is pretended to perform. In addition, data from exteroceptive sensors can be included to improve the detection of maneuvers wherein other vehicles are implied in the maneuver performing.

4. Methodology for laboratory stress induction and collecting of physiological signals

In the scope of affective computing, physiological signals have driven multiple trials to determine the distinct emotional, and altered consciousness states that persons can go through. In the last decade, many studies have dedicated their efforts to collect reliable data about the responses triggered by sympathetic nervous system (SNS) activity. In this way, stress induction paradigms rely on exposure to physical, emotional, or cognitive stimuli to provoke physiological arousal due to sudden mood changes. Despite the growing trend test based on real-life events, widely accepted laboratory tests have proven significant differences between relaxation, basal, and stressful periods in measurable bioindicators. Stroop color-word test (SCWT) and Trier Social Stress Test (TSST)[144],[145], among others, have demonstrated by statistical analysis of heart rate that cognitive interference and social exposure are stressors of moderate magnitude. Accordingly, both time and frequency ultra-short-term metrics (within the 30s and 5 mins) have denoted the existence of significant variations in the heart rate utilizing electrocardiogram (ECG) sensors during stress episodes [20].

This chapter highlights the applied methodology for collecting physiological signals by cognitive stress test and driving-inducted stress in the laboratory. Mixed methods were conducted to collect the data; therefore, this chapter exposes the material and research methods for gathering qualitative and quantitative data. Finally, the resulting sample is outlined by analyzing the gathered qualitative data, including psychosocial, sociodemographic, and self-assessment level information.

4.1. Experimental protocol

The experimental protocol relies on the designing of two time-limited acute stressors. Experiments were conducted at the laboratory, and the challenges are based on two approaches: a widely accepted cognitive stressor and a task-related stressor.

Firstly, the cognitive stressor exposure experiment relied on a Stroop Color-Word Test (SCWT). Since John Stroop reported his research work in 1927, multiple studies focused on explaining the Stroop phenomenon by analyzing its effects on different physiological indicators have been conducted. Leung et al. [146] tested the active brain areas during the different stages of the Stroop Color test by analyzing magnetic resonance imaging (MRI). Satish et al. [147] found significant differences on several metrics, both time and frequency domain of the heart rate variability during the Stroop exposure across different gender groups.

Accordingly, a website was modeled for conducting the SCWT interactively, thus leading the distinct stages of the designed experiment.

Secondly, a car-driving stressor was modeled in a driving simulator system. In this case, participants were exposed to several critical driving situations to provoke arousals related to this application domain. Participants performed both challenges, and these processes led to the recollection process of quantitative data, including physiological data such as EDA and PPG, among others, and the timestamps that bound the limits of each experimental phase.

Moreover, although the labeling of physiological signals through qualitative information could be questionable, qualitative data can provide contextual information which describes certain observable phenomena that emerge on quantitative data. Different widely accepted self-assessment surveys provide scores for evaluating the subject's emotional and stress state concerning psychometric instruments to analyze the emotional state. First, the Positive and Negative Affect Schedule (PANAS) proposed by Watson et al. in 1988 [148] consists of 20 positive and negative terms rated in the Likert scale. Posteriorly, different versions of PANAS, such as International PANAS short (I-PANAS-SF), were proposed to reduce the ambiguity aroused from the polysemy of specific terms.

Widely used Perceived stress scale (PSS) [105] allows qualifying the overall stress that subjects were passing through because of their issues.

After a brief overview of the methodology, the following sections broadly detail the several aspects of the conducted experiments for gathering all the data described above.

4.1.1. Materials

Sensors

The physiological measures were gathered using unobtrusive devices. Empatica E4 wristband [149] was the principal device because it includes multiple sensors for collecting different signals during the experiments. Alternatively, the Contec CMS50D+ pulse oximeter was used as a control device because it is a gold-standard device used in the clinical environment and provides a more accurate BVP measurement. However, this sensor was only used in the Stroop task due to this task does not involve big and continuous hands movements, and the sensor location on the index fingertip does not affect the performance of the experiment. Table 4.1 resumes the features of the acquired signals from each device, specifying the device connectivity requirements, the sort of sensors that the devices incorporate, and the sampling rate of the signals.

TABLE 4.1. PHYSIOLOGICAL DEVICES AND SIGNALS FEATURES.

Device	Connectivity	Sensor	Sampling frequency	Signals
Empatica E4 Wristband	Bluetooth	Photoplethysmography	64 Hz	PPG waveform
			1 Hz [1]	Interbeat interval
			1 Hz [1]	Heart rate
		3-axis accelerometer	32 Hz	Motion across x, y, and z axes
		Electrodermal activity	4 Hz	Skin conductance
		Infrared thermopile	4 Hz	Skin temperature

Device	Connectivity	Sensor	Sampling frequency	Signals
Contec CMS50D+	COM Port	Photoplethysmography	60 Hz	PPG waveform
				Heart rate
				Oxygen saturation (SpO2)

[1] As long as motion artifacts do not corrupt the PPG waveform

Web application

A web structure based on Java Server Pages (JSP) was locally deployed on a laptop computer (Intel i7) using Apache Tomcat 9 Server. This webserver hosted the designed application for conducting the Stroop experiment.

Driving simulator system

The driving simulator STISIM Drive based on M300WS configuration with VDANL Drive capability was used as a driving platform. This system was deployed on a desktop computer (Intel i7). Additionally, the scenario modeling was designed using the scenario description language (SDL) provided by the STIMSIM drive distribution.

4.1.2. Computerized Stroop color-word test (C-SCWT)

The design of the experiment is based on the Group Stroop Color-Word Test proposed by Vanitha et al. [150]. In this work, the authors defined different stress levels regarding the congruency of the displayed words and their printing colors. Consequently, color terms displayed on the same color are congruent words, whereas those displayed in distinct colors are incongruent words. In this study, the stressor exposure relies on presenting sets of incongruent words in which the participant must say aloud the displayed color of each term. Figure 4.1 shows two examples of congruent and incongruent words presented to the subject during the experiment. As can be observed, 30 each kind of color terms addressed to Spanish speakers are presented.

(a) (b)

Figure 4.1. Examples of Stroop color words sets. (a) Congruent Stroop color words (b) Incongruent Stroop color words.

An ad-hoc web application led the participants across around 30 minutes of the Stroop test. This test experiment consisted of several phases as is explained as follows:

- ***Recovery phase***. It consists of stabilizing emotions before the test and between the test phases. During this phase, the subject did not perform any activity. Moreover,

relaxing music was played during this phase to facilitate the subject's mental relaxation.

- ***Baseline phase***. In this phase, the subject must read aloud four sets of 30 correctly painted color terms. Simultaneously, the subject's EDA signal was measured during the reading without perceptive interference.
- ***Stroop phase***. The subject must read aloud four sets of 30 wrongly painted color terms. The perceptive interference was provoked. This reading corresponds with the stress phase.
- ***Combined phase***. The subject had to read aloud eight sets of 30 words, a baseline set of words was presented, alternating with a Stroop set of words each time.

Figure 4.2 shows how the phases explained above are disposed of along the experiment. As it can be noted, the test transits on increasing difficulty with the intention of provoking different stress stimuli. Additionally, recovery phases interpose reading phases, thereby isolating the stressor exposure.

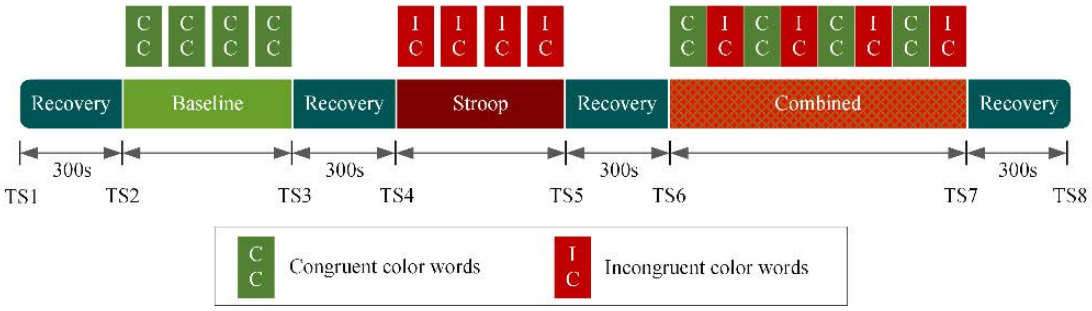


Figure 4.2. Experimental designing of Computerized-SCWT. The distinct experimental phases are delimited with their corresponding color, where TSx are the moments in which the timestamps were collected. IC and CC correspond with the set of congruent and incongruent words presented to the participants.

Experiment deployment

Figure 4.3 overviews the interactive process deployment that leads the C-SCWT experiment. As it can be observed, the participant interacts with a web application deployed on a laptop computer.

This application guides the participant through the distinct phases of this experiment. Moreover, the web application manages the creation of timestamps corresponding to the starting of each experimental phase to relate the acquired signals with the exposure of the event. Furthermore, the participant wears both devices to capture their physiological measurements produced during the experiment performing. All information derived from each experiment was stored in a folder repository. Thereby, a file-based database centralizes all the information about the conducting of the C-SCWT experiment.

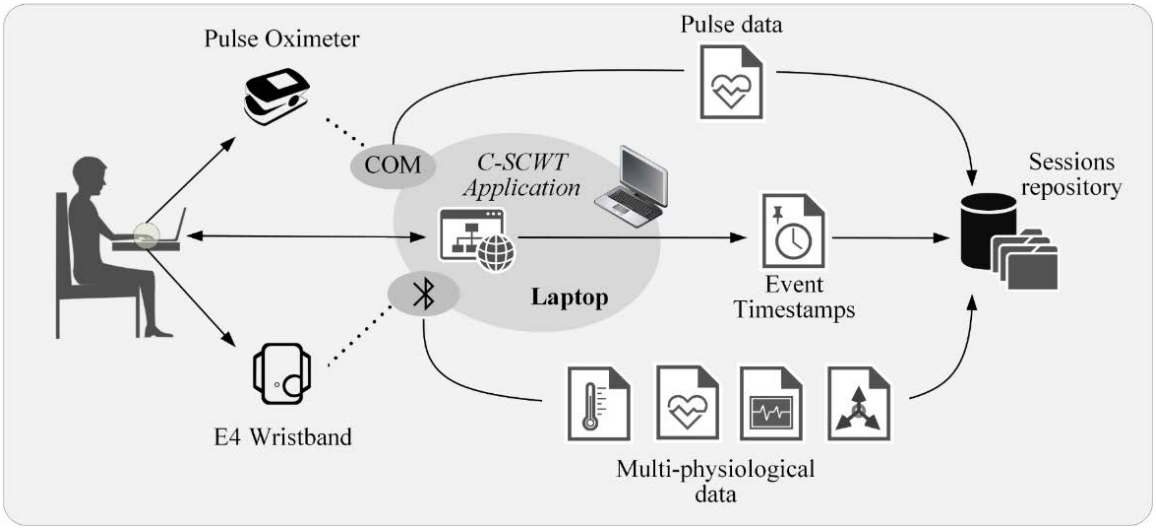


Figure 4.3. Deployment scheme of interactive SCWT experiment.

4.1.3. Driving event stressor exposure test (DESET)

The driving event stressor exposure test (DESET) is based on stress elicitation by presenting situations that likely raise fight-or-flight responses owing to their connotation in the driving context. These situations are triggered suddenly so that the driver only has a brief and limited period to cope with them before suffering a driving crash. DESET relies on different driving stages described as follows:

- ***Recovery phase***. The participant drives without environmental stimuli. Only the road is displayed at this stage, empty of traffic signs, other vehicles, pedestrians, and even urban elements such as buildings and green areas.
- ***Normal driving.*** The participant performs a driving task in an urban driving environment without any driving conflicts.
- ***Critical events.*** At this phase, the driver had to face five critical driving situations. Some of these critical events involve emotional components to induce emotional arousal in the driver.
- ***Divided attention.*** On this task, the driver must keep the distance respect with a leading vehicle which modifies its speed constantly. At the same time, the driver must keep attention to some symbols that appear at the left and right of the bottom of the screen to press the corresponding button located on the dashboard.

Table 4.2 describes the five critical events presented to the drivers in this stage. Each critical event is detailed according to the situation scenery, the dynamic actor involved, and the major handicap to deal with it.

TABLE 4.2. DRIVING CRITICAL EVENTS DESCRIPTION.

	Short name	Scenery	Handicap	Description
a	Absent-minded pedestrian	School area and the speed limit at 30 km/h.	Brief-time to reaction.	An absentminded teenage girl goes across the roadway and gets paralyzed in the middle.

	Short name	Scenery	Handicap	Description
b	Ball missing	Residential condos, One-lane roadway, narrow street, and the speed limit at 30 km/h.	Low visibility on sides.	Two kids are playing until they throw the ball in the middle of the roadway. After that, one kid runs to take the ball.
c	Unproperly merge	Commercial area and the speed limit at 50 km/h.	Brief-time to reaction.	A parked vehicle tries to merge into the traffic without checking the arrival of vehicles.
d	Dog crossing	Residential Condos, one-lane roadway, narrow street and, the speed limit at 30 km/h.	Low visibility on sides, brief time for reacting.	A dog that stays on the left sidewalk goes across the roadway.
e	Emergency vehicle	Interurban Road, Two-lane roadway, the speed limit at 70 km/h.	Sudden noise.	A police vehicle starts an emergency service, speeds up, and turns on the emergency lights and sirens.

Figure 4.4 illustrates how the distinct previously exposed driving phases are established, including their approximate duration periods when the timestamps were collected for upcoming data analysis. As it can observe, the DESET (Experimental designing of the Driving event stressor exposure test) experiment consists of six delimited phases such as detailed above. Recovery phases enclose the experiment to mitigate the influence of non-related stimuli on the physiological signals. Normal driving phases interlayer defined stressors to avoid prolonged exposure to cognitive stimuli, thus returning to the basal level.

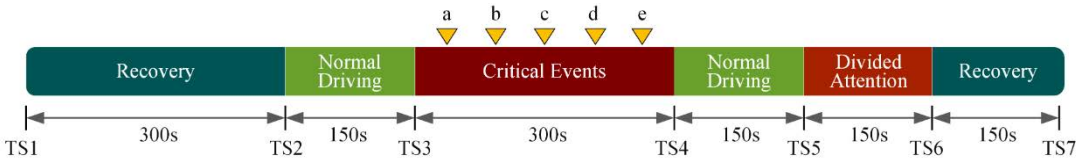


Figure 4.4. Experimental designing of the Driving event stressor exposure test (DESET). The distinct experimental phases are delimited with their corresponding color, where TSx are the moments in which the timestamps were collected (yellow triangular marks: each hazard presented during the driving).

Experiment deployment

Figure 4.5 illustrates the deployed resources to conduct the DESET experiment. The psychological measures were analogously collected with the methodology described for SCWT, except for the exclusion of heart data gathering with the pulse oximeter owing to the motion requirements.

The designed experiment phases were adequately implemented using the driving scenario definition language provided by STISIM Drive. Accordingly, the stage bounds were established from the covered distance, and the timestamps were extracted after achieving each stage limit.

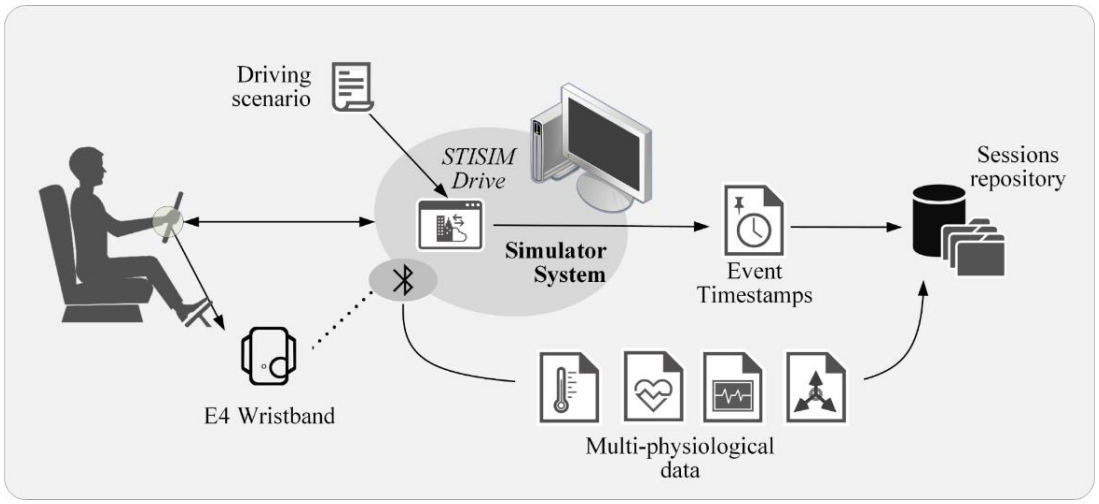


Figure 4.5. Deployment scheme of ERDT experiment.

4.1.4. Self-assessment surveys

To conclude the experiments, some psychometric instruments were used to obtain subjective perceptions of participants' mood states, the stress caused by events and experiences in their lives, and those experimented with during the experiment session.

First, I-PANAS-SF [151] questionnaire (see Annex I) was designed to assess the subject's mood before conducting the designed Stroop-based experiment. Accordingly, subjects rate their feelings on a five-level Likert scale regarding five positive terms (active, determined, attentive, inspired, alert) and five negative terms (afraid, nervous, upset, hostile, ashamed). As a result, the summation of positive terms ratings and negative terms ratings provide the positive (PA) and negative (NA) scores, respectively. PA and NA scales are independents, and each one ranges from 5 and 25, where the highest scores represent a higher level of positive and negative affect, respectively. Accordingly, three affective categories were proposed from the score range terciles as follows:

- Low PA (or NA) scores are within the interval [5, 11].
- Mid PA (or NA) scores are within the interval [12, 18].
- High PA (or NA) scores are within the interval [19, 25].

Second, the PSS-10 questionnaire (see Annex I) relies on ten questions that the subject must answer using a 5-point Likert scale (from 0 – never to 4 – very often). The PSS-10 score is obtained, reversing the score of the positive items and, subsequently, calculating the summation of all the ratings. The PSS score ranges from 0 to 40, where the highest value represents high-stress levels. Additionally, three stress levels are defined from subranges as follows such as was reported in [152]:

- PSS within [0, 13] is considered a low-stress level.
- PSS within [14, 26] is considered a moderate stress level.
- PSS within [27, 40] is considered a high-stress level.

Concerning ERDT, a questionnaire (see Annex I) about the self-perceived stress regarding the different events was presented after the driving test experiment.

4.2. Experimental setup

Participants were dated on separate days to conduct both experiments. Participants were dated in a meeting room without an audience on the first day, where the experiment materials were deployed. Before starting, participants completed a profile form, including information about their age, gender, education rate, employment, and driving skills, including driving license years, and driving distance per year. After that, the I-PANAS-SF form was filled out. Finally, the C-SCWT experiment was conducted.

On the second day, participants were dated in the driving simulator room. Before starting, the E4 wristband was placed on the participant's wrist, checking the correct location and tightness. Moreover, participants filled out the PSS questionnaire during this period. After that, the participant performed the DESET experiment. Finally, participants completed a questionnaire, reporting their subjective impressions about the perceived stress regarding the presented events.

During both experiments, only the laboratory technician was present to control the well-outgoing of the experiment. Subjects were adequately informed about the recording and storing of their physiological data elicited during the experiment and, they signed a data use agreement (see Annex I).

4.3. Results

Twenty-five healthy and non-colorblind subjects (6 females and 19 males, age: 24 ± 2.99 years (mean ± SD)), with at least one year of driving experience, agreed to participate in the experiment sessions. The experiment was conducted between April and May of 2019. In this section, the sociodemographic and psychosocial characteristics of the sample are detailed, and the perceived stress related to driving stressors.

The sample was taken in college; therefore, the population was composed of master's students and Ph.D. candidates. Concerning occupation, 24% of participants indicated be students, 20% partial-time employers, and 56% remain declared be full-time employers.

Figure 4.6 depicts the psychometric outcomes of a statistical sample, describing how emotions are related to daily life stress. Concerning PSS Score, three PPS ranges are represented in the sample where three probability distributions are differentiable, and most of the participants were coping with moderated stress levels.

In terms of scores, NA and PSS share a weak linear correlation (Pearson's r = .303), whereas both PA and PSS scores and NA and PA scores share a very weak linear correlation (Pearson's r =-.110; Pearson's r =.034). However, regarding scale ranges, NA and PSS ranges are moderately correlated (Pearson's r = .485), whereas PA and PSS ranges and PA and NA ranges maintain a very weak correlation (Pearson's r = -.057 and Pearson's r = -.011, respectively).

A conditional distribution table was constructed to establish a relation between occupation and PPS Range.

Accordingly, Table 4.3 shows the dependency between the PSS range and the occupation.

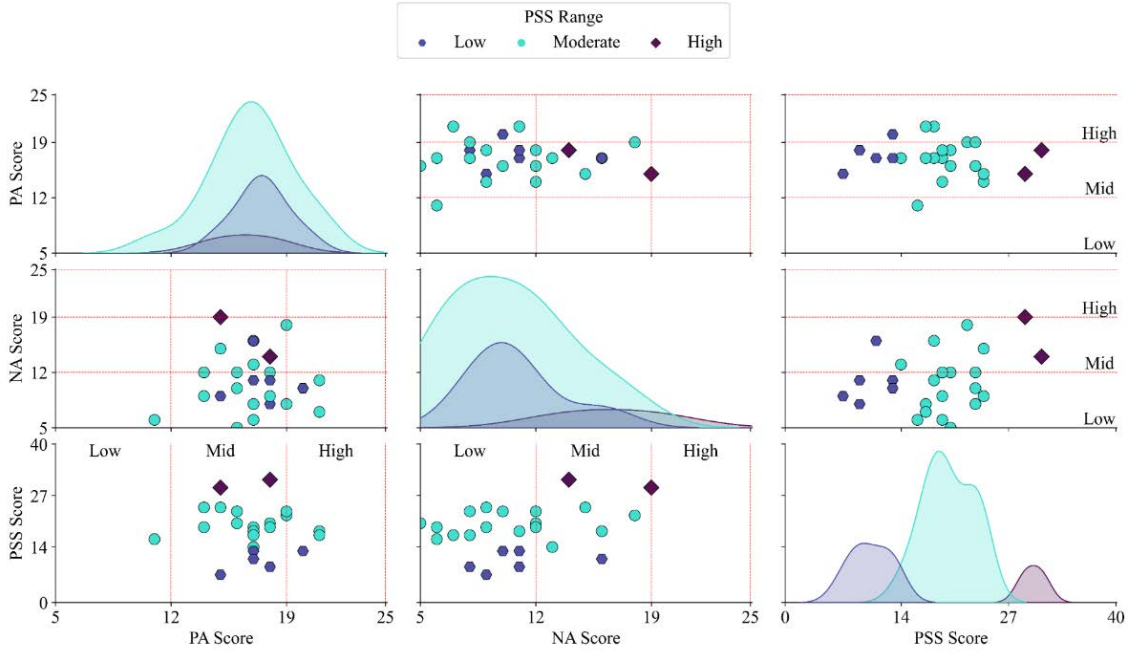


Figure 4.6. Relation between the range of Perceived Stress Scale (PPS), the score of Positive Affect (PA) and the score of Negative Affect (NA). Diagonal plots: Kernel density estimate (KDE); upper and lower corners plots: Scatter; Dotted lines: PANAS ranges bounds.

TABLE 4.3. CONDITIONAL PROBABILITY OF PPS RANGE GIVEN THEIR OCCUPATION.

	Low	Moderate	High
Student	.333	.5	.167
Partial time	0.	1.	0.
Full time	.286	.643	.071

Regarding driving ability, Table 4.4 summarizes the driving skills based on the covered distance yearly. In this case, age and the driving license years are the discriminant factors. As can be observed, drivers with the highest driving skills are also the oldest and with more driving license years. Furthermore, the female/male ratio indicates the absence of women in the two highest profiles. Figure 4.7 graphically overviews the driving skills where both discriminant variables are presented regarding the participants' genre.

TABLE 4.4. DRIVING SKILLS PROFILES BASED ON THE COVERED DISTANCE PER YEAR AS REGARDS AGE AND DRIVING LICENSE YEARS.

Kms/year	Count	F/M	Age	Driving license years
(, 1000)	9	.8	23.44 ± 3.32	4.00 ± 2.50
[1000, 5000)	6	.5	23.67 ± 2.66	4.83 ± 3.06
[5000, 10000)	3	0.	23.67 ± 1.53	3.67 ± 2.31
[10000,)	7	0.	26.71 ± 2.43	7.71 ± 2.29

About the DESET experiment, Figure 4.8 resumes the overall assessments of the perceived stress. Figure 4.8 (a) indicates that overall stress follows a normal distribution across the sample. Figure 4.8 (b) shows that only around 2 out of 10 individuals experimented stressful moments in driving stages without sudden events. Moreover, participants were asked about their stress perception related to the different experimental stages. "Dog crossing" and "Ball missing" obtained the highest stress ratings. Figure 4.9 shows how the stress rates regarding each presented stressor are distributed. As it can observe, the "Police in Emergency service" event obtained the lowest stress rating, just like participants indicated in Figure 4.8 (c).

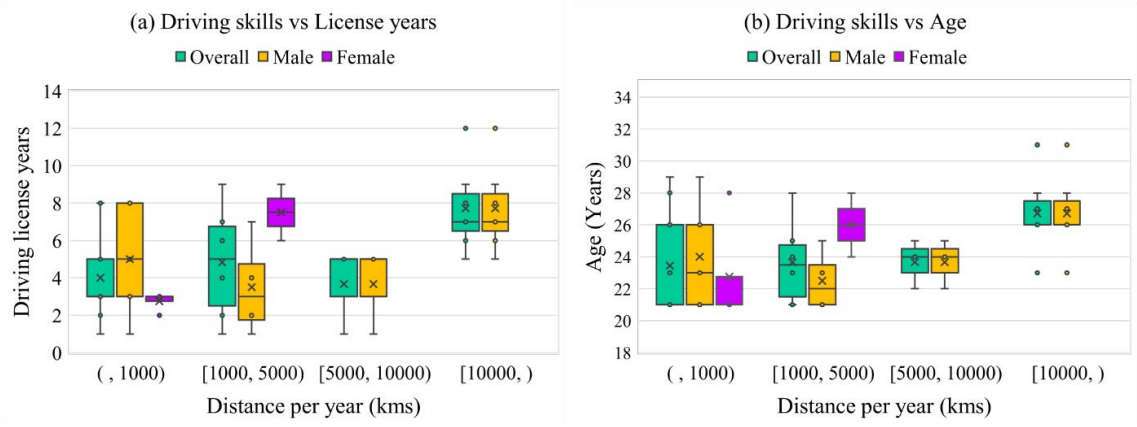


Figure 4.7. Boxplots of driving skills profiles based on covered distance per year by (a) driving license years and (b) driver's age.

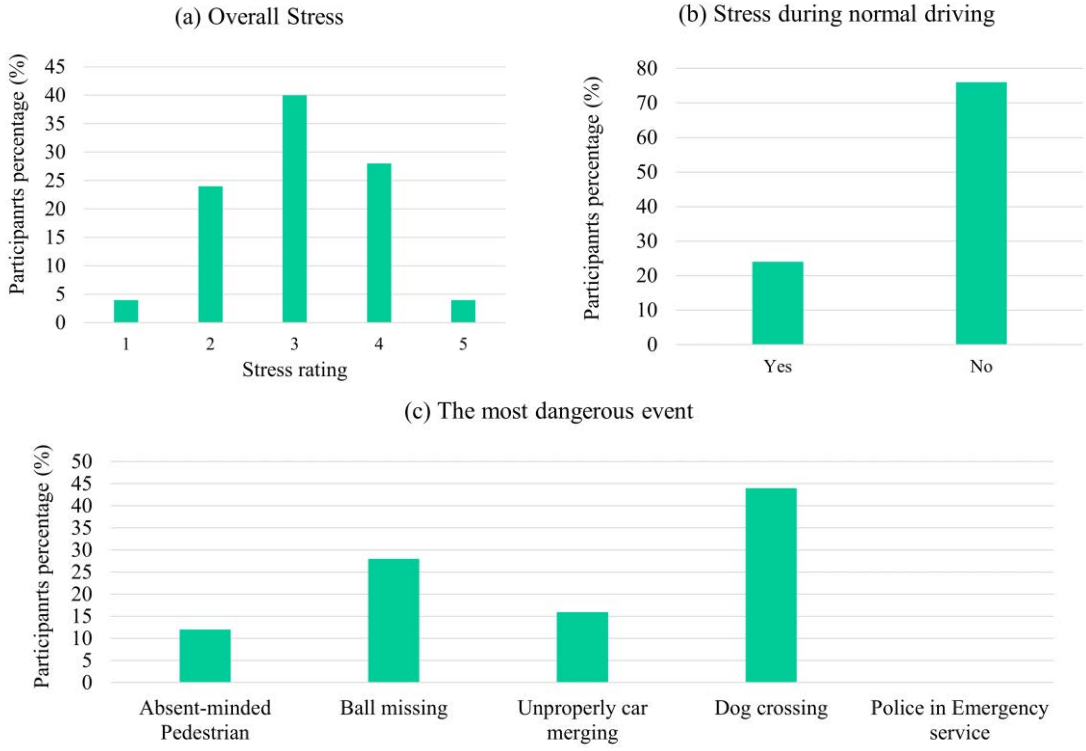


Figure 4.8. Overall stress assessment about the DESET experiment. (a) Overall stress experimented; (b) stress experimented during normal driving; (c) selection of the most dangerous event.

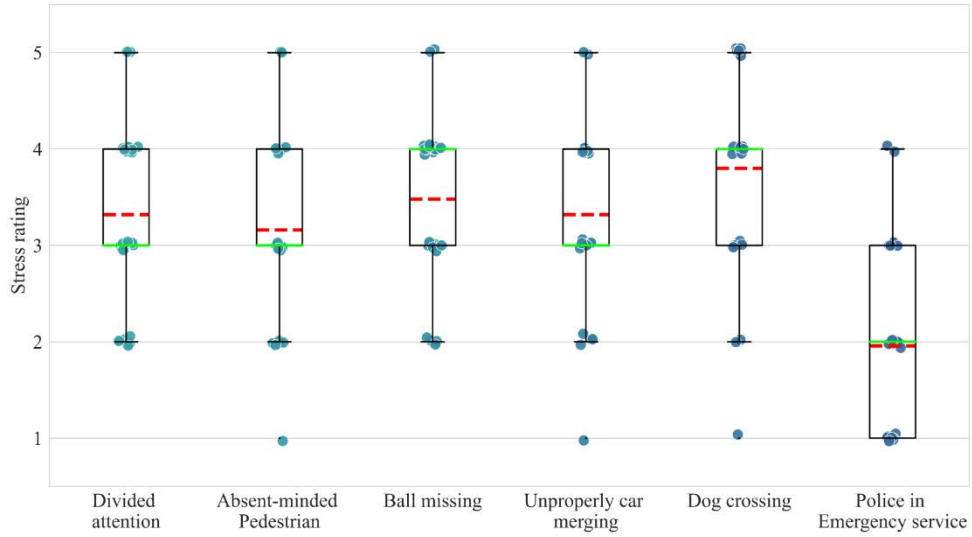


Figure 4.9. Stress rating outcomes of the driving events.

4.4. Limitations

Overload issues complicated the correct performance of the driving simulation. Because of this, only it could retrieve the event traceability of 17 out of 25 samples accurately. Another limitation was that the long extension of the survey dilates the required time to complete it appropriately; thereby, questions about habits that can affect the physiological signals such as smoking, frequency intake of caffeinated beverages, physical activity, eating, and sleep schedules were not included. Moreover, the inclusion of other psychometric instrumentals like the Connor-Davidson Resilience Scale (CD-RISC) [153] could have improved the gathered psychosocial details.

Regarding the gathered physiological signal, a priori, signals collected during the driving test are likely motion corrupted owing to the E4 wristband sensibility, especially PPG signals.

4.5. Summary

This chapter introduced the methodology for collecting data linked to stress patterns. Twenty-five participants were exposed to two acute-time limited stressors.

Regarding psychosocial results, most participants indicated going through a stressful moment in their lives. This fact suggests that the course period when the sample was toke was crucial. Likely, students with final exams and final project deliveries of their unrolled subjects were coping with a stressful moment.

Regarding self-assessment of experimented stress, the unusual event with highly risky connotation provoked high-stress levels in most participants. However, sudden but common events did not provoke stress on participants

5. Exploring unsupervised learning to characterize stress patterns on EDA signal

Nowadays, wrist-worn wearables are moving towards multi-path sensor designs, incorporating various sensors to track activities and measure health and wellness state along the day. Besides advances in heart rate monitoring, novelty designs, such as newly released Fitbit devices, include sensor technology to gather the electrodermal activity (EDA) because it is one of the most widely used physiological measures to determine arousal periods and different emotional states [154]. Electrodermal activity refers to the alteration of the electrical properties of the skin due to the secretion of eccrine sweat glands [90]. The perspiration located in eccrine sweat glands and other physiological effects including pupil dilatation, increasing heart rate and blood pressure, and increasing cortisol concentration in salivary and blood are the consequence of Fight-or-Flight response, which is the biological reaction of dealing with stress episodes [155]. The autonomous nervous system is responsible for increasing (sympathetic nervous system; SNS) or mitigating (parasympathetic nervous system; PNS) these physiological effects produced by stress episodes. The constant interaction of both systems forms part of the self-regulation process to maintain the chemical stability of the organism, known as homeostasis.

Some studies which address stress detection propose a multivariate analysis based on sensor fusion, including information derived from EDA and heart monitors, e.g., electrocardiogram (ECG) and photoplethysmography (PPG) [156]. However, despite technological improvements of wrist-worn wearable's PPG sensors to collect the volumetric changes in blood profusion, the heart rate (HR) measurement still presents several limitations in real-world conditions. Motion artifacts inducted by sudden arms movements, wrist placement, and device tightness are the principal factors contributing to a pretty much lower accuracy of the PPG signal, thus deriving in missing data and high error rates on the HR measurement [157]. Another limitation lies in how persons' demographic features can decrease the accuracy of PPG sensors based on green light reflection. Even green light PPG sensor has shown to be more robust to motion artifacts than red light [158], it has demonstrated that darker skin tones reduce the amount of green light travel through tissue [157].

Due to the facts exposed above, the presented study explores the stress detection based on EDA analysis, introducing user segmentation, time series clustering, and sequence mining as a pipeline of approaches to obtain a model that can represent the most common patterns regarding how EDA behaves during stress episodes and be able of recognizing them in real-time.

The adoption of this approach against supervised learning methods is based on the complexity induced by three primary factors: individuals' variance, the lack of datasets labeled exhaustively, and the resolution of sensors plugged in wrist-worn wearables.

From the perspective of biological personality research, Carver and Scheier [159] defined the concept of personality as "*a dynamic organization, inside the person, of psychological systems that create the person's characteristic patterns of behavior, thoughts, and feelings*." Stemmler and Wacker [160] reviewed the factors which intervene in individuals' variance, being biological variables the most participating in the manifestation of these differences.

In laboratory-based stress induction experiments based on sociocultural abilities such as the Trier Social Stress Test (TSST) [161], physiological data are labeled according to the experimental phase. Nevertheless, two factors can negatively affect this kind of examples labeling. First, First, the variability of personality traits within experiment individuals can compromise the reliability of the labels, especially for phases that theoretically do not provoke stress in the subjects. Second, stress responses are momentary and do not remain stable along the time. Therefore, a subset of examples will assign a wrong label when the window width taken for analyzing the signal is too narrow. The exposed problems increase the variance of the observations set, thus affecting the global accuracy of classification algorithms based on supervised learning.

In terms of ultra-short-term analysis, EDA signal modeling is still an open question. Real-world applications present a sensor resolution handicap. Wrist-worn wearable devices present a much lower sampling rate (from 1 to 10 Hz) than ambulatory sensors that collected data at sample range within 200 and 400 Hz [162]. Therefore, assessing SNS activity by features extraction [111] and frequency spectrum analysis such as proposed by Posada-Quintero et al. [163] could not be suitable approaches for real-time analysis in the scope of wearable devices due to the number of samples that these techniques require.

In summary, the presented study addresses the stress characterization based on a one-dimensional analysis of EDA signal obtained with a wrist-worn wearable device. The proposed framework tackles three problems, including user segmentation, data reduction, and pattern finding from the perspective of unsupervised learning, applying different algorithms according to the nature of the problem.

The experimental process has been conducted using the public access dataset called Wearable Stress and Affect detection dataset (WESAD) [164]. This dataset has been extensively used in stress detection due to the experimental protocol that the authors proposed merging different effective psychological techniques for stress, emotional, and relaxation induction, including TSST, video-based emotion elicitation [165], and mindfulness exercises used as an emotional regulator [166]. Although chapter 4 was entirely dedicated to stress elicitation and a dataset was collected as a result of the conducted study, using the WESAD dataset allows validating the proposed method and comparing the outcomes with other studies.

This chapter mainly contributes to the characterization of EDA using representation data models, robust to magnitude, diversity, and applicable in systems that require a fast response, such as active safety systems and autonomous driving functions in the scope of

driving safety and ergonomics. The goal is to identify stress patterns on the driver, thus providing real-time information about the driver's cognitive state as part of a driver monitoring system.

After this introduction, the remainder of this chapter is organized as follows. Section 5.1 describes the proposed framework for stress characterization, delving into the underlying issues of managing physiological signals and how to address each of them from an unsupervised learning perspective. Section 5.2 widely describes the method applied from discovering nested structures in EDA signal and a hierarchical variance level that groups users according to similarities. Section 5.3 introduces time-frequency as a preprocessing method for characterizing fine-grain EDA signals by partition clustering models. Section 5.4 explores the paradigm change towards sequential pattern analysis through the symbolic representation of the EDA signal. Section 5.5 analyzes the performance of supervised algorithms guided by all gathered information by unsupervised-based analysis and, the achieved outcomes are compared as far as possible with reviewed literature. Section 5.6 discusses the findings across the distinct framework stages, exposes the study's limitations, and presents the future guidelines derived from this research.

5.1. Proposed framework: General overview

Addressing stress detection from the perspective of unsupervised learning and hybrid models involves defining a machine learning-based framework whose processes have to be designed to tackle each of the underlying issues of this problem. In this way, the proposed framework is aligned with the "*Divide & Conquer*" method wherein a complex problem is divided into subproblems, and intermediate results lead to the starting of the following subproblem resolution. Figure 5.1 depicts how original EDA data is processed recurrently to fulfill subproblem and global requirements.

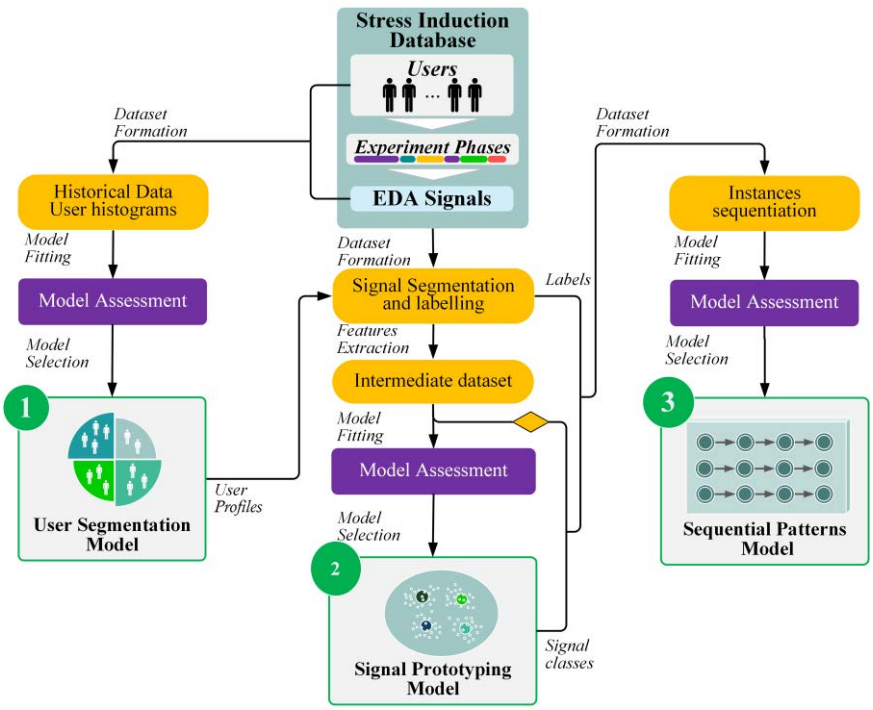


Figure 5.1. Unsupervised-based framework for stress patterns characterization.

As can be observed, physiological data obtained from the stress elicitation experiment fed the distinct parts of the framework. Therefore, the deployment of this framework requires a well-formed dataset in which experimental protocol had been defined meticulously. Details, such as defining the experiment phases limits accurately, locating the irrelevant regions to ignore, and identifying the corrupted excerpts because of sensor failures, are crucial for achieving accurate models, mostly when models are targeted to real-time signal analysis.

Stress characterization from an unsupervised perspective tackles the underlying problems that inhibit achieving high accurate general models with supervised learning approaches, especially in ultra-short-term analysis. High variability across individuals' physiological measures and the lack of instances labeled exhaustively constitute the critical challenges in stress detection.

Here, the presented framework addresses the previously mentioned issues with the statement of the three following subproblems:

The first subproblem to tackle is subdividing the entire population into subsets of individuals with common characteristics. Due to intra-individual differences that inescapably introduce variance to the collected physiological sample, finding similarities across user population that allows identifying user segments could support the achievement of a model approach halfway between personalized and general models.

Nested data structures can be present in the sample during any analytical work. Hierarchical-based models such as multilevel modeling address the analysis of data that presents complex patterns of variability, determining which factors produce a distinct level of variance in the sample [167]. Splitting individuals by similar physical characteristics provide a hierarchical level data according to an aspect of variance similarity. According to this statement, this study raises the user characterization based purely on users' physiological behavior during stress elicitation, leaving aside sociodemographic aspects to segment the set of individuals. Accordingly, the development of a user segmentation model aims to approach stress detection as a hierarchical model. Several models are obtained according to the distinction of subsets of individuals that share a similar EDA behavior during the elicitation of different mental states.

The second subproblem deals with the issue of determining the EDA state from short signal excerpts. Handling real-time sensor streams, the main problem of characterizing signal states lies in the lack of reference values due to inter-individual variations and within-person effects. Temporal variance on individuals provoked by internal and external factors makes uncertain the bounds of EDA signal during the exposition of the same stimuli, even when individuals belong to the same variance cluster. This fact impedes that historical EDA statistics are a trustworthy resource to determine a current state. Vildjiounaite et al. [168] highlighted how physiological data's temporariness affects the reliability of machine learning approaches. They argued that stress models should evolve in the same way that human behavior and physiological condition changes along the time, thus questioning the strictly supervised approaches in the scope of stress recognition.

Accordingly, this study explores the feature analysis to determine which fine-grain EDA waveforms are the most representative in the signal landscape corresponding to each defined user segment. Further, this model should be fitted purely unsupervised, regardless of the labels defined by the distinct experiment phases. Mainly, this approach aims to reduce the data dimensionality of the EDA fragments, thus allowing the symbolic transformation of an EDA signal.

The third and last subproblem contemplates the study of transitions across the signal states, evaluating how the joint occurrence of distinct states is linked with the different physiological states induced during the experiment protocol and thus addressing the labeling inexactness. Even experiment phases are temporarily delimited, the elicited states, especially the stress, probably do not occur along with all experiment phase duration, happening in a punctual moment in which the maximum peak is reached. Personality traits and sociocultural abilities are significant factors because tasks that become very stressful for some people, other ones can perform naturally. The finding of relevant patterns is extensively applied in diverse scopes such as activities and behavior recognition and, it can be a proper method when the labeling process has not been conducted exhaustively.

All models generated in this framework should be obtained following the CRISP-DM methodology [169]. Thus, the achievement of each model includes data understanding, data preparation, modeling, and evaluation phases.

5.1.1. Model assessment methods

From the point of view of unsupervised learning, evaluating models' performance implies considering different evaluation criteria when the ground truth of data is unknown. As Figure 5.1 shows, each fitting process of the distinct kind of models presented in this framework is succeeded by an assessment stage. Here, heuristics support the choice of the most fitted model according to objective information. Different metrics and methods for evaluating unsupervised learning models are exposed as follows:

- ***Bayesian information criterion (BIC).*** Proposed by Schwarz, BIC is a statistical method that determines which is the best model within many different models [170]. BIC computation is defined with the equation as follows:

$$BIC = k\ln(n) - 2\ln(\hat{L}) \tag{1}$$

  Where $(\hat{L})$ is the log-likelihood of the model, *n* is the number of training examples and, *k* is the number of terms of the model. Model selection is based on minimizing the BIC function; hence, models with lower BIC are generally preferred. BIC favorites the selection of simpler models, penalizing models with more parameters.

- ***Within-cluster sum of squares (WCSS)***

  It measures the sum of squared distance between each point and its cluster centroid. Its computation is formulated as (5.1):

$$WSCC = \sum_{k=1}^{K} \sum_{i \in S_k} \sum_{j=1}^{P} \left(x_{ij} - \bar{x}_{kj}\right)^2 \tag{5.1}$$

where S_k is the set of observations in the k[th] cluster, $\bar{x}_{kj}$ is the j^{th} attribute of the centroid for the k^{th} cluster and, P is the number of attributes of the sample. WCSS is an internal validation measure that quantifies the compactness of the cluster, i.e., it determines how closely the objects in a cluster are connected [171]. Therefore, lower WCSS values provide better partitions.

- ***Between-cluster sum of squares (BCSS)***

  It measures the squared average distances between all centroids. BCSS is defined as (5.2):

$$BCSS = \sum_{k=1}^{K-1} |C_k| \sum_{j=1}^{P} \left(\bar{x}_j - \bar{x}_{kj}\right)^2 \tag{5.2}$$

  where $|C_k|$ is the size of the cluster k, $\bar{x}_j$ is the j^{th} attribute of the sample mean. BCSS is an internal validation measure as well; however, BCSS is a separation metric that indicates how distinct or well-separated one cluster is from others [171]. Consequently, Better partitions are obtained with larger BCSS values.

- ***Silhouette coefficient***

  Rousseeuw [172] stated silhouette coefficient to measure the cluster consistency by determining the degree of within-cluster dissimilarities. Let an object of the dataset be denoted by i, the cluster assigned to i denoted by A and, the cluster with the second smallest dissimilarity denoted by B; silhouette coefficient $s(i)$ is defined as (5.3):

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \tag{5.3}$$

  where $a(i)$ is the average distance of i to all objects of A and $b(i)$ is the average distance of i to all objects of B. Value of $s(i)$ is between interval [-1, 1] and is interpreted as follows. When $s(i)$ is close to 1 indicates that i has been correctly clustered, values near to zero means that it is not clear whether i should be assigned to A or B, and a value near to -1 indicates that i has been wrongly classified, being B its correct cluster assignment. Its computational complexity is $O(n^2)$ complexity.

- ***Knee method***

  The well-known Knee (or Elbow for increasing and convex functions) method [173] determines the optimal number of clusters graphically by location of maximum curvature point in the function defined by WCSS of k. The curvature point indicates that the next value of k does not improve the clustering results significantly.

- **Term frequency-Inverse document frequency (TF-IDF)**

  From the perspective of information retrieval and text mining, TF-IDF allows measuring the relevance of a term in a collection of documents. TF-IDF of a term in a document involves computing the frequency of the term in the document (TF) and the number of documents wherein the term appears across documents collection (IDF).

5.2. User segmentation based on physiological information

Stress detection models can be classified regarding their generalization ability, distinguishing individual-specific, semi-personalized, and general models. In contrast with individual models, few studies proposed user segmentation as a data processing strategy to create a semi-personalized mode, avoiding implementing a person-specific approach. Vildjiounaite et al. [174] and posteriorly, Tervonen et al. [175] adopted the same strategy to distinguish groups of similar individuals for creating a semi-personalized model. They fitted the *k*-Means model with individual statistics such as mean and standard deviation extracted from physiological data collected during their respective experiments. Das et al. [176] distinguished two groups of subjects applying *k*-Means clustering and Manhattan distance as similarity measurements over qualitative data collected by self-assessment questionaries about how their emotional state evolved during the experiment. In this case, the authors chose this value of *k* to discriminate high and low values in all data dimensions.

The proposed method for creating a semi-personalized model is focused on grouping the users from EDA signal distributions. The proposed method for creating a semi-personalized model is focused on grouping the users from EDA signal distributions. For the achievement of this purpose, a Gaussian mixture model (GMM) was fit employing high-resolution histograms taken as each user's characteristic data. The fitting process will adjust a joint Gaussian distribution model through a specified number of components and the EDA histograms dataset. Bayesian Information Criterion was employed to estimate the adequate number of components and, consequently, the minimum number of the distinct user profiles can characterize the entire sample reliably. The resulting model will be the foundation of the data hierarchy that will be applied in the successive processes of the framework. The details about how EDA was processed and how GMM models were obtained are widely described below.

5.2.1. Dataset preparation

For each subject, the entire EDA signal collected in the experiment was processed to obtain a density histogram. Figure 5.2 overviews the process to obtain each data instance of the EDA histogram dataset.

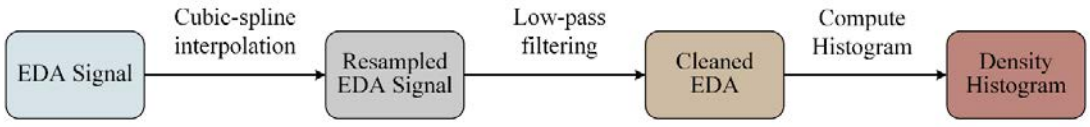


Figure 5.2. EDA processing pipeline for obtaining a density histogram dataset.

First, Cubic-spline interpolation is used to resample the original EDA signal. Due to the low sampling rate of wrist-worn wearables, this step is oriented to increase the number to process the EDA signal properly.

Next, the signal filtering method cleaned the signal, mitigating the noise captured by sensor electrodes. Several EDA processing platforms such as Neurokit and BioSPPY use a fourth-order lowpass Butterworth filter with a cutoff of 3 and 5Hz, respectively.

Finally, the resulting EDA signal from previous processes led to the calculation of density histograms employed to extract the frequency distribution own of each subject. Histogram construction implies choosing the number of bins (also called classes or intervals) within a range of values that define the histogram resolution. In this case, the selected range is a portion of the sensor range specified by the sensor manufacturer because it surely is larger than the reported skin conductance levels typically range from 1 to 20 μS [92]. The bin width value will determine the number of attributes that compose each instance. Although bin width is selected arbitrarily, it must be adequately chosen because it will affect the model's accuracy. For instance, the choice of wide bins will obtain a poorly detailed histogram, and too narrow bins can drive towards performance issues during the model fitting. Consequently, the bin width selection supposes a tradeoff between thoroughness of histograms and resources consumption.

5.2.2. Distribution-based clustering of EDA Histograms

This process aims to determine how users can be grouped by the frequency distribution of their EDA signals. Accordingly, some clustering algorithms allow determining the characteristic data distribution existing in a dataset. As Reynolds described [177], Gaussian Mixture Model (GMM) represents the underlying densities distributions in D-dimensional data as a weighted sum of Gaussian component densities.

Given histogram data from all users whose dimension is determined by the selected bin width and the measurement range, GMM was fitted to estimate the different Gaussian distributions and, thus, determine the user segments. As a parametric model, GMM computation has different variants according to the number of selected components, the weights of each component, and the kind of covariance matrix chosen to estimate the component densities. GMM has four distinct ways of computing its covariance matrix with distinct restrictions levels as follows:

- ***Full***. Each component has its general covariance matrix and, graphically, distributions may adopt any position and shape.
- ***Tied***. All components share the same general covariance; thus, components have the same shape, but the shape may be anything.
- ***Diagonal***. Each component has its diagonal covariance matrix.
- ***Spherical***. Each component has its single variance.

The choice of the GMM's parameters is sensitive to the domain; therefore, obtaining the GMM that maximizes the estimation likelihood from the training data depends on exploring the combinations of GMM parameters.

Once the model has been fitted, experiment users can be grouped into subsets concerning the existing similarity patterns in their EDA histograms. This approach based on histograms facilitates the segmentation of new users because histograms can be built in real-

time from short periods of EDA signal. However, as a disadvantage, the distribution-based model's reliability greatly depends on the number of individuals used for fitting the model.

5.2.3. Experimentation

Evaluation of this method was conducted using the WESAD dataset, in which 15 users conducted a stress elicitation experiment. As EDA data, only data collected with Empatica E4 Wristband were used for the study to validate wrist-worn wearable devices. Table 5.1 overviews the periods corresponding to each experiment phase and a brief description of the activities conducted.

Concerning the construction of histograms, bin width was $0.05\mu S$, and the value range was between 0 and $20\mu S$ to obtain histograms of 400 classes. Furthermore, histograms were normalized according to a probability distribution function; consequently, the integral under the histogram is 1.

TABLE 5.1. DESCRIPTION OF THE DISTINCT PHASES OF WESAD EXPERIMENTAL PROTOCOL.

Phase	Description	Duration (s)
Baseline	Neutral reading (magazines)	1200
Amusement	11 funny videos	392
Meditation 1	Mindfulness exercise of breathing and closed eyes	420
Stress	Speech about their personal traits and countdown from 2023 to zero doing steps of 17.	600
Rest	Rest after stress elicitation	600
Meditation 2	Idem. Meditation 1	420
Total		**3632**

The experimentation consisted of converging all models resulting in combining the intrinsic parameters for the GMM configuration, including the number of components, and computing the covariance matrix of the model. For several components, values within 1 and 7 were used, being this the number of probability distribution functions for computing the maximum likelihood function, and it will correspond with the number of user groups. Considering the number of users, the evaluation of the number of components higher than seven likely produces groups of one individual. Model fitting parameters were selected as follows:

- ***Number of algorithm initializations:*** Number of times that the algorithm is performed to find the best likelihood approximator (4)
- ***Random state:*** Random seed for initialized the algorithm (current system time in nanoseconds mod $2^{32} - 1$)
- ***Maximum number of algorithm iterations:*** 5000

- ***Tolerance:*** variation threshold with determining when the algorithm has converged ($1e^{-3}$)
- ***Training process:*** Cross-Validation with 15 folds.
- ***Covariance matrix types:*** Full, tied, diagonal, and spherical.

Although the nature of data suggests that diagonal was the most indicated according to data, all possible combinations were deployed. As a result, 28 GMMs were converged using EDA histograms to determine the optimal number of groups.

As it is an unlabeled dataset, Bayesian Information Criterion (BIC) was used for determining the preferred model among the combinations of parameters exposed above, taking as the final model the one that obtained the lowest BIC value.

5.2.4. Results

Table 5.2 shows a statistical overview of the recorded EDA signals from 15 users during the WESAD experiment. As the dataset specification describes, the data collected from some users were removed due to sensor issues.

TABLE 5.2. STATISTICAL SUMMARY OF EDA SIGNALS OF WESAD DATASET.

User ID	Samples	Mean (μS)	Min (μS)	Max (μS)	Std (μS)	Skewness
S02	24316	0.392	0.045	1.717	0.329	1.419
S03	25972	1.231	0.292	8.792	1.645	2.424
S04	25692	0.475	0.099	3.119	0.689	2.246
S05	25032	1.599	0.352	4.739	0.934	1.394
S06	28284	3.354	0.199	9.862	2.580	0.824
S07	20952	5.230	3.687	8.174	1.061	0.852
S08	21864	0.401	0.205	1.282	0.217	1.512
S09	20892	0.562	0.055	1.744	0.291	0.878
S10	21984	0.982	0.266	3.508	0.767	1.317
S11	20932	3.185	1.843	6.096	0.855	1.228
S13	22148	7.094	2.841	15.921	3.443	1.053
S14	22192	0.314	0.060	1.034	0.025	2.497
S15	21008	0.635	0.151	1.643	0.364	0.998
S16	22524	0.716	0.243	2.656	0.571	1.663
S17	23680	1.080	0.643	2.586	0.248	2.485

Table 5.3 shows the results of the experimental process to obtain the fittest GMM concerning BIC. As model selection outcomes show, the GMM diagonal and spherical covariance types obtained better results than other options. Diagonal covariance exhibits the best set of BIC values across the different number of components, founding the lowest value of BIC at four components. Figure 5.3 depicts the progression of BIC values of diagonal covariance matrix GMMs.

TABLE 5.3. RESULTS OF MODEL SELECTION OF GMMS BASED ON BAYESIAN INFORMATION CRITERION (BIC) ACROSS THE DIFFERENT NUMBER OF COMPONENTS AND COVARIANCE MATRIX TYPES.

Components	Spherical		Tied		Diagonal		Full	
	k	***BIC***	***k***	***BIC***	***k***	***BIC***	***k***	***BIC***
1	400	3478.691	80199	148548.122	798	-32356.719	80199	148548.122
2	801	-2304.750	80599	149399.879	1597	-39900.418	160399	364195.586
3	1202	-3400.183	80999	150275.044	2396	-39390.622	240599	581084.969
4	1603	-4276.560	81399	151136.051	3195	**-41816.452***	320799	798057.823
5	2004	-8999.304	81799	152001.401	3994	-39777.462	400999	1014962.293
6	2405	-13343.625	82199	152864.392	4793	-40760.294	481199	1231999.689
7	2806	-16956.236	82599	153727.444	5592	-41595.164	561399	1449072.096

* The lowest BIC value among all GMM models

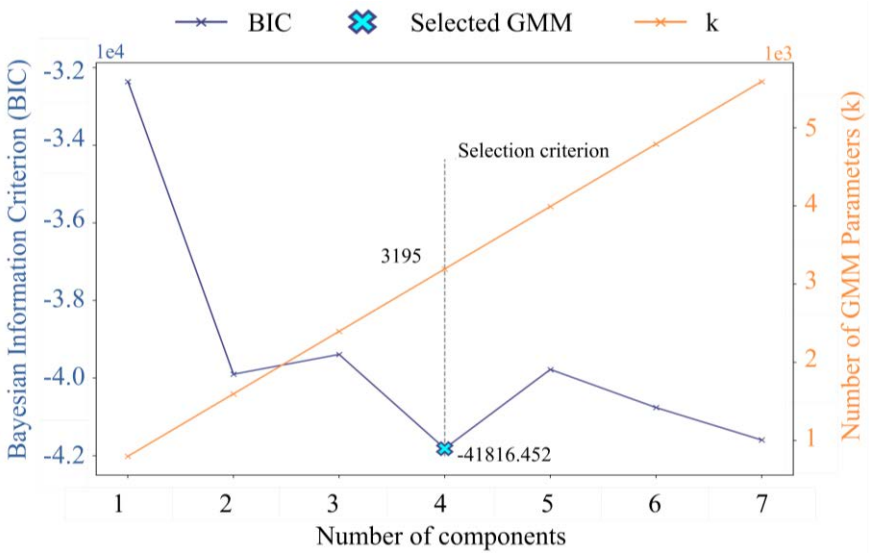


Figure 5.3. Model selection among BIC values of Diagonal covariance matrix GMMs.

Figure 5.4 depicts how individuals were grouped using the GMM model of 4 components and diagonal covariance calculation. As it can be noticed, the graphical representation of histograms shows both similar wideness and distribution shapes over the EDA classes across the same profile's users. Profile A users show variated averages but a pretty similar wide deviation. The users of Profiles B presented similar skewed EDA histograms. The particularities of user S04 made that not were grouped with other users.

Table 5.2 verifies that user S4 has a standard deviation too low, and the subject does not share similar characteristics with any of the other groups. Profile D distributions show some similarities with Profile B in terms of mean and standard deviation; however, GMM had determined that differences are not significant enough to merge them in only one group.

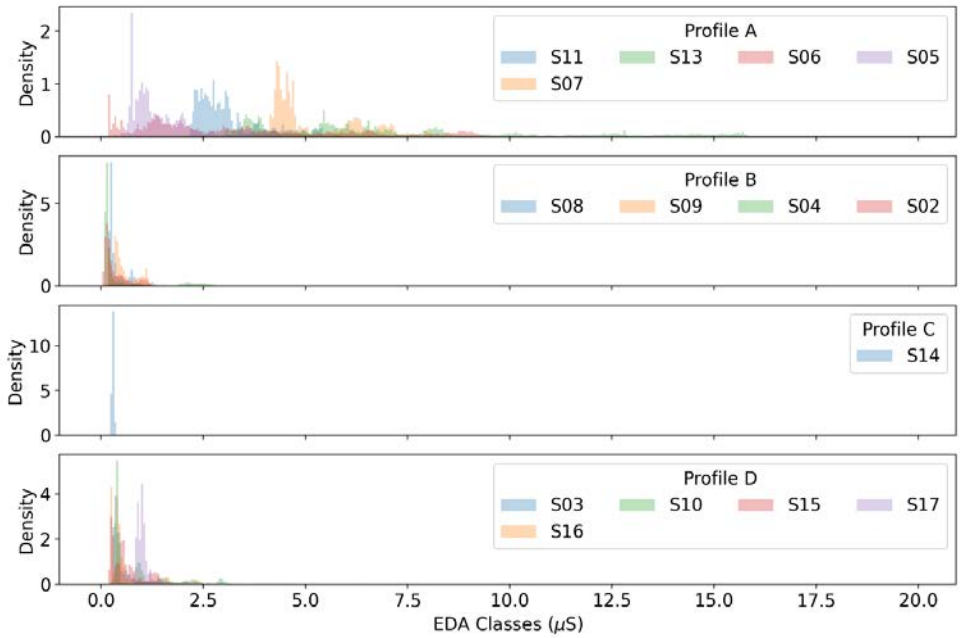


Figure 5.4. Result of user assignation concerning the result GMM for user segmentation.

Figure 5.5 illustrates how EDA signals are distributed across the distinct profiles defined by GMM. By the distinction of the experiment, phases reveal that the grouped users had a similar EDA behavior, except for the differences of magnitude such as histograms shown in Figure 5.4.

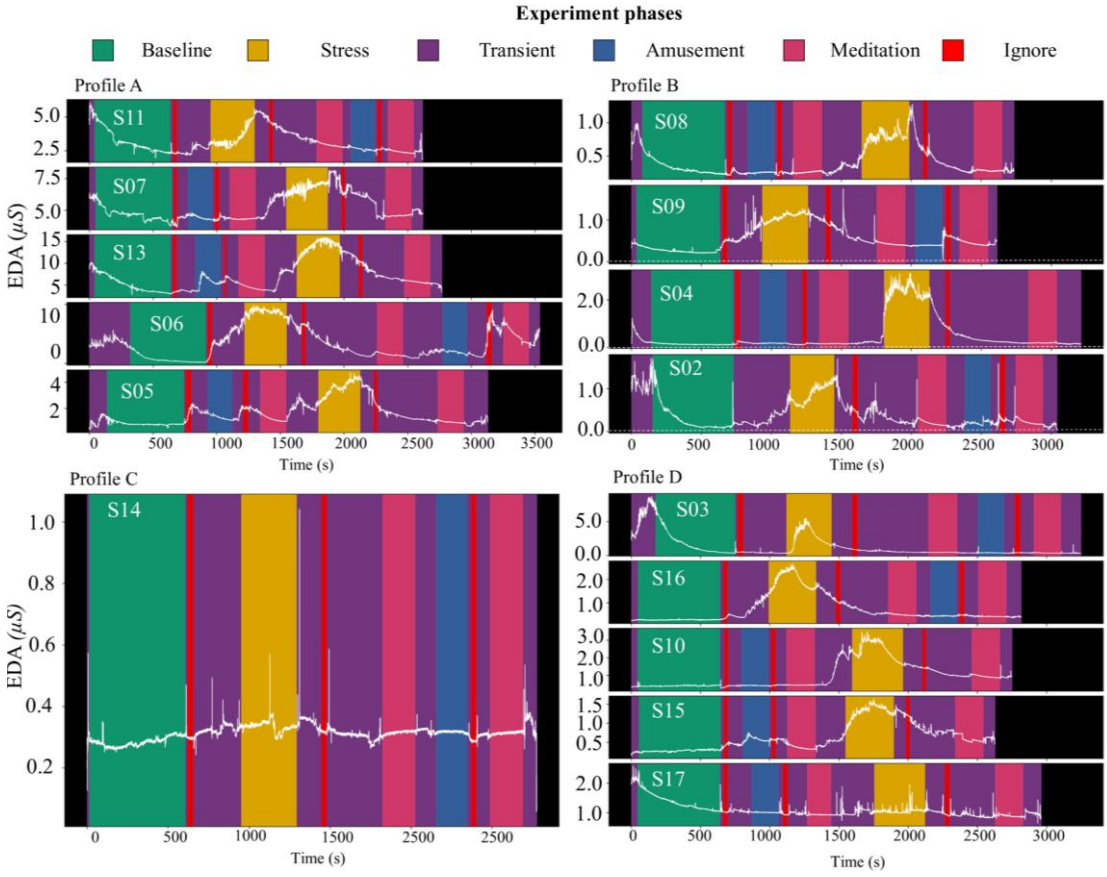


Figure 5.5. Set of user's EDA signals of Profile A.

5.2.5. Take-home messages

The experimentation about user segmentation provides the conclusions as follows:

- Distribution-based clustering was able to extract a user segmentation model.
- BIC was used to determine the number of components and covariance types that maximize the estimator likelihood.
- Among the clusters of users defined by GMM, a user segment named *"Profile C"* could correspond with sensor failures during collecting.
- The rest of the user's clusters exhibited a similar shape on the user's distributions which belong to the same profile.
- In contrast with the user segmentation proposed by Tervonen et al. [136] over the WESAD dataset, the number of user profiles was determined with a heuristic method. Moreover, the EDA signal simplification based on statistics denotes the loosing of discriminant factors, thus obtaining an inaccurate data hierarchization.
- Hereafter, the obtained user profile model leads the subsequent framework steps, preserving this data hierarchy in the conformation of the intermediate datasets and models.

5.3. EDA signal prototyping based on time series clustering

Different studies focused on statistical EDA features in both time and frequency domain; however, within-subject effects and intra-individual variance raise a fundamental issue that becomes critical in real-time analysis. Most machine learning models require data to be processed by scaling or standardization methods to improve data consistency in both the fitting phase and predicting new incoming data. The correct application of this method implies that samples must be enclosed in fixed and known limits and be homogeneous regardless of their context, i.e., that identical samples taken randomly have the same meaning throughout the data set. Nevertheless, these usage restrictions are not accomplished in the scope of EDA analysis. Although user segmentation has provided one hierarchical level in data, individual differences persisted within the formed groups. Consequently, the prototyping of EDA signal was addressed from the viewpoint of signal preprocessing techniques more robust to magnitude variability.

As explained in this chapter's introduction, the characterization of the EDA signal is based on unsupervised learning. In this case, partition clustering was addressed for determining the different states what signal transits. Moreover, the proposed approach must fulfill the real-time constraints, being able to determine the signal state with narrow analysis windows up to ten seconds and, the processing must be transferable to deployment environments.

This section explains how this approach was conducted, exploring the adopted solution for segmenting the EDA signal to simulate the stream processing. After, the processing based on time-frequency space is outlined. Finally, the applied techniques for obtaining a clustering model at level signal are presented.

5.3.1. Processing strategy for conforming stream-based EDA dataset

The first process of the framework involves the transformation of the collected EDA database to construct both training and test datasets, simulating the arrival of streaming sensor data. Figure 5.6 overviews how EDA signals were processed for composing a dataset whose instances are similar to those obtained in real-time conditions. This processing strategy comprises two segmentation levels for treating the signal: initialization windows and fine-grain windows.

As shown in Figure 5.6, the initialization window relies on obtaining a larger signal fragment to make some data processing that requires a minimum signal length for working correctly, such as digital filters.

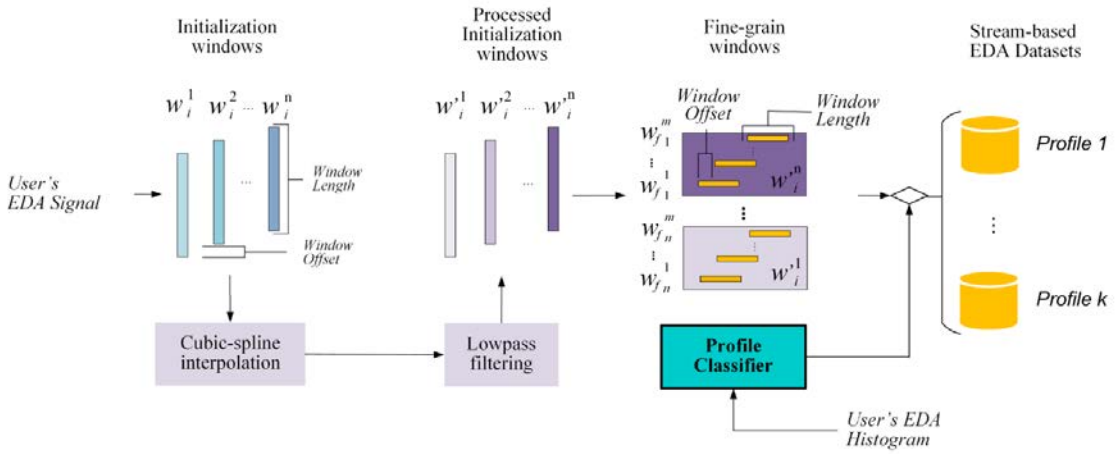


Figure 5.6. EDA preprocessing strategy for the formation of the stream-based EDA dataset.

In this way, each initialization window is resampled with a cubic-spline interpolation to smooth the signal. After that, the resampled signal is cleaned by applying a low pass filter to remove the high-frequency noise from electromagnetic radiation captured by the electrodes [178], thus isolating the EDA signal from alien electromagnetic signals.

The second level of processing relies on splitting each initialization window to obtain small fragments of EDA signal to conduct an ultra-short-term analysis, thus naming "*Fine-grained window*". The processing of this kind of window only introduces an overlapping factor in the extraction of fine-grained windows to avoid the loss of the temporal character of the signal.

This proposed segmentation method exposes a remarkable aspect to consider when generating the dataset. The window offset selection will directly affect the size and diversity of the resulting dataset. At the fine-grain segmentation, window offset will indicate how much difference exists between two consecutive instances. At the initialization segmentation, window offset is inversely related to the oversampling factor, i.e., a window offset of 0.25s will displace the initialization window a quarter of a second with respect to the previously analyzed window.

Furthermore, window offset should be chosen according to the window length of the fine-grain window in order to improve the diversity of the dataset. In this case, selecting

window offsets that keep a proportional relation will produce repeated instances throughout the dataset.

After processing an entire EDA signal, the individual's instances are included in their corresponding user profile dataset, creating a hierarchy of data regarding the determined user profiles.

Experiment labels treatment

Although the experimental phase does not guarantee the reliability of labels, keeping the trace of where a signal window occurred is essential. In this case, each window has its corresponding window of labels. Therefore, it should adopt a strategy to determine which label belongs to the signal window, especially for frontier instances that belong to two experiment phases. The adopted strategy to solve this issue relies on the election of the majority label because of the sequential nature of the data and the time delay of physiological systems [179]. In the opposite case in which two different labels had obtained an equal count value, the first occurred label will be a priority against others.

Parameters and resulting dataset specification

Python 3.7 and some libraries were used for generating the dataset. *Sktime* [180] was used to conduct the proposed signal segmentation based on two levels, *Scipy* [181] was employed to perform the signal resampling, and *Neurokit2* [182] was applied for cleaning the EDA signal.

As the first approach, the parameters from creating the stream-based dataset will be described as follows. Concerning the initialization window, window length and window step were set to 30 seconds and 5 seconds, respectively. The window length and window step were set to 10 and 2 seconds for fine-grain windows, respectively. The sample was upsampled by interpolation from 4 Hz to 8 Hz.

About the EDA preprocessing, the signal was resampled from 4 Hz to 8 Hz, and the cleaning was conducted using the method proposed by *Neurokit2* based on a lowpass Butterworth filter with a cutoff frequency of 3 Hz. Figure 5.7 depicts the instances used for conducting the subsequent analysis distinguishing different waveforms and conductance magnitude across the distinct phase and user profiles.

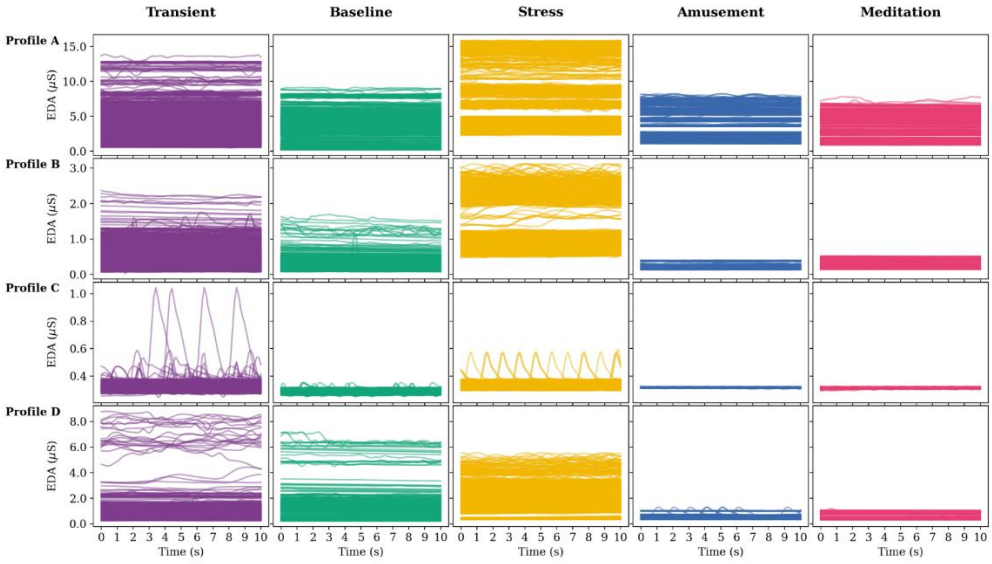


Figure 5.7. Randomly selected examples of the stream-based dataset.

5.3.2. Time-frequency space analysis-based prototyping

Spectrum-based prototyping addresses the partitioning of stream-based EDA through information extracted from analyzing the signal in time-frequency space. The transient and non-stationary nature of the EDA signal raises the use of time-frequency transform as a leading processing method in the stress detection framework. Among time-frequency analysis, transform functions, such as Short-Time Fourier Transform, Wigner-Ville distribution function, and Wavelet transform, provide a frequency analysis over the time represented in two dimensions, thereby allowing to analyze signals whose frequency varies along the time. Unlike Fourier transform, time-frequency transforms are suitable to analyze the non-stationary signal in the application domain with short-term requirements.

This study used continuous wavelet transform (CWT) to preprocess the stream signals, thus converting EDA fragments into two-dimensional energy spectra. Figure 5.8 illustrates how a Wavelet transform enables the analysis of different frequency spectra at each instant of time. Depending on the frequency level of analysis, Wavelet transform varies the time resolution, where high frequencies have better time resolution than low frequencies.

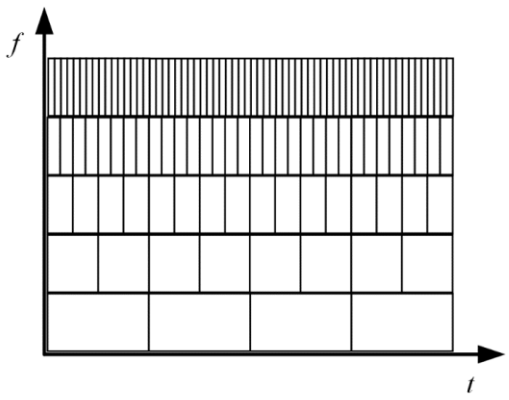


Figure 5.8. Time and frequency resolution of CWT.

The application of wavelet analysis depends on selecting the analysis scale and the mother wavelets. CWT coefficients widely vary when CWT is calculated for the same signal using different wavelets. Consequently, it is necessary to evaluate which set of configuration parameters is the most to tackle the stress detection problems in quasi-real-time.

The Wavelet transform has been applied to address diverse aspects of EDA analysis from the perspective of supervised learning. Concerning distinct mental states recognition, Slukla et al. [183] studied the influence of multiple and diverse features extracted from EDA, including discrete and stationary wavelet transforms for emotions recognition. Feng et al. [184] obtained a model based on complex-Morlet wavelet transform, principal component analysis (PCA), and support vector machine (SVM) to distinguish children's emotional states, including joy, boredom, and acceptance. About EDA signal assessment, Taylor et al. [185] investigated the artifact detection of 5 seconds windows extracting, among others, features from Haar wavelet transform that could indicate sudden changes on the EDA signal. As a result, 96.95% of validation accuracy was obtained applying a workflow that includes feature selection and SVM.

Further, in the study of stress detection based on data gathered from multiple sensors, Hsieh et al. [186] employed chest and wrist sensors from the WESAD dataset to address a feature extraction approach. Among the considered features, both tonic and the phasic component of EDA were decomposed through a Daubechies-3 wavelet to extract some statistical and entropy features from the wavelet coefficient.

One of the advantages of frequency analysis is that input signals do not require any normalization to homogenize to scale in time domain across different subjects. However, Wavelet transform is a parametric function, and its configuration is sensitive to the application domain. Such as was described by Addison [187], a continuous wavelet transform of a signal *s(t)* in real numbers domain applying the mother wavelet $\psi(t)$ at scale *a* and location *b* is defined as (5.4):

$$T(a,b) = \frac{1}{\sqrt{a}} \int s(t)\psi\left(\frac{t-b}{a}\right) dt \tag{5.4}$$

Analytically, computation of CWT coefficients refers to the similarity of the signal with the wavelet function at a specific scale [188]. Consequently, obtaining the most suitable wavelet transform according to the EDA signal characteristics collected with a wearable sensor requires exploring multiple wavelet settings, including the wavelet family, wavelet order, and the examined frequency spectra.

The reviewed literature highlighted the need for conducting feature selection methods when wavelets are applied due to the dimensionality of the resulting energy spectrum. When machine learning techniques are applied, instances with too many attributes drastically reduce the ability to converge to optimal or sub-optimal solutions in a reasonable time; therefore, dimensionality reduction is required. Among dimension reduction methods, widely used principal component analysis (PCA) determines a set of linearly uncorrelated variables by analyzing a set of observations, thus creating a new low-dimensional space that preserves the original data's variance as much as possible [189].

The new projection of data leads to applying partition clustering techniques to distinguish the different states of the signal regardless of data labeling. In this way, k-Means characterize the EDA time-spectrum on the PCA space by a set of prototypes. This approach has been addressed in a wide diversity of domains.

Proposed methodology

According exposed above, a methodology has been proposed to introduce time-frequency analysis and the distinct aspects that involve obtaining EDA signal. Figure 5.9 overviews the methodology that addresses cited aspects above, where PCA and k-Means intervene in the CWT analysis of the EDA signal. Notice that this workflow defines procedures in which multiple datasets and intermediate models are generated.

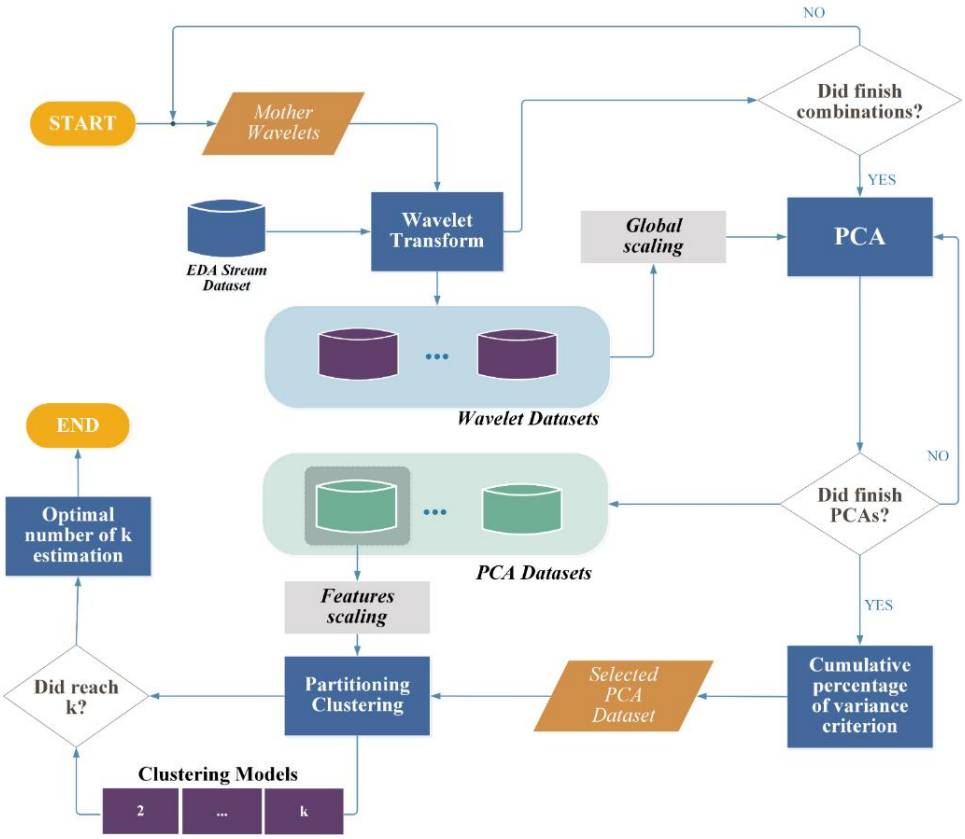


Figure 5.9. Methodology for EDA characterization based on time-frequency space analysis.

A wavelets transformation database of EDA signal fragments is generated by exploring the wavelet settings involving the wavelet family and the wavelet order. From that, each wavelet is evaluated with PCA with the proposed of determining which wavelet mother minimizes the information loss.

Decision-making about the best-fitted model among several models of the same kind is led through statistical heuristics that allow for rating them according to objective criteria. PCA models regarding each wavelet transform setting were obtained. The criterion for selecting the best orthogonal basis for producing the wavelet projections was conducted by analyzing the cumulative percentage of the total variation, as explained by Jolliffe [190]. It

relies on analyzing which PCA model conserves in the first *m* principal components most of the information of its corresponding wavelet data.

Third and finally, determine how the sample space can be partitioned, thus finding a set of prototypes that accurately represent the wavelet projection landscape. In such a manner, partition clustering models within several numbers of partitions are fitted using the resulting dataset from PCA transformation, hereafter called PCA dataset. Before the training process, PCA attributes should be scaled to facilitate the convergence of the model.

Once all models have been fit, a model selection criterion should be applied to determine which model provides better clustering results. For example, the study conducted by Raj et al. [41] proposed a method based on assessing elbow and silhouette methods parallelly. However, due to the high number of clustering models across the distinct user-profiles and the computational complexity of silhouette index ($O(n^2)$), it proposed another way for determining the optimal number of clusters from the location Knee point that avoids exceeding the *k* value that significantly improves the partition results in relation with the previous one.

Figure 5.10 describes the proposed algorithm called "*Optimal Estimation of k for a large number of clustering models*".

```
Inputs:
    -   KM(n): List of fitted clustering models
    -   WCSS(n): within-cluster sum of squares of the n models
    -   S: Dataset

Output: Optimal number of K

INIT

    K1 = ElbowLocation(WCSS, polynomial_degree=2)
    K2 = ElbowLocation(WCSS, polynomial_degree=3)

    best_k = K1
    best_silhoutte = -infinite

    FOR i=K1 TO K2 INCREASE BY 1:
        labels = predict (KM(i), S)
      silhoutte_index = Silhoutte(S, labels)

        IF silhoutte_index > best_silhoutte:
          best_k = i
        END IF

  END FOR

  RETURN best_k

END
```

Figure 5.10. Optimal number of k estimation algorithm based on optimist and pessimistic Elbow location.

According to the Knee/elbow location algorithm proposed by Satopaa et al. [191], the proposed method is based on the optimistic and pessimistic location of the curvature point defined by the function WSCC of k. The curvature point detection algorithm largely depends on the shape described by the WSCC curve and the polynomial degree used for interpolating

the WSCC curve. A higher-order degree will locate the optimal k in the lower value (optimistic estimation), and lower order will locate optimal k in the higher value (pessimistic estimation). In this way, optimistic and pessimistic locations are obtained with polynomial interpolation of 3rd order (K_1) and 2nd order (K_2), respectively.

The range within K_1 and K_2 provides a limited number of candidates to determine the optimal value of k for which the Silhouette index is computed. The optimal number of k will be the maximum value of silhouette s(i) within the interval $i \in [K_1, K_2]$.

This algorithm allows examining the silhouette values of a reduced number of k in order to avoid computing the silhouette index for all range of trained models, thus significantly reducing the computational time for obtaining the best estimation of k.

Once the optimal value of k has been estimated across all user profiles, the obtaining process of EDA characterization models is concluded

Experimentation

The entire process was developed with Python 3.7. The EDA stream dataset was generated for this experiment following the methodology explained in section 5.3.1. However, level 1 of segmentation was not normalized in this case because the spectral analysis was invariable to signal magnitude. As parameters for dataset formation, first-level windows length and windows interval were set to 30 and 5 seconds, respectively; second-level window length and window interval were set to 10 and 2 seconds, respectively. As a result, the dataset was formed for 184927 instances. Table 5.4 describes the distribution across the different experimental phases and user profiles.

CWT over EDA stream dataset instances was applied in this study, obtaining the wavelet dataset that includes the different selected wavelet transforms. In order to do that, *Pywavelets* library [192] was used to generate the different transform all segmented EDA windows. Among the different mother wavelets, only wavelets that provide transform in real subset were considered because, in complex wavelets, the imaginary domain is usually employed to analyze the signal phase.

TABLE 5.4. STATISTICS OF INSTANCES USED FOR OBTAINING THE CWT DATASETS.

Profile	Transient	Baseline	Stress	Amusement	Meditation	Total
A	29466	13033	7214	4110	8706	**62529**
B	23356	10232	5643	3248	6955	**49434**
C	5120	2596	1485	818	1747	**11766**
D	28077	12883	7585	4088	8565	**61198**
Total	**86019**	**38744**	**21927**	**12264**	**25973**	**184927**

Accordingly, wavelet families including Gaussian, Ricker (also known as Mexican hat), and Morlet wavelets were selected as mother wavelets. The selection of wavelet scale is a determining factor in computing transformations. Due to its signal nature, the spectrum of low-frequency spectrum carries a big part of information about how EDA behaves. Hence, eight voices per octave with 2^4 spacing between scales. As a result, each EDA signal was converted into an 80x16 matrix, having 1280 attributes per instance.

Figure 5.11 reveals the behavior differences of the distinct wavelet families applied over the same signal. As observed, signals taken from distinct experiment phases exhibit highly significant differences concerning the magnitude of the wavelet coefficients.

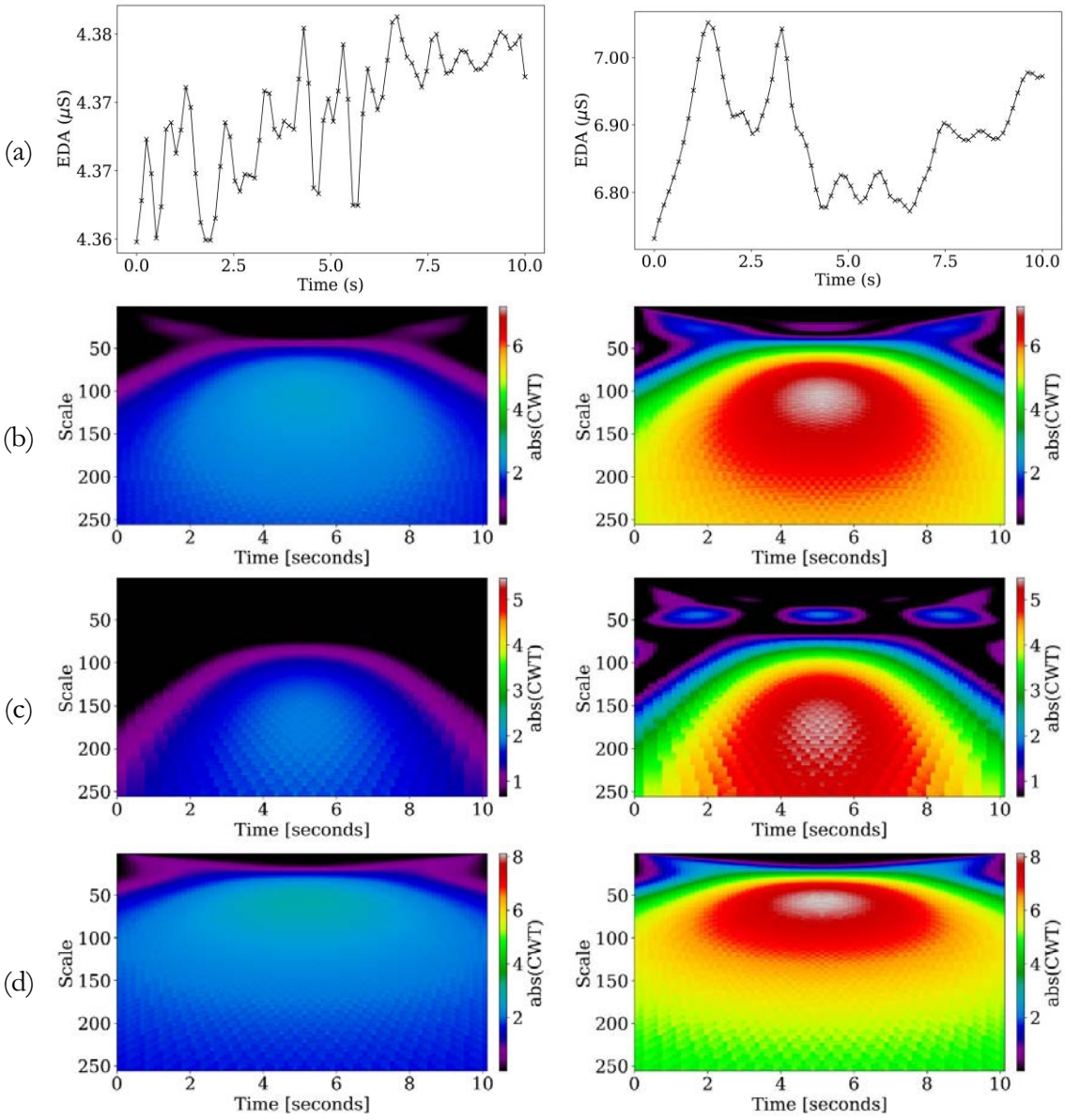


Figure 5.11. Two examples of EDA signals (a) with their corresponding scaleograms using distinct wavelet mothers ((b) Gaussian 4th order; (c) Morlet; (d) Mexican hat) with 256 voices per octave taken from meditation phase (left column) and stress phase (right column).

Moreover, considering that scale is inverse to the frequency, range of scale between 4 and 128 corresponding to mid-range and low-frequency spectrum provide highly relevant information about how signal behaves. As a result, a 19 GB wavelets database was generated with this process.

After, PCA has applied over all wavelets datasets previously scaled with the purpose of limiting the range. In this case, maximum absolute value scaler was considered among the wide variety of scaling methods because it does not shift center and consequently does not

modify the sparsity of the original data. Furthermore, this scaler was applied through global statistics of CWT coefficients, leaving aside the feature-based scaling approach.

Scikit-learn [193] was used for conducting the dimensionality reduction analysis. Given a fixed number of components, the explanatory ratio was the selection criteria to determine with the wavelet family better represents the EDA signal from each user profile. The selection of best fitted PCA led to transform the wavelet datasets into their principal components.

Thus, k-Means were applied to obtain the sample landscape partitions in space defined by resulting PCA and, *Scikit-learn* was used for computing these models as well. As a previous step of the fitting process, the PCA dataset was scaled applying a feature maximum absolute value scaling. Fourteen clustering models were fit varying k values from 2 to 16, aiming to determine which k is the most appropriate for representing the sample space. *Kneed* package [191] was used to detect the optimistic and pessimistic Knee points over the resulting WCSS across the distinct values of k and the proposed algorithm determined the optimal number of k.

Finally, the selected k-Means were used to classify the spectrum PCA projections of EDA to their corresponding partition, thereby characterizing all EDA fragments for the next framework steps.

The whole described process was conducted across distinct user profiles.

5.3.3. Results

Concerning the principal component analysis, Figure 5.12 compares which mean percentage of variance the PCA model preserves with the three principal components across the user profiles. Notice that all results presented an unsubstantial information loss, and the worst result reached a mean cumulative explained variance ratio over 98%. As it can observe, the explainable rate from Mexican hat wavelet transformation is quite significantly higher than the others wavelet mother transformations.

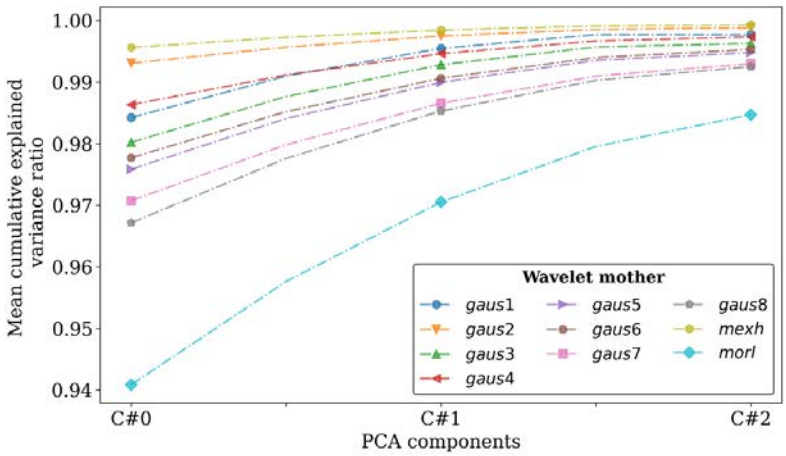


Figure 5.12. Mean values of cumulative explained variance ratio of the wavelet mother across the distinct user profiles.

Table 5.5 indicates that the best PCA model was obtained with the Mexican Hat wavelet across all profiles. Although all results are pretty similar, minimum value and standard deviation of cumulative explained ratio across was determinant to decide across all wavelet mother functions.

TABLE 5.5. STATISTICS OF THE CUMULATIVE EXPLAINED RATIO OF PCA USING THE DISTINCT WAVELET MOTHER ACROSS USER PROFILES.

Wavelet mother	Cumulative explained ratio statistics			
	min	max	mean	std
Gaussian 1st	0.991	1.000	0.998	0.004
Gaussian 2nd	0.995	1.000	0.999	0.002
Gaussian 3rd	0.986	1.000	0.996	0.007
Gaussian 4th	0.990	1.000	0.997	0.005
Gaussian 5th	0.980	1.000	0.995	0.010
Gaussian 6th	0.982	1.000	0.995	0.009
Gaussian 7th	0.973	1.000	0.993	0.013
Gaussian 8th	0.971	1.000	0.993	0.014
Mexican hat [1]	0.997	1.000	0.999	0.001
Morlet	0.942	0.999	0.985	0.029

[1] *Selected wavelet mother*

Despite PCA models were not fitted using labels, Figure 5.13 illustrates how EDA is distributed across experiment phases after data reduction obtained with PCA. The shape of data concentrations is easily distinguishable between amusement, meditation, and stress phases. The baseline's cloud of points overlaps considerably with the other experimental phases.

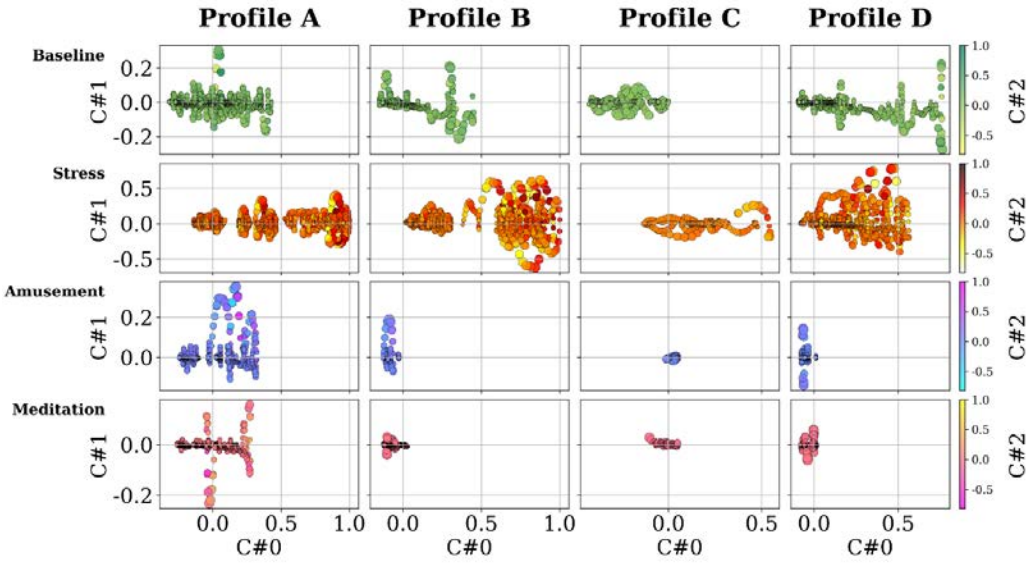


Figure 5.13. PCA projections of the wavelets transform of EDA signal with Mexican hat wavelets of 16 per octave (c#0: x position; c#1: y position and size; c#2: marker color tone).

Regarding selecting the best-fitted k-means that partitions PCA projections, Table 5.5 shows the within-cluster sum of squares (WCSS) across the k-means models obtained with k values between 2 and 16. As it can observe, the method of detecting the elbow point in the

function defined by the progression of WCSS along k provided different results according to the selected interpolation method, thus estimating the range in which the optimal number of partitions is. After the fitting process, centroids were sorted according to the distance from the origin of PCA space to facilitate the later data interpretation.

TABLE 5.6. RESULTS OF K VALUE ESTIMATION WITH ELBOW METHOD APPLYING A POLYNOMIAL INTERPOLATION FUNCTION.

	Within Cluster Sum Squares (WCSS)			
k	**Profile A**	**Profile B**	**Profile C**	**Profile D**
2	1930.809	921.542	244.199	608.286
3	1233.459	446.956	118.161	394.323
4	865.613	358.180	93.990	301.492
5	702.338	325.598	80.267	266.969
6	622.781	**298.513**[b]	**66.539**[b]	232.046
7	**586.498**[b]	280.413	57.465	**208.708**[b]
8	**489.938**[a]	**244.537**[a]	**53.882**[a]	187.272
9	487.613	227.377	42.031	**170.192**[a]
10	447.395	218.514	43.147	165.732
11	425.135	202.569	35.430	149.308
12	401.330	187.116	34.424	146.550
13	379.492	179.834	31.200	137.760
14	367.448	174.332	29.294	137.270
15	349.693	165.100	27.282	129.381
16	338.839	161.340	24.856	116.549

[a] 2nd order degree, [b] 3rd order degree

After evaluating the results of Elbow location applying a pessimistic and optimist location of elbow points, the elbow points were situated within the indicated ranges of *k* for each profile. Accordingly, the Silhouette coefficient was applied for these specific values of k to determine which number of partitions is the most consistent. Table 5.7 exposes the silhouette for a reduced range of k values across the distinct user profiles. As can be observed, the higher silhouette value varies across the examined ranges of k across the distinct user profiles.

TABLE 5.7. AVERAGE SILHOUETTE SCORE FOR THE REDUCED RANGE FOR SELECTING THE OPTIMAL VALUE OF K.

	Average Silhouette Score			
k	**Profile A**	**Profile B**	**Profile C**	**Profile D**
6	-	0.404	0.486	-
7	0.378	0.407	**0.486***	0.595
8	**0.450***	**0.518***	0.455	0.595
9	-	-	-	**0.596***

[*] Select values of k

Figures 5.14, 5.15, 5.16 and, 5.17 depict how separated the partitions are for the highest scored *k*-Means models of Profile A, B, C, and D, respectively. For cases wherein different k values obtained the same silhouette score, the highest value of k was selected to increase the number of partitions.

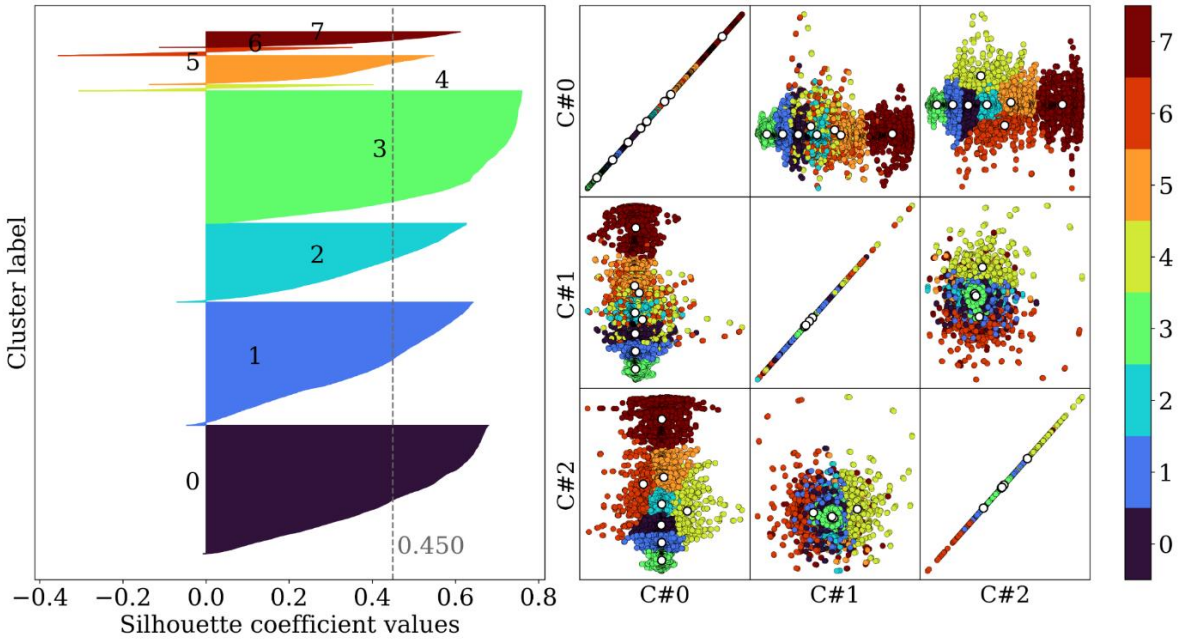


Figure 5.14. Silhouette coefficient and clustering classification of PCA projections for profile A.

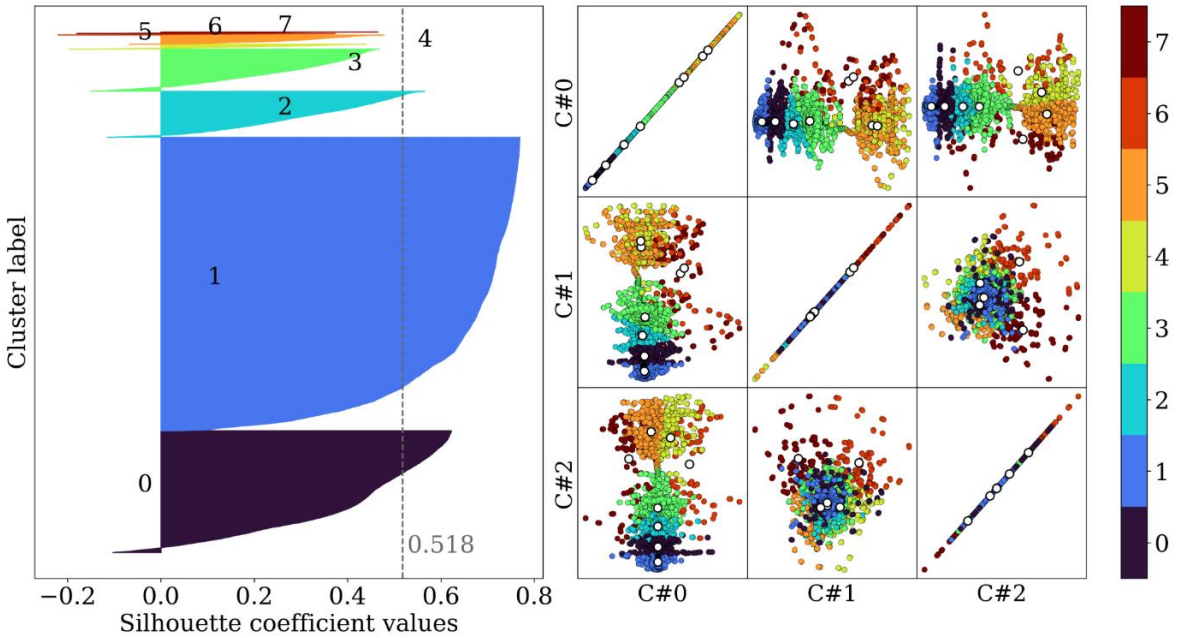


Figure 5.15. Silhouette coefficient and clustering classification of PCA projections for profile B.

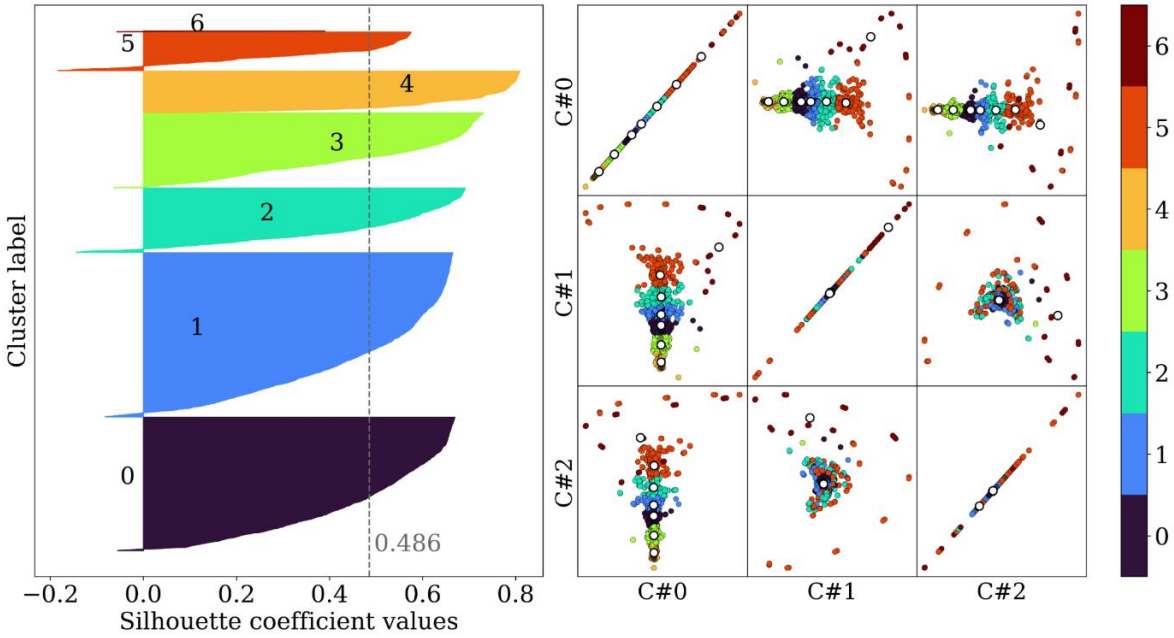


Figure 5.16. Silhouette coefficient and clustering classification of PCA projections for profile C.

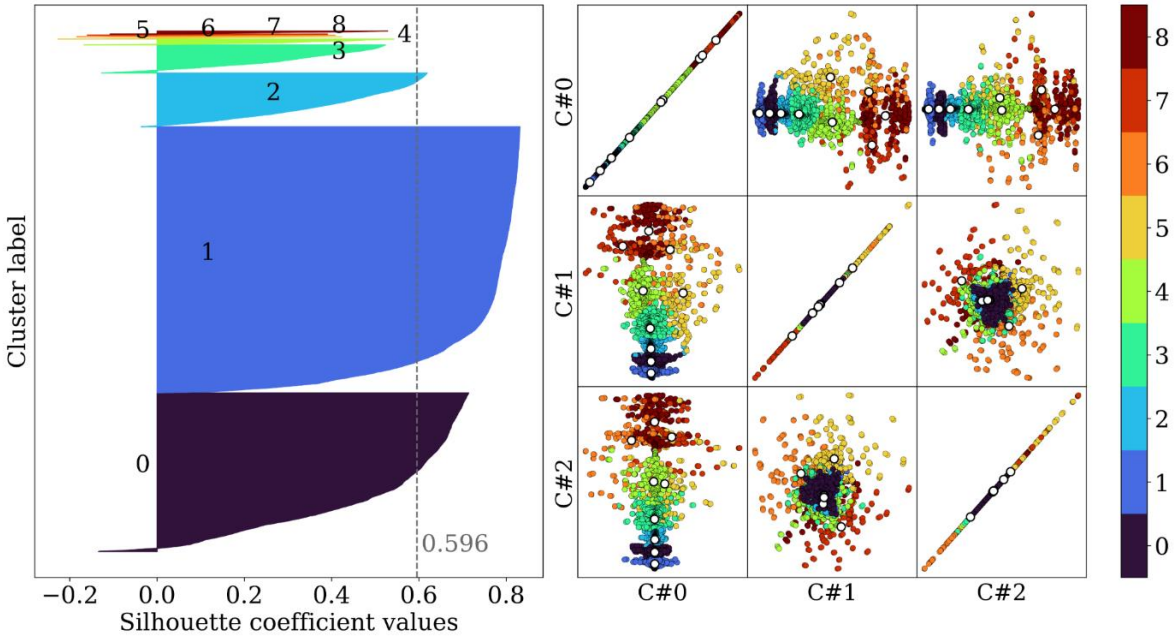


Figure 5.17. Silhouette coefficient and clustering classification of PCA projections for profile D.

The bi-dimensional representation of PCA projection corroborates that sample space was appropriately partitioned. Concerning silhouette samples representation, across all *k*-means models, the instances near the partition limits were either misclassified or undetermined between the assigned cluster and the second nearest cluster, thus indicating that partitions boundaries are pretty diffuse. However, considering the overlapping of

instances in the distribution, partition results determine the different states of the EDA signal.

5.3.4. Take-home messages

The findings of the process of characterizing the signal state are summarized as follows:

- Clustering is addressed to the characterization EDA fragment of 10 seconds.
- Time-frequency analysis was conducted to transform EDA, and several wavelet mothers were assessed with PCA to determine which wavelet produces the lowest loss of information in the representation of the whole dataset of EDA fragments.
- K-Means were applied to create several partitions in the PCA projections of the time-frequency space transformations that determine different states on the EDA signal.
- The variability of results regarding the characterization of EDA signals across the distinct user profiles corroborates maintaining the hierarchical level of variance.

5.4. Sequential pattern mining

Independently partial labeling provided by the experiment temporal bounds, centroid-based methods have demonstrated that different partitions exist on the sample space of EDA signals. User segmentation and all data processing across the different framework stages to characterize EDA signal provides a snapshot of where an EDA window is located at each evaluated signal period, thus defining a finite set of states. However, considering each instance as an isolated piece of information eliminates the temporal nature of the signal, losing information about which state had the signal in previous periods or how transitions between states behave in specific moments. Therefore, preserving the data sequentiality could more accurately distinguish stress elicitation from other cognitive states.

Despite the relevance of Sequence analysis in Bioinformatics, searching relevant patterns given a sequence of discrete symbols is a barely explored field in stress recognition. Applying time series clustering facilitates the paradigm change from signal analysis towards a frequent string mining problem such as was described by Vilo [194]. Accordingly, this approach studies how a finite set of symbols that characterize the most representative events in EDA signals are sequenced to form patterns and which event patterns are statistically associated with either stress elicitation or common activities.

Sequential classification of symbolic representation of EDA signal has been applied in different contexts. Jambigi et al. [195] addressed the reliability of labeling outcomes in the crowd working, analyzing the EDA signal to determine how difficult each labeling task for the annotators was. For analyzing of EDA signal, dictionary-based classifiers in the context of time series, including Bag of Symbolic Fourier Approximation Symbols (BOSS) [196] and word extraction for time series classification (WEASEL) [197], were applied to determine stress levels and distinguish between simple and hardly labeling tasks, being tested both individual-centric and general models.

Henriques et al. [198] addressed the emotion recognition with EDA signal in the human-robot interaction. Their approach included Symbolic ApproXimation (SAX) [199] to transform an EDA signal into an equiprobable discrete set of alphabetic symbols.

This presented study introduces Trie (retrieval tree) to structurally represent the transitions across the set of finite states resulting from partitions of EDA determined with clustering methods. Due to its efficiency, Tries are commonly used to construct frequency-based patterns in dynamic environments. Independently the partition clustering methods used, each defined data partition was designed with an alphabetic symbol; thereby, the classification of all EDA signal provide an analyzable set of symbol sequences categorized by the experiment phases where they were extracted. The distinct Tries construction was conducted to evaluate the relationship between a symbol and its prefixes by Chi-square test, such as was implemented by Iglesias et al. [200]. Moreover, some metrics based on information retrieval have been adapted to evaluate the relevance of the set of subsequences regarding stress identification.

5.4.1. Tries formation

Considering a reliable method for describing the different states through which small fragments of signal transits, each experiment phase can describe the distinct sequences of states that occurred. Partition methods of EDA signal, such as described in Section 5.3, allow transforming the signal into a symbolic representation using the partition identifiers. This domain transformation facilitates analyzing the dependency within the finite set of EDA states and determining the sequences of states more related to a specific physiological state. In this way, the symbolic transformation was conducted with the formation as follows (5.5):

$$argmin_c \sum_{c=0}^{K} d\,(X, c),\ \forall\, X\ =\ (x_1, x_2, \dots, Xn) \in D \tag{5.5}$$

where function *d* is the distance function used to measure the similitude between the centroid c and the short-term EDA signal *X* from the dataset *D*. By means the symbols representing the nearest centroids corresponding to each instance, a discrete set of events can represent the short fragments of EDA.

The methodology proposed by Iglesias et al. [200] described the process for analyzing sequences. First, given a sequence of symbols $S = S_1, S_2, \dots, S_n$ extracted from a specific experiment phase conducted by one user, the first step consists of segmenting into *m* subsequences of fixed length *L* that established the number of states whose dependency is analyzed. Accordingly, events were taken and ordered in sequences of specific length to compose a dataset regarding user profile as follows (5.6):

$$S\ =\ \left\{\left\{\left\{e_j\right\}_{j=1}^{L}\right\}_i, y_i\right\}_{i=1}^{n}, \forall e_j \in K, \forall y \in L \tag{5.6}$$

where e_j will be the event that represents the assigned cluster to each EDA signal, y_i is the label corresponding to the sequence of events and whose value will be the experimental phase, and K is the identifiers set corresponding to the centroids of the clustering model.

Extending this procedure to all experiment phases and profile users permits collecting a database of event subsequences.

Second, all event subsequences of each experiment phase and profile are stored in prefix Trie. This tree structure is constructed by the hierarchy defined by the prefix of the occurrence of an event taken as the parent node and its successive events considered as child nodes. During the Trie construction, the insertion of each event kind is accounted for.

Once Tries are constructed, the definition of path length is required to fulfill the ultra-short-term requirement. In this way, the selection of *l* events of length paths was extracted to determine their relevance.

5.4.2. Proposed metrics

Interpreting the obtained subsequences model requires the definition of some metrics that provide comprehensive information about the subsequences, allowing rank the

relevance of sequences to identify stress patterns. Some direct statistics such as the occurrence frequency and the rate of individuals producing a subsequence were combined with other proposed metrics adapted to this domain. These proposed metrics are formally presented as follows.

a. Interphase collision rate (ICR)

According to this domain, some metrics are proposed to quantify how relevant the sequences of the subsequences model are, especially the stress model. In the same way that TF-IDF index [37] is used in text mining and information retrieval to measure the relevance of the set of terms that contains a corpus, Interphase collision rate (ICR) has been stated as an index to determine the specificity of each instance of the subsequence model. The ICR of a subsequence that appears in a specific experiment phase is defined as follows (5.7):

$$ICR_q^h = 1 - \frac{q(s,h)}{\sum_{i=1}^{H} q(s,i)}, \tag{5.7}$$

where *q(s, h)* and *q(s, i)* are the numbers of appearances of the sequence *s* in the phases *h* and *i*, respectively. *H* is the total number of experiment phases. ICR is in the closed interval [0, 1] in which the value 0 implies the highest degree of specificity because sequence *s* is tightly linked to the phase *h*, whereas the value 1 implies that the occurrence of sequence s is fairly common.

b. Inter-individual occurrence rate in profiles (PIOR)

With the purpose of quantifying how many distinct individuals have experienced a specific subsequence of a model, Inter-individual Occurrence Rate in profiles (PIOR) was proposed to measure the generalization ability of a subsequence, thus indicating the degree to which a subsequence is linked to a specific phase across individuals with belongs to the same user segment. PIOR value only involves the individuals of the corresponding user profile and is defined by the formula (5.8):

$$PIOR_q^h = {i_q^h}\Big/{(i_q^h + \sum_{k=1}^{P} i_q^k - (i_q^k \cap i_q^h))}, i \subseteq I_P \tag{5.8}$$

where i_q^h is the size of the subset of individuals who belong to the profile P, and which experienced the subsequence at least once in the phase h. One key point in the computation of PIOR is the attenuation of the impact that the individuals i_q^h cause in this rate. The individuals who produced the subsequence q are more susceptible to reproducing it in other phases. Consequently, the occurrence of the subsequence q in the phase *k* for individuals who had already produced this subsequence in the phase p were excluded.

c. Relevance rate of subsequences and stress patterns identification (r-Rate)

As previously explained, the definition of ICR and PIOR metrics allows quantifying the singularity of a subsequence across the experiment phases and its commonness across the individuals of the same user profile. The combination of both factors constitutes the

criteria for rating the relevance of each subsequence. Accordingly, the rating of a subsequence q extracted from the experiment phase h is defined as (5.9):

$$r\ Rate_q^h = \begin{cases} 0, PIOR_q^h = 0 \cap ICR_q^h = 1 \\ \alpha\ PIOR_q^h \times p + \beta(1 - \ ICR_q^h),\ otherwise \end{cases} \quad (5.9)$$

where α and β are arbitrary values that weigh the subsequence's relevance rate computation and p is the rate of persons who experience the subsequence. A higher value of α will prioritize the uniqueness over the generalization ability, whereas a higher value of β will prefer the set size of individuals that experienced the assessed subsequence over its specificity.

d. Metrics interpretation for generating results

The validation stage relies on quantifying the number of unique individuals represented by the resulting stress subsequence patterns models. Thus, the computation of the disjoint set of individuals across the subsequences with an r-Rate value higher than a threshold determines the rate to identify stress patterns.

Despite ICR and PIOR indexes providing an accurate description of each subsequence individually, these metrics do not guarantee that isolated subjects are major contributors to the set of subsequences formation regarding stress patterns. Accordingly, the validation stage focuses on determining the ratio of unique individuals represented with the stress model. Figure 5.18 summarizes the process to validate the resulting set of subsequences that compose the stress model.

The process is led by the classification task of all individuals' event subsequences, thus extracting each subsequence's ICR and PIOR indexes, besides the interrelation between the distinct subsequences and the individuals who underwent each of them. After that, the computation of R-Rate allows selecting the *nth* first subsequences with the highest R-Rate values, thus extracting the most relevant subsequences regarding stress a priori. Later, the cumulative users set was computed to assess the generalization ability of the stress subsequence models. Graphing stats, including cumulative user rate, individual rate, and r-Rate, provide a global perspective about how many subsequences are needed to represent the stress elicitation of all users.

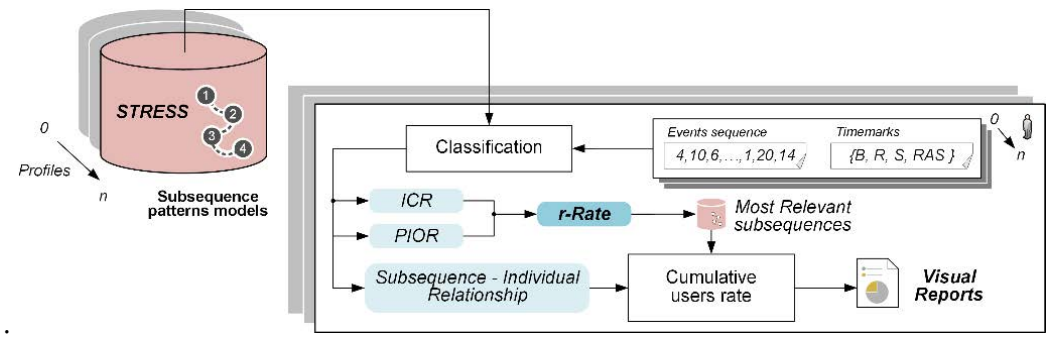


Figure 5.18. The validation process of obtained stress subsequences models.

5.4.3. Construction of directed weighted graphs

The proposed r-Rate measure states the relevance of all sets of subsequences found at each experiment phase. The construction of directed weighted graphs from sequences that pass a minimum threshold enables representing how EDA states and their transitions behave across the distinct experimental phases. Accordingly, the subsequences subset with an r-Rate above a specific threshold is used as inputs to compute the adjacency matrix in each profile. Disposing of all possible EDA states as a square matrix, each matrix position indicates the weight of transits from state A to state B. The weight of transition weight is determined by the occurrence rate of transitions along the subsequences set of the examined experimental phase.

5.4.4. Experimentation

From the signal state model previously presented, all instances across profiles were classified to obtain a symbols dataset whose values correspond with the predicted partition identifiers. After, symbol sequences were differentiated regarding the experiment phase for each profile's user and, each sequence was segmented in subsequences of 12 symbols, corresponding to 30s of signal retrieving the original window length used for segmenting EDA signal. The generated dataset was stored in separated text files in CSV format regarding the experiment phase where they were extracted and the user profile wherein subsequences belong. Although sequences regarding transient labels were obtained as well, for this experiment, they were excluded because it corresponds to periods in which subjects did not perform a specific task. Table 5.8 resumes the amount following the procedure explained above.

TABLE 5.8. THE NUMBER OF SUBSEQUENCES ACROSS EXPERIMENT PHASES AND USER PROFILES.

	Profile A	**Profile B**	**Profile C**	**Profile D**
Baseline	1185	930	236	1171
Stress	685	513	135	691
Amusement	374	296	74	370
Meditation	792	631	159	779

As was explained, the implementation of the algorithm reported by Iglesias et al. [200] was used for generating the prefix Tries, and all paths of 2,3,4 and 5 symbols were obtained. The length selection depended on the diversity of ICR across the found subsequences, being three symbols the selected length for conducting this analysis. Three symbols of length correspond to six seconds of analysis; therefore, it accomplishes the initial requirements.

For computing the r-Rate metric, α and β were set to 0.25 and 0.75, respectively. Accordingly, subsequence uniqueness regarding a phase prevails over the generalization ability of that subsequence.

Regarding the construction of the adjacency matrix, the selected 0.5 was the selected r-Rate threshold to consider relevant the subsequences related to specific phases. In some cases, such as mediation and amusement, r-Rate does not reach 0.5. Consequently,

subsequences were decreasingly sorted by r-Rate, and the first 50% of the subsequences were selected.

5.4.5. Results

Figure 5.19 shows how the ICR was distributed for all subsequences of 2, 3, 4, and 5 symbols of length. As observed, 2-symbols subsequences exhibited a flattened distribution, indicating that all subsequences found in one specific phase have the same rate of occurrences in the other phases. However, 3-symbols and successive lengths presented a more considerable variability regarding ICR; thereby, recognizable patterns appear in stress and baseline phases across profiles because some subsequences got a low rate of ICR. This fact demonstrates that a 3-symbols length subsequence is the minimum information unit for interpreting and analyzing the subsequences incidence during cognitive states elicitation.

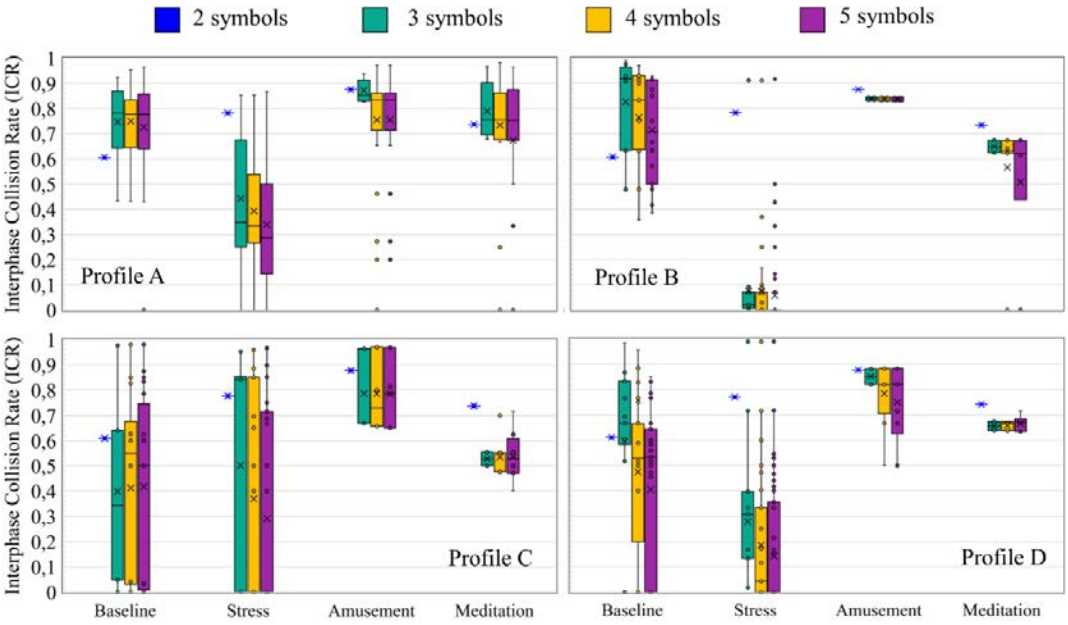


Figure 5.19. Interphase collision rate (ICR) distributions of subsequences of length 2, 3, 4, and 5 symbols across experiment phases.

Figure 5.20 shows the 25 highest r-rates subsequences corresponding to the stress phase. As it can observe, the shown subsequences obtained an r-Rate between 0.6 and 0.95, being a value still far away from the estimated threshold of 0.5. ICR indicates that shown subsequences presented high specificity, whose occurrence in other phases barely exceeds the 0.25. Obtained PIOR indicates a high significative generalization ability in which the set of individuals which produce the shown subsequences set occurred in the stress phase most of the time. Overall, the constituted subsequences present differences between their consecutive elements or stay in the highest EDA states.

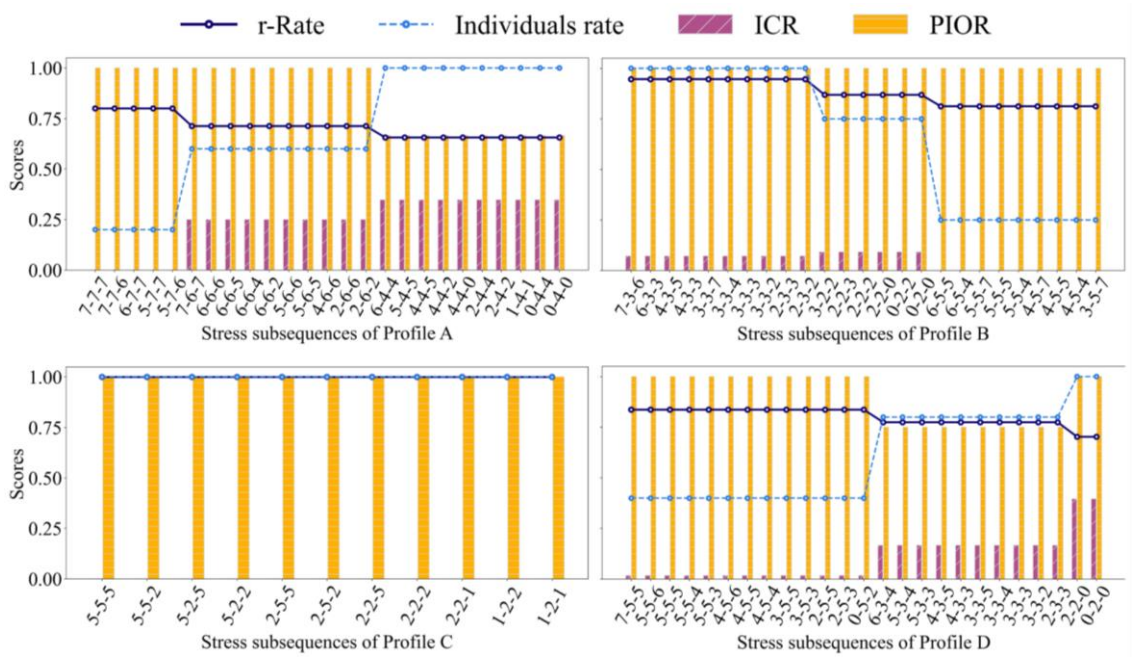


Figure 5.20. Stress subsequences across the distinct user profiles.

Concerning baseline subsequences, Figure 5.21 shows 25 (in possible cases) highest r-Rate of this phase. As it can observe, the r-Rate of subsequences dropped significantly concerning the stress subsequences due to the high value of ICR that indicates the low specificity of subsequences found in the baseline phase. Peculiarly, Profile D showed elevated levels of specificity in the subset of subsequences formed with high-level EDA states.

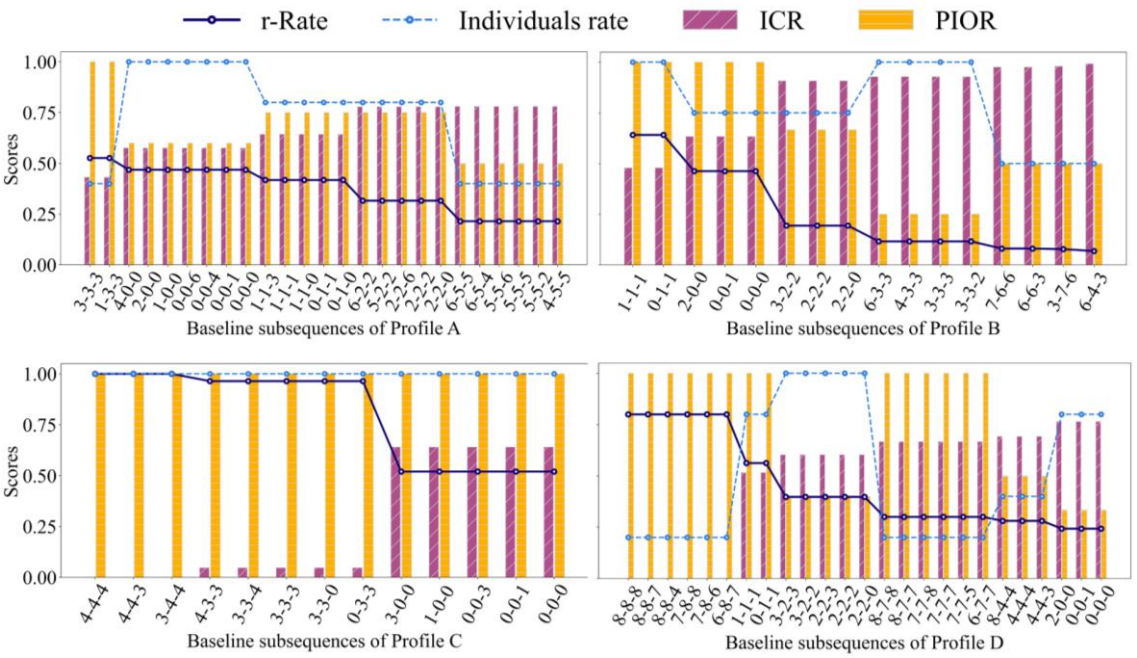


Figure 5.21. Baseline subsequences across the distinct user profiles.

Figure 5.22 shows the 25 (in possible cases) highest r-Rate subsequences of the amusement phase. In general, subsequences in the amusement phase are very homogeneous across all user profiles presenting low values of r-Rate because of their high values of ICR.

All subsequences are formed with low and mid-range EDA states that indicate this high collision rate with other phases.

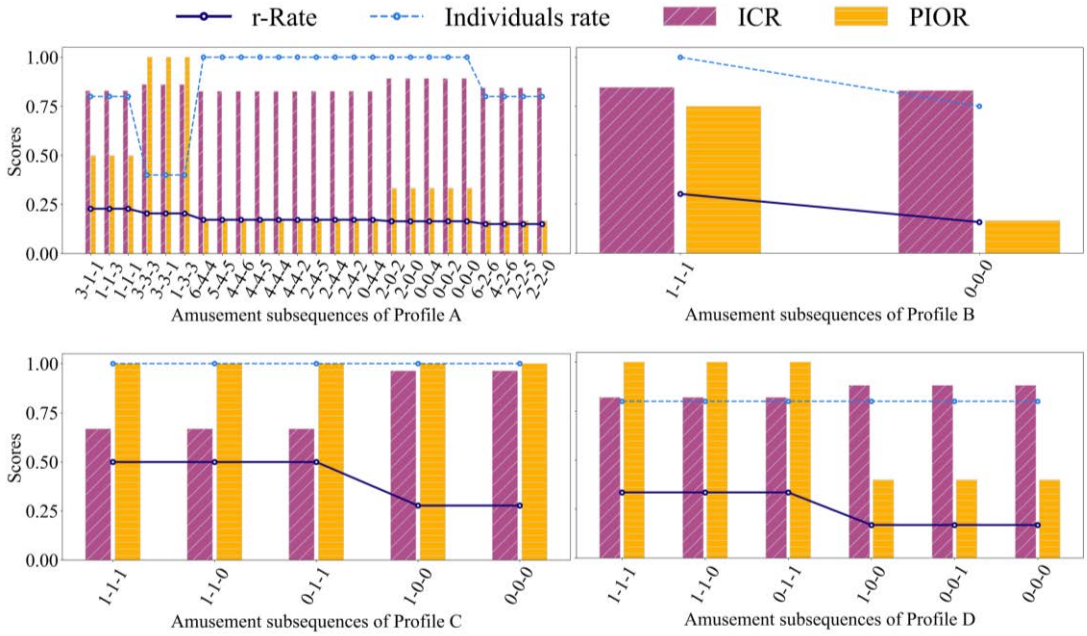


Figure 5.22. Amusement subsequences across the distinct user profiles.

Concerning meditation models, Figure 5.23 shows the 25 (in possible cases) highest r-Rate. Majorly, symbols corresponding to the nearest partitions from the PCA space origin intervene in the formation of meditation sequences.

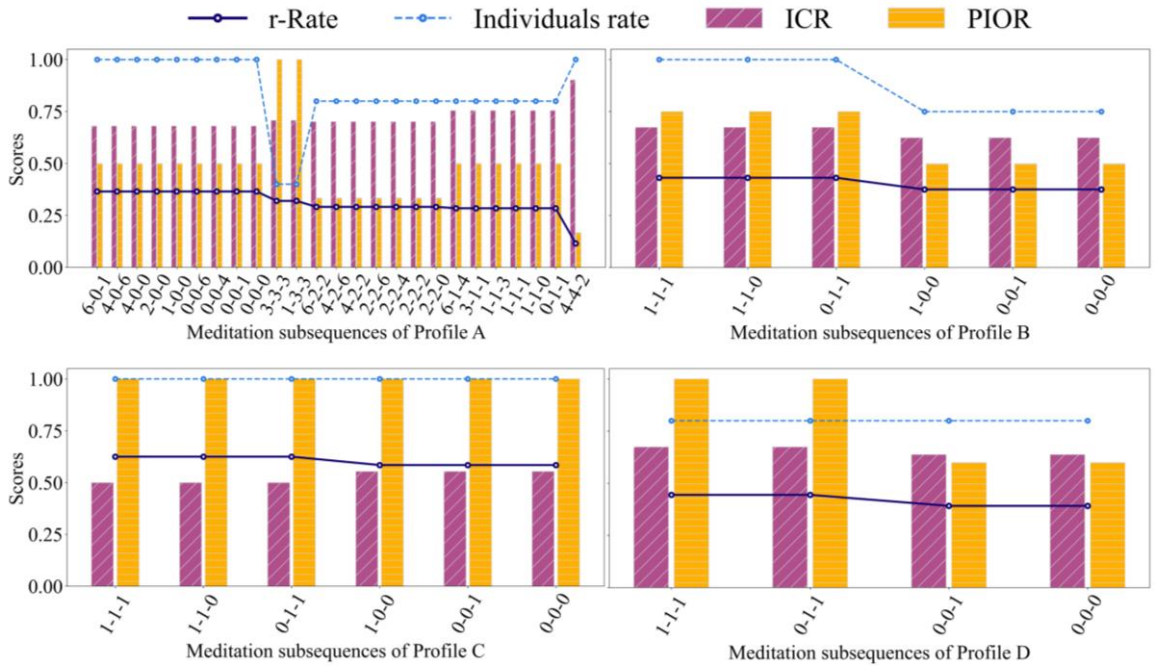


Figure 5.23. Meditation subsequences across the distinct user profiles.

The assessment of the stress subsequences concerning the generalization ability was conducted by evaluating the cumulative users set who undergo each pattern. Figure 5.24

indicates how the transition across the stress subsequences sorted by r-Rate increases the cumulative sum of represented users rate in each profile.

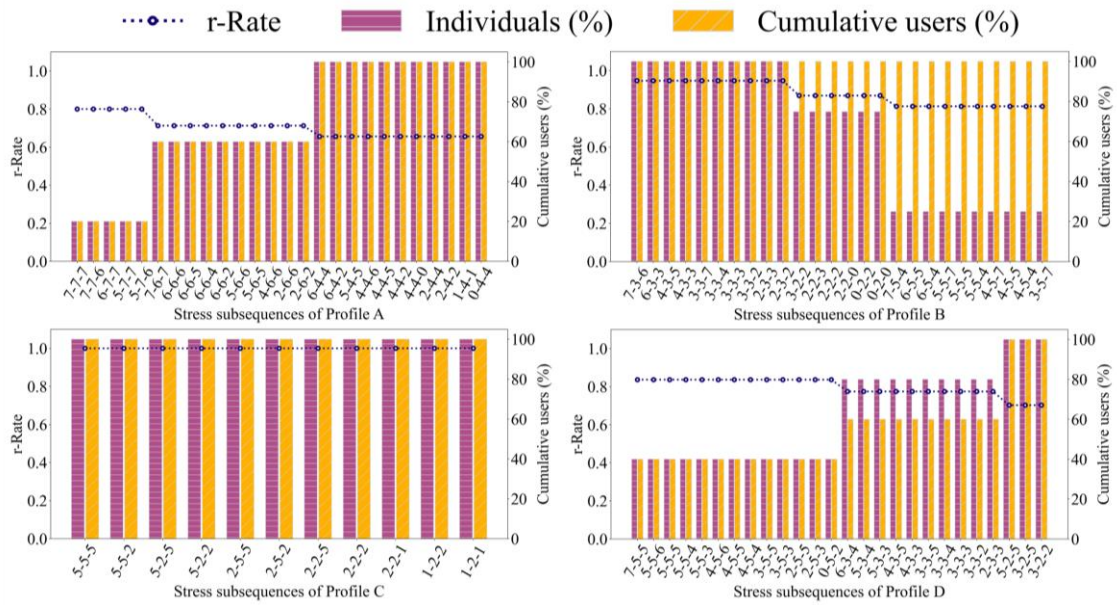


Figure 5.24. Progression of the set of cumulative users across highest rated stress subsequences.

Profile A shows an equal increase between the individuals and cumulative user rates. The first and most rated subsequences in Profile B already represent all users. Profile C only contains one user; therefore, cumulative is maximum across all subsequences. Profile D reflects an increasing progression of the cumulative sum of users, achieving the 100% of users near 0.7 r-Rate.

Adjacency matrices were computed to determine how transitions across the EDA state centroids are arranged. The relation between partitions is determined with the ratio of occurrences wherein pairs occur consecutively, thus constructing transition probabilities matrices. Figures 5.25, 5.26, 5.27 and, 5.28 depict the directed weighted graphs constructed with subsequences of length three whose either r-rate was higher than 0.5 or with 15 highest r-rate subsequences for cases in which r-rate did not reach that score. In this case, the prefix of states was not considered for graphs formation. States of the graph represent the EDA states obtained from the characterization model of the EDA signal where a blank node represents the starting node and node area, and the color tonality represents the in-degree of the node. Transition weights are represented with the color and thickness of edges, indicating for each node which is the probability of transiting to its related nodes.

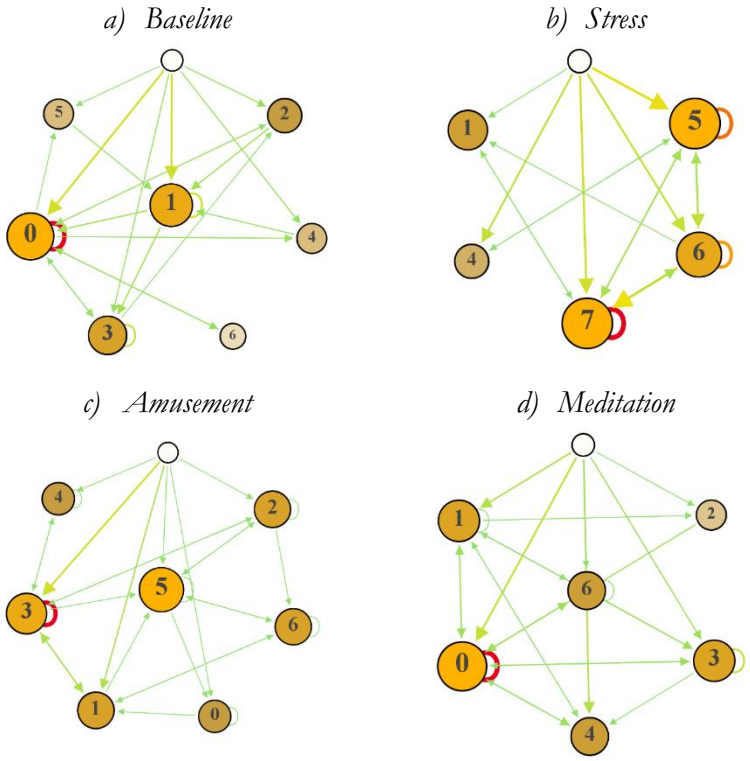


Figure 5.25. Weighted directed graphs of the 15 highest r-Rate Profile A subsequences.

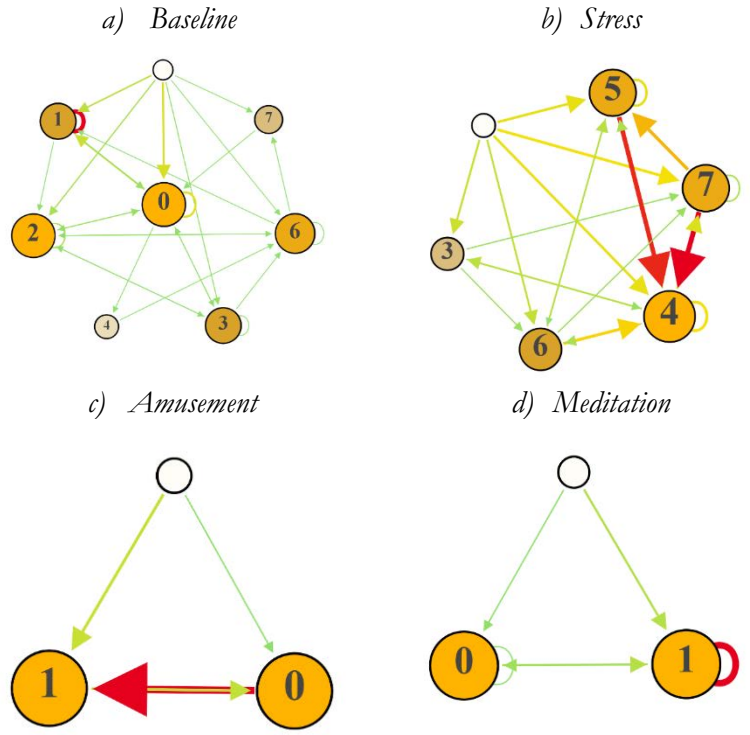


Figure 5.26. Weighted directed graphs of the 15 highest r-Rate Profile B subsequences.

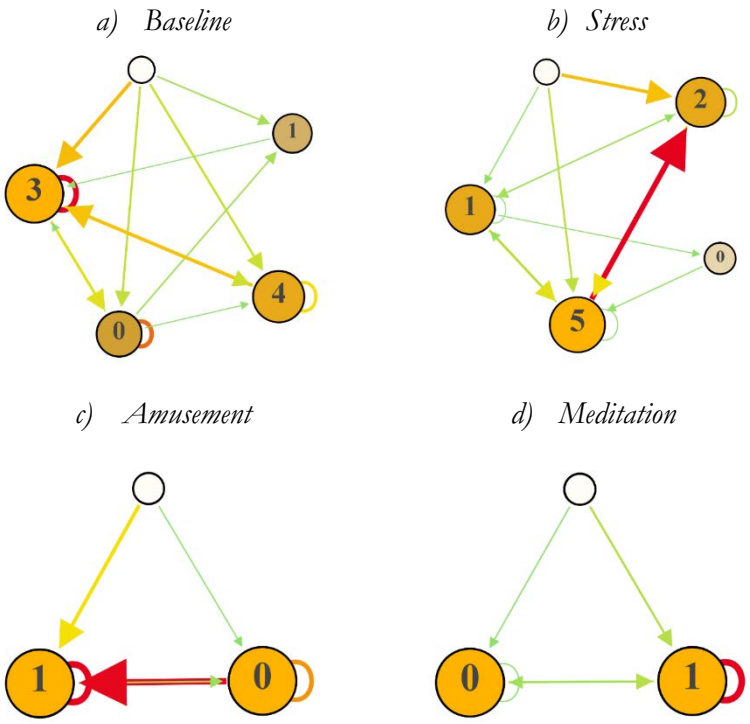


Figure 5.27. Weighted directed graphs of the 15 highest r-Rate Profile C subsequences

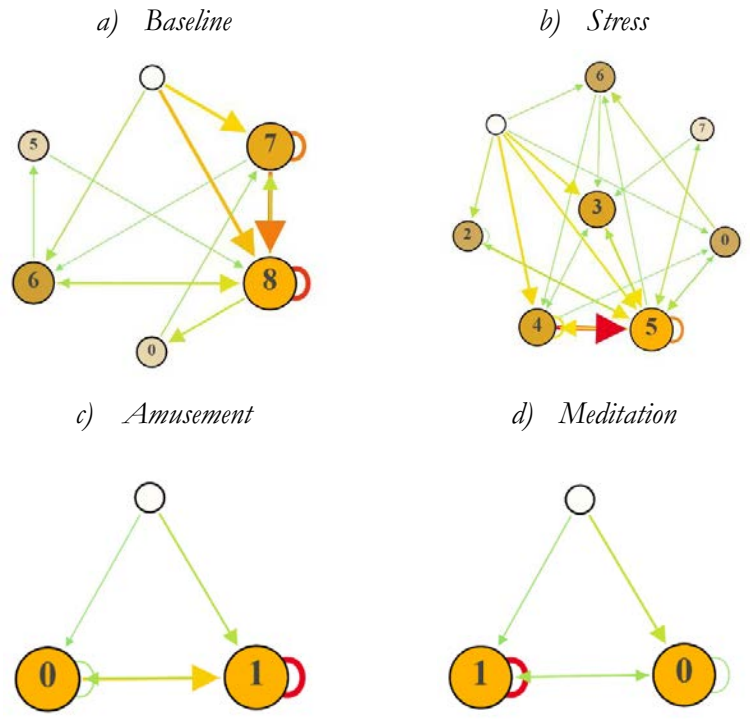


Figure 5.28. Weighted directed graphs of the 15 highest r-Rate Profile D subsequences

5.4.6. Take-home messages

The analysis of symbolic sequences provides the conclusions as follows:

- 3-symbol length is the minimum information unit that allows distinguishing unique or very representative patterns across the distinct user profiles.
- Meditation graphs have demonstrated sharing similar behaviors across the different user profiles.
- The resulting adjacency matrices can be used to construct Markov chain models that provide the probability of staying in distinct cognitive states.

5.5. Model validation

Although supervised learning is not covered in this chapter initially, the aggregated and extracted information from EDA signal through the applied unsupervised learning techniques raises the question of how the proposed framework has diminished the signal entropy for recognizing stress patterns classification tasks. It is worth highlighting that using hyperparameter optimization methods such as Grid Search, Random search, and Bayesian Optimization, among others, which tune the models to solve optimally stress classification is not the aim of this section.

In this section, how the dataset is structured for its use for classification tasks takes more relevance. As explained in the introduction of this chapter, most of the approaches focus on extracting features in all possible domains to maximize the amount of information that describes. This study surveys the impact of signal segmentation, transformations, and variance hierarchization level, including user similarities and signal states in the stress recognition task.

Accuracy metrics should be obtained to validate the proposed framework with the published literature. Accordingly, two perspectives were adopted to generate the dataset for training some supervised-based learning models. First, the formation of instances using the transformed data provided by PCA, including as attributes the user profile which user belongs and the partition identifier that signal belongs.

Furthermore, supervised learning was applied taken consecutive instances to create a new dataset based on multivariate time series. The multiclass classification approach was rejected due to the target of this study only concerns stress detection. In this way, a One-vs-All strategy was adopted, thus converting the problem into a Stress-vs-non-Stress classification problem. The classification results were compared with studies that tackled stress classification using the WESAD dataset to validate their proposed approach.

5.5.1. Dataset formalization

The study on symbolic EDA patterns demonstrated that an isolated event could occur at any part of the experiment, but a sequence of more than two symbols could uniquely characterize stress patterns. Obviously, the amount of information that a snapshot taken in an instant of time contains is lesser than a short clip or a GIF can provide, even though it does not have the frequency rate of high-definition videos. In other words, uncertainty is higher when an information source lacks context, regardless of its nature.

The dataset was conformed following this principle exposed above so that instances are composed by a finite succession of information pieces to reduce EDA data's entropy regarding stress. Formally, an instance is a tuple described as:

$$\langle a_{t-n}^{0}, a_{t-n+1}^{0}, \dots, a_{t-1}^{0}, a_{t}^{0}, \dots, a_{t-n}^{p}, a_{t-n+1}^{p}, \dots, a_{t-1}^{p}, a_{t}^{p}, C_t \rangle$$

where a_{t-n}^{p} is the p^{th} considered variable that occurred n lapses of time previous to the target event to be predicted (C_t).

According to this description, the dataset can be formatted in four distinct ways: only with PCA projection and combining PCA projections with either the state of the signal, the user profile, or both.

5.5.2. Experimental protocol

Dataset details

For this experiment, three consecutive instants of time were taken for composing the instances. Considering that initially dataset was segmented with two levels in which fine-grained window had 10 seconds of length and 2 seconds of offset, formed instances differ by 6 seconds. Further, under the perspective of deploying the model in a real-world environment, the model functioning only would require an initialization period of 14 seconds to provide predictions every 2 seconds.

Concerning the data preprocessing, categorical data provided both by the user profile model (A, B, C, and D) and by the EDA characterization model (values within 0 and 8) was binarized, converting each category in a new binary attribute.

With respect to label treatment, the multi-class classification problem was converted into a binary classification problem. In this manner, the Stress-vs-non-stress problem was addressed due to the main objective of this study relies on differentiating stress patterns from other elicited emotions. As a result of One-vs-All processing, the proportion of classes is imbalanced (stress: 22.17%; non-stress: 77.83%); however, no method was applied for treating this issue. Furthermore, it is relevant to mention that instances labeled as "*Transient*" were excluded from the dataset in the same way that found literature was done.

As with it was specified above, the four datasets addressed to conduct the validation process considering three units of information are described as follows:

- ***Only PCA***: A sequence of the PCA projections of EDA time-spectrum space forms this dataset. Accordingly, 9 numerical attributes correspond with sequencing the three components of PCA transformation.
- ***PCA + UP***: In this dataset, categorical information concerning user profiles is aggregated to PCA attributes. In this way, this dataset comprises 13 attributes (9 attributes for PCA transformations and 4 attributes corresponding to the binarization of the possible values of the user profile).
- ***PCA + EDA-States****:* This dataset introduces a sequence of the categorical information about the states of EDA determined by the clustering methods, getting as results 36 attributes (9 attributes for PCA transformations and 27 corresponding with the binarization of the possible EDA states).
- ***PCA+UP+EDA-States:*** All information explained above was gathered for this dataset, thus obtaining 40 attributes.

Regarding dataset size, 98908 (stress: 21927; non-stress: 71981) instances formed the datasets used to conduct this experiment.

Experiment description

According to the dataset described above, experiments were deployed from two perspectives: general models that consider all subjects for fitting the models and semi-personalized models in which only users of the same profile were used for fitting the models. Individual models are discarded; therefore, Profile C, which only contains one subject, was included in the semi-personalized experiment.

Scikit-Learn [193] leads all experiment processes, including the deployment of models and model selection by assessing model scoring obtained by cross-validation.

The strategy applied for fitting the model was Leave-one-person-out cross-validation (LOOCV), in which the one individual's data is used as a test set, thus allowing to compare results between general and semi-personalized models.

Stratified K-Fold was the applied method for the semi-personalized approach to conducting the cross-validation. This approach was considered to preserve the class distribution across the training and test dataset, thus solving the issue concerning the reduced number of instances when user profiles are taken. Table specifics the number of instances and how classes preserve the distribution of the entire instances set.

TABLE 5.9. INSTANCES USED FOR VALIDATING THE SEMI-PERSONALIZED STRESS-VS-NON-STRESS MODELS.

Profile	Stress	Non-stress	Total
A	7214 (21.8%)	25849 (78.2%)	33063
B	5643 (21.6%)	20435 (78.4%)	26078
D	7585 (22.9%)	25536 (77.1%)	33121

Only were considered the not too many complex models to deploy. In this way, models such as neural networks were not considered to conduct this experiment. Among the deployed models, different families of machine learning models were studied, including statistical models (Linear discriminant analysis (LDA), Logistic Regression), kernel-based classifiers (Linear Support Vector Classifier (SVC), Stochastic Gradient Classifier (SGD)), and ensemble models (Random Forest, Gradient Boosting).

Several scoring metrics were used to evaluate fitted models' performance. Due to dataset unbalancing of classes, accuracy, precision, recall, F1-Score were computed to evaluate the models trained by cross-validation. Even the area under the curve of receiver operator curve (ROC AUC) is widely used to assess models' performance; it can become too optimistic in imbalanced datasets. Therefore, ROC AUC did not take too much relevance in the model selection analysis.

5.5.3. Results

Regarding the outcomes of general models (see Annex III), Figure 5.29 shows the performance results about how accurate the models were to predict stress and non-stress instances correctly. Considering that the default classifier, which employs the majority class

in the dataset to classify all instances, has around 77% accuracy, most validated models improved this classification results. The ensemble-based model achieved the best classification results, exceeding the 95% of classification accuracy.

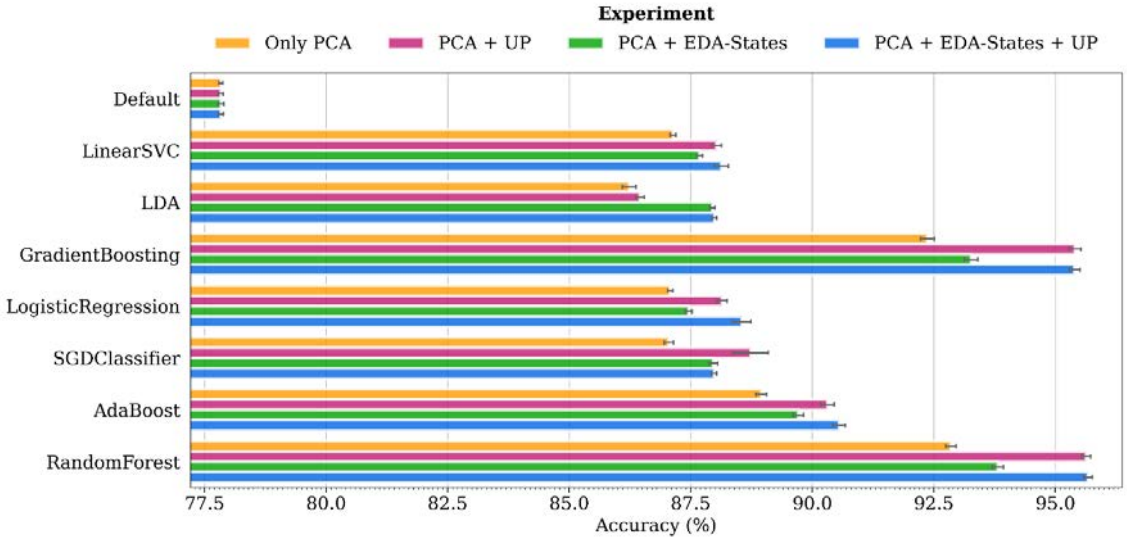


Figure 5.29. Accuracy score of general Stress-vs-non-Stress models across the deployed experiments regarding the dataset composition.

As can be observed, accuracy outcomes across experiments had noticeable differences. All experiments in which dataset contains aggregated information had very significant accuracy improvements respecting the experiment in which models were trained with the "*Only PCA*" dataset. Deployed models had diverse behavior concerning the aggregation of categorical information; however, most models were more accurate in the experiment that integrates all information ("*PCA + EDA-States + UP*"). Particularly, ensemble models (Gradient Boosting and Random Forest) did present very low significant accuracy differences between "*PCA+UP*" and "*PCA + EDA-States + UP*" experiments.

Table 5.10 indicates more detailed results of the four most accurate models, including Precision and Recall scores that are much more suitable for evaluating models trained with imbalanced data. Comparing the highest accuracy models, Random Forest models obtained in the "*PCA + UP*" experiment had significantly higher recall, thus indicating better classifying non-stress instances. In contrast, Random Forest models obtained in the "*PCA + EDA-States + UP*" experiment better predicted stress instances.

TABLE 5.10. SCORE METRICS OF THE 4-TOP GENERAL STRESS-VS-NON-STRESS MODELS OBTAINED WITH LEAVE-ONE-PERSON-OUT CROSS-VALIDATION (LOOCV)

Experiment	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (x100)
PCA + EDA-States + UP	Gradient Boosting	95.39±0.11	89.62±0.48	89.61±0.60	89.61±0.25	98.54±0.12
PCA + UP	Gradient Boosting	95.40±0.12	89.57±0.45	89.69±0.55	89.63±0.29	98.54±0.11
PCA + UP	Random Forest	95.63±0.09	89.17±0.41	91.39±0.49	90.27±0.21	99.07±0.03

Experiment	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (x100)
PCA + EDA-States + UP	Random Forest	95.67±0.09	89.91±0.40	90.64±0.70	90.27±0.24	99.09±0.03

To evaluate the classification performance, the spent time of the validation stage during cross-validation was measured for each model throughout the distinct experiments. Figure 5.30 shows the differences in prediction time per instance across the evaluated models. Random Forest and Gradient Boosting obtained performance time likely acceptable of accomplishing real-time constraints. Notice that AdaBoost models obtained the worst time performance, and the deployment of this model could be questionable, notwithstanding its mean prediction time per instance remains short.

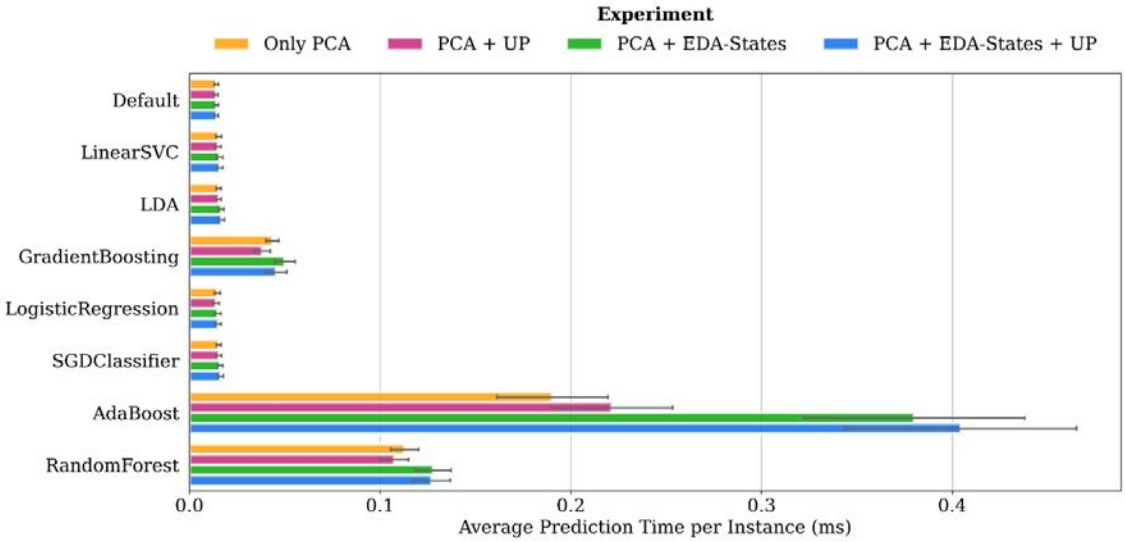


Figure 5.30. Prediction time performance of general Stress-vs-non-Stress models during cross-validation.

Concerning semi-personalized models (see Annex III), Figure 5.31 shows how overall classification results were significantly better than general models.

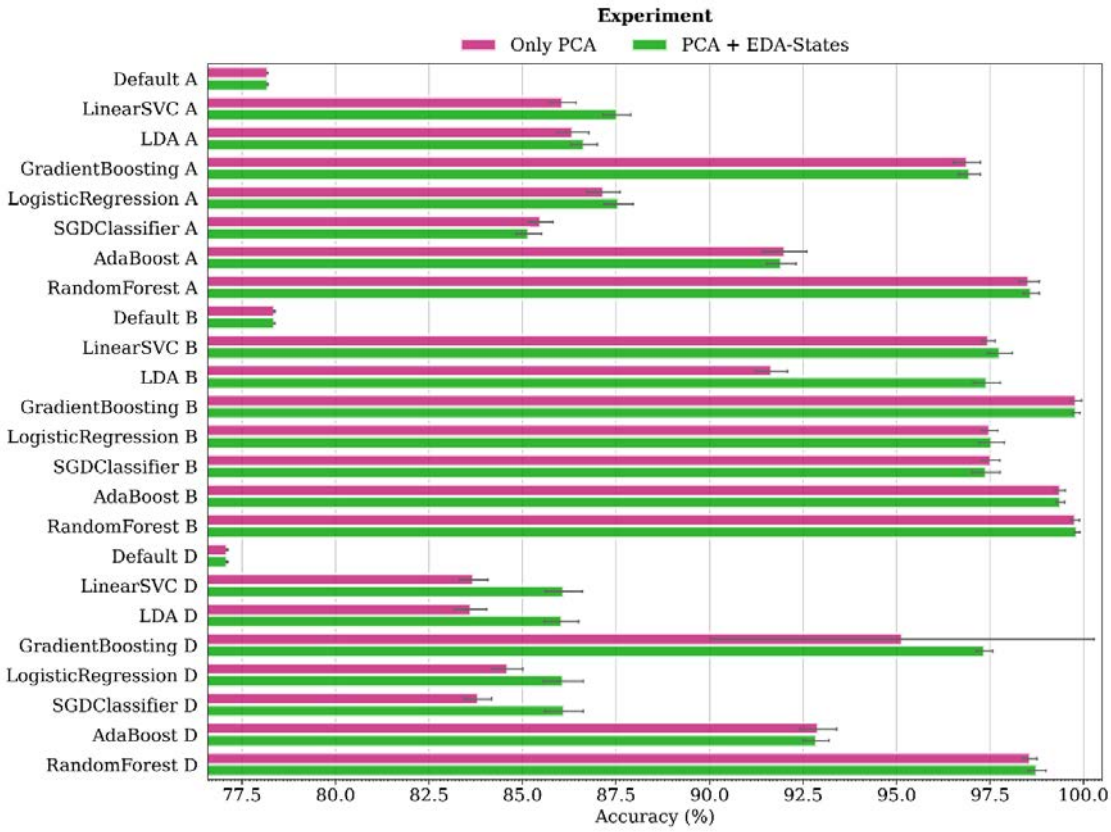


Figure 5.31. Accuracy score of semi-personalized Stress-vs-non-Stress models across the deployed experiments regarding the dataset composition.

Random Forest and Gradient Boosting were the best accurate models again, achieving accuracy rates close to 99.9% in some cases. Except for models from Profile B, in which all models improved their accuracy very significantly respecting the general model on the distinct experiments, the accuracy of the rest of the models remains stable, and they did not exhibit significant differences.

Table 5.11 shows the validation metrics used in the model selection process. Regarding sensibility, "*PCA + EDA-States*" models presented a lower rate of false positives (Recall) than "*Only PCA*" models, i.e., a lesser number of non-stress instances were misclassified as stress, thereby showing a higher sensibility. Concerning specificity, "*PCA + EDA-States*" models exhibited a higher rate of false positives (Precision) than "*Only PCA*", thus manifesting a higher specificity as well. Particularly, the model of profile B did not exhibit notable differences in terms of sensibility and specificity.

TABLE 5.11. SCORE METRICS OF THE 2-TOP SEMI-PERSONALIZED STRESS-VS-NON-STRESS MODELS ACROSS USER PROFILES OBTAINED WITH STRATIFIED K-FOLDS CROSS-VALIDATION

	Experiment	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (x 100)
A	Only PCA	Gradient Boosting	96.88±0.38	93.09±0.64	92.58±1.5	92.83±0.92	99.04±0.23
		Random Forest	98.54±0.28	96.62±0.68	96.69±0.8	96.65±0.65	99.84±0.06

	Experiment	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (x 100)
	PCA + EDA-States	Gradient Boosting	96.95±0.31	93.56±0.81	92.38±0.91	92.96±0.72	98.97±0.33
		Random Forest	98.6±0.22	96.58±0.51	97.02±0.67	96.8±0.51	99.85±0.06
B	Only PCA	Gradient Boosting	99.8±0.15	99.3±0.47	99.77±0.28	99.53±0.36	99.87±0.15
		Random Forest	99.78±0.13	99.09±0.57	99.89±0.17	99.49±0.29	100.0±0.0
	PCA + EDA-States	Gradient Boosting	99.8±0.11	99.33±0.48	99.73±0.25	99.53±0.25	99.95±0.06
		Random Forest	99.83±0.08	99.4±0.26	99.8±0.34	99.6±0.17	100.0±0.0
D	Only PCA	Gradient Boosting	95.15±5.4	88.38±12.48	94.03±1.05	90.66±8.15	96.55±6.72
		Random Forest	98.57±0.19	95.57±0.81	98.33±0.32	96.93±0.41	99.88±0.03
	PCA + EDA-States	Gradient Boosting	97.35±0.23	93.69±0.75	94.82±0.56	94.25±0.5	98.96±0.15
		Random Forest	98.75±0.25	96.23±0.93	98.42±0.45	97.31±0.54	99.89±0.04

Table 5.12 resumes the overall results of the best semi-personalized models. Analyzing the dispersion throughout all metrics, the Random Forest model evidenced higher stability than Gradient Boosting, and specifically "*PCA + EDA-States*" model exhibited the highest accuracy, sensibility, and specificity.

TABLE 5.12. OVERALL RESULTING OF TOP-2 SEMI-PERSONALIZED STRESS-VS-NON-STRESS MODELS MODEL SELECTION

Experiment	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (x100)
Only PCA	Gradient Boosting	97.28±3.59	93.59±8.32	95.46±3.32	94.34±5.97	98.49±4.01
	Random Forest	98.96±0.62	97.09±1.64	98.3±1.42	97.69±1.38	99.91±0.08
PCA + EDA-States	Gradient Boosting	98.03±1.3	95.53±2.82	95.64±3.17	95.58±2.94	99.3±0.51
	Random Forest	**99.06±0.59**	**97.4±1.57**	**98.41±1.25**	**97.9±1.31**	**99.91±0.07**

Prediction time of semi-personalized models was assessed as well. Figure 5.32 illustrates the average prediction time of all models obtained during cross-validation. The outcomes of personalized ensemble-based models manifested a speed up close to two orders of magnitude respecting general ensemble-based models. This performance improvement confirms that smaller datasets for fitting generate models with fewer tree estimators, and consequently, it obtains lighter models.

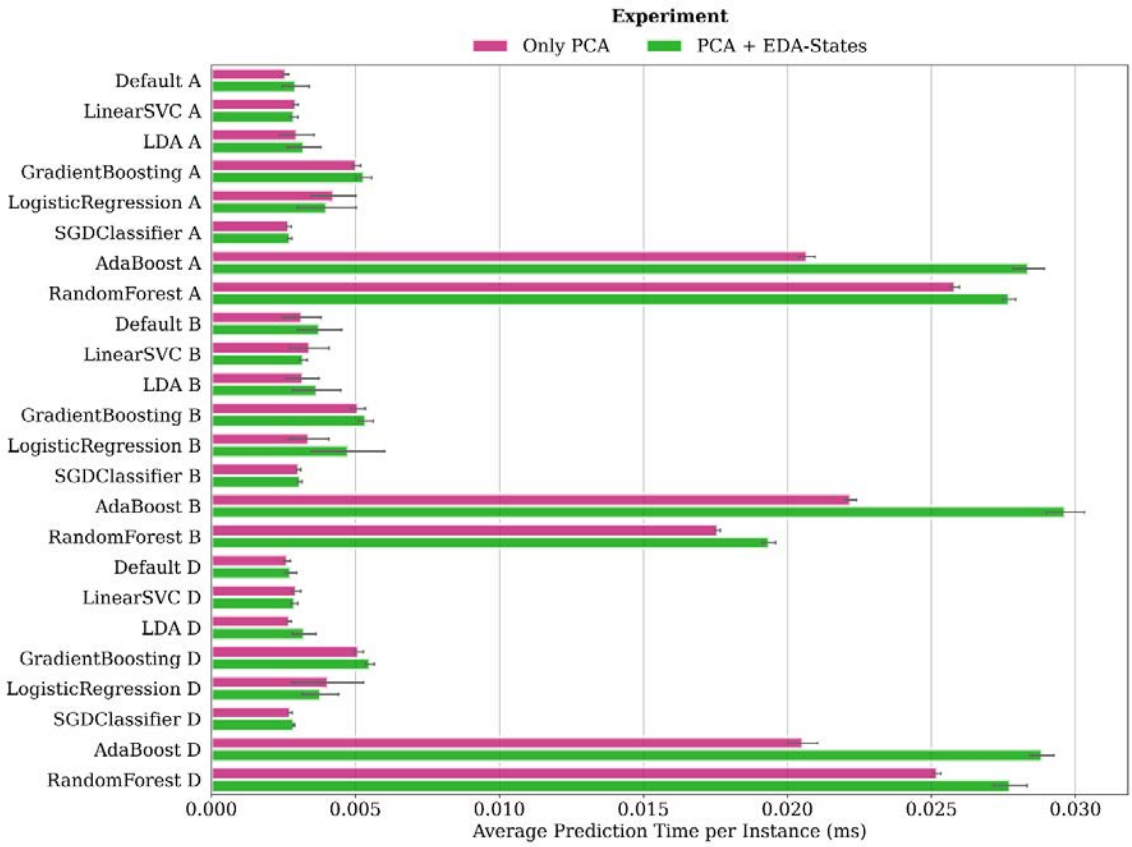


Figure 5.32. Prediction time performance of semi-personalized Stress-vs-non-Stress models during cross-validation.

The best outcomes of the proposed approach were compared with the reviewed literature in which the WESAD dataset was used to validate their prosed methodologies for stress detection. In this comparative analysis, only binary classification results were considered to fairly evaluate and compare the outcomes. Table 5.13 resumes the results of several studies in terms of score metrics, including accuracy and F1-score, as well as the used sensors to achieve the reported results.

As it can be observed, both proposed perspectives (general and semi-personalized) accomplished better results than all the reviewed studies, even when they used multiple sensors to classify stress. These studies were addressed from the perspective of feature extraction through signal analysis in diverse domains, including time, frequency, and even time-frequency space. However, even when algorithm outcomes should be worse theoretically, e.g., Random Forest vs. Extra trees and Random Forest vs. XGBoost, the proposed framework achieved the best accuracy and F1-Score.

TABLE 5.13. RESULTS COMPARISON WITH FRAMEWORKS THAT USED THE WESAD DATASET.

Study	Model	Sensor	Accuracy	F1-Score
Proposed General approach	Random Forest	Wrist EDA	**95.67±0.09**	**92.51±0.26**
Proposed Semi-personalized Approach	Random Forest	Wrist EDA	**99.06±0.59**	**97.9±1.31**

Study	Model	Sensor	Accuracy	F1-Score
Schmidt et al. [164]	LDA	Chest (ECG, EDA, EMG, RESP, TEMP)	93.12	91.47
Hsieh et al. [186]	Dominant Features + XGBoost	Chest EDA	N/A	92.38
		Wrist EDA	N/A	89.92
Marugappan et al. [201]	ExtraTrees	All Chest + All Wrist Sensors	93.	93.
Aqajari et al. [202]	kNN	Wrist EDA	91.6	N/A
Föll et al. [203]	ADABoost	Wrist EDA	N/A	51.96±0.32
Tervonen et al. [140] (Fully personal model)	SOM	Wrist EDA	92 ± 9	89±13

5.5.4. Take-home messages

- Although supervised learning was not the target of this chapter, stress detection was addressed in a hybrid way wherein supervised algorithms were guided for all information recruited through unsupervised analysis.
- Unsupervised learning leads to both the data enrichment and the design decision for structuring the data model employed for feeding the supervised algorithms.
- Aggregated data resulting from unsupervised learning have demonstrated that their inclusion significantly improves the classification results.
- Ensemble models have shown that decision trees are the most appropriate models to tackle stress detection addressed as a binary classification problem.
- Semi-personalized models have obtained very high accuracy outcomes to classify instances extracted from the stress phase using only the EDA sensor, substantially improving the reported results of the reviewed stress detection studies that validated their proposed frameworks with the WESAD dataset.

5.6. Summary

This chapter introduced the stress detection problem, presenting the underlying issues that become the real-time detection of stress patterns in a complex problem. Inter-individual and within-person variations, and the lack of datasets exhaustively labeled, complicate the achievement of a generalized solution; thereby, stress detection in real-time is still considered an open issue.

Following the "*Divide and Conquer*" methodology, a framework based on unsupervised learning approaches was proposed for tackling these issues individually. In this way, a pipeline of techniques including user segmentation, fine-grain signal characterization, and sequence analysis was applied; thereby, EDA signal was subjected to various transformations, having as the main purpose of reducing the entropy of determining stress condition in a group of subjects. For that, the WESAD dataset was used because the authors

applied a widely accepted stress elicitation methodology and other studies used it to validate their approaches.

The experimentation conducted with supervised algorithms achieved 95.67% and 99.06% accuracy, substantially improving the best Stress-vs-non-Stress classification results reported in the reviewed literature that employed the WESAD dataset to validate their approaches. Considering that the examined supervised models were neither tunned for improving their accuracy performance nor overtrained to be specialized in recognizing the instances of the employed input dataset, the high accuracy achieved outcomes evidenced the relevance of framework stages rather than selecting a specific supervised machine learning model. This fact leads to discuss the distinct findings of this study.

In general Stress-vs-non-Stress models, the influence of user profile information was notable due to this aggregated information improved the classification accuracy up to around 2.5 and 3% with respect to experiment wherein this categorical information was not included. The semi-personalized stress-vs-non-Stress classification models overcame the accuracy results of the generalist approach. This fact reveals that the user profile model subdivided the set of subjects very truthfully according to the subject's EDA histograms. Comparing with state-of-art related results, Tervonen et al. [140] distinguished two groups of subjects applying k-means on the WESAD dataset. Nevertheless, this paper did not expose the selection criterion to determine which value of k optimizes inter-cluster and intra-cluster distance.

Another critical factor was related to the effort to avoid losing the data sequentiality and the selection of window length and the window offset considered to analyze the EDA signal. Unlike most reviewed studies that chose 0.25 seconds of offset, being this the minimum possible value, an offset of two seconds was taken. Despite real-time constraints, this time separation between instances can reflect more significant gradient changes on the EDA signal, thus being probably more in concordance with the time delays in physiological systems.

Applying Wavelet transforms and PCA was fundamental for characterizing the EDA signal. The obtained accuracy combing both techniques was similar and even higher than studies that employed multiple features extracted from several sensors and devices. Inspecting mid and low spectra with Mexican hat wavelets, an orthogonal basis of three components explained the 100% of the variance of thousands of fine-grain EDA waveforms. Moreover, with the signal analysis in the time-frequency space, the signal scaling problem across subjects was eluded entirely.

Determining the EDA states by partition clustering of the PCA projections of the EDA spectrum contributed to the achievement of the presented results.

The analytical study about symbolic EDA representation was crucial in selecting the correct number of consecutive windows to design the data structure used to feed the models. The statistical analysis based on sequence occurrence demonstrated that three symbols were the minimum length in which patterns either exclusive or highly related with stress phase appears.

6. Designing and assessment of multi-agent architectures for data Fusion

Data fusion in active driving safety design is crucial for achieving a functional human-centered approach that enhances driving safety and ergonomics. In this way, the driver's characteristics, needs, capabilities, and limitations govern all aspects of vehicle and environmental design considerations [204]. That applies to the elements that likely improve the support systems' functioning and, consequently, the driving experience. The first issue concerns the driver's cognitive load mitigation produced by driving support features, especially when the car equips several. The second issue concerns the scope of Automated Driving Systems (ADS). In this case, the driver must supervise driving automation system engagement to intervene in case of system failure.

Consequently, the monitor must assess several aspects of driver behavior according to their chosen role during the dynamic driving task (DDT). Stanton et al. stated that six relevant factors compose the psychological model which defines the driver's behavior concerning automated systems operation [20]. Thus, the quality of the interaction process initiated through feedback depends on and affects the state of the locus of control, trust, stress, situation awareness, mental model, mental workload, and task demands.

Currently, the driver modeling analyzes physical and physiological aspects to determine signs of deprecated states of the psychological factors cited above. In cases such as models that assess the driver's situation awareness, the complexity of the applied technique is transferred to the deployment in the experimental platform because, in general, complex models use information from heterogeneous data sources concerning both the driver and the environment.

Most of the time, the models under development are composed of a mixture of both owner- and third-party systems developed on different platforms and programming languages. Therefore, orchestrating the information flows that feed the subsystems implies an elevated cost. This chapter presents the insights and technical concepts for deploying models based on multi-agent architectures that require data fusion techniques. The main advantage of the proposed deployment is the flexibility because it allows the deployment of a specific test to be completely decoupled from the test platform, being extremely useful when several experimental processes share the same driving simulator.

The main idea of this chapter is the proposal of a highly scalable integration model whose objective is to achieve an experimental platform able to host the functionality of multiple models, running without mutual exclusion and maintaining a low dependencies level between them.

6.1. Conceptual model

Figure 6.1 depicts the hierarchical structure which supports the architecture design. The shown architecture consists of four stacked tiers corresponding to the agents involved in this domain. Moreover, establishing different abstraction levels allows defining the bottom-up information flow between the agent levels. Additionally, it can observe the interaction between the three entities (car, driver, and environment) and the system that implements the described architecture.

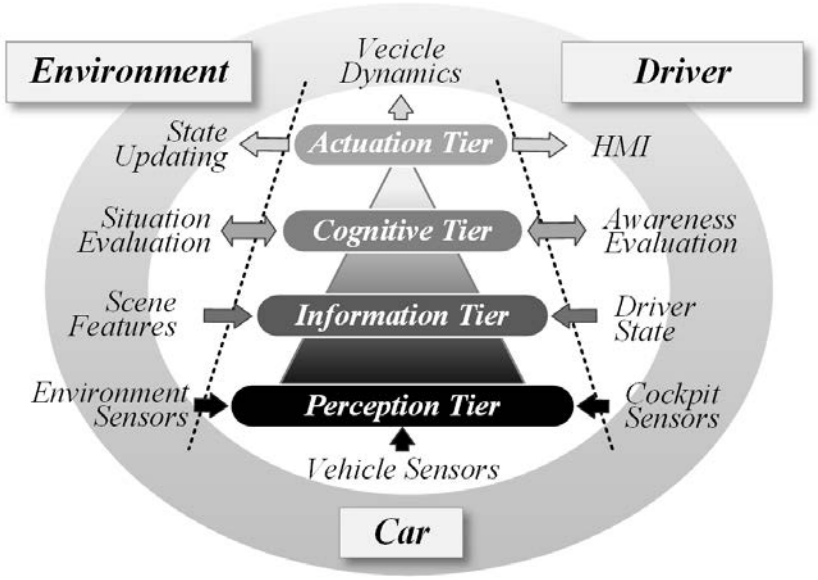


Figure 6.1. Conceptual model of the interactions between entities in a driving scene.

The four tiers of the architecture are described as follows:

a) ***Perception Tier***. It holds the in-vehicle technology in charge of gathering raw data concerning all the factors involved in driving tasks. Both vehicle sensors and any kind of vehicle communication technologies (Vehicle-to-Vehicle, Infrastructure-to-Vehicle, Pedestrian-to-Vehicle) can provide this information.

b) **Information Tier**. This tier provides the necessary information to define the driving scene. It covers reactive warning systems, including current safety offerings and those in development phases. These systems extract high-level information from the Perception Tier elements, both individual and by combining multiple signals. As can be observed in Figure 6.1, it can distinguish three system groups:

- Systems oriented to the environment feature extraction such as Frontal Collision Warning, Rear Collision Warning, Traffic Sign Recognition, Pedestrian Detection, Blind Spot Detection.
- Systems that assist in fulfilling the traffic regulations as speed limit detection, traffic light detection, and lane-keeping assistance.
- Systems intended to determine the driver's condition aspects like physical-based (Gaze Estimation, 3D Head Pose Estimation, Maneuver Detection) and psychic-based analysis (Fatigue Detection, Aggressiveness).

c) **Cognitive Tier.** In this tier, information fusion takes place to formalize the knowledge derived from the heterogeneous outputs provided by the information tier's agents. Information maps about driving scenes can be created and analyzed using a knowledge base derived from this data formalization. Accordingly, systems of this tier will be able to determine whether a given driving scenario belongs to the catalog of use cases. The use case identification also leads to the driver situation awareness assessment, establishing a direct relationship between the driving scene and the driver's faculty to deal with it.

Such as was defined by Ensley [29], the stimulus perception, its comprehension, the projection of its future state, the decision-making process, and finally, the performance of an action compose the driver situation awareness (SA) stages.

The information tier in charge of determining the driver state allows the estimation of each SA stage. The main aim of this tier is to establish a relationship between the driver, profiled by driver monitoring systems, and the warning in-vehicle systems that support their decision-making.

d) **Actuation Tier**. This tier represents the interaction between the system and the entities involved and where entity actuation happens, affecting the behavior and state of other entities. In the driver's context, an interaction interface reports a warning delivered by Cognitive Tier, starting their decision-making process. The actions performed by the driver over the driving controls, or even control signals sent from the Cognitive Layer, modify the vehicle dynamics parameters. Consequently, the result of the actuation tier performance updates the environment state.

6.2. Model of Architecture

This section describes the technical details of the proposed integration architecture for the deployment and testing of data-fusion models in simulation environments. The proposed architecture encloses the main concepts: firstly, the multi-agent paradigm conceptualizes the diverse models that analyze involved aspects in a dynamic driving task. An agent is a distributed system node that executes a machine-learning process to fulfill a stage of a more complex task. Second, a mediation engine (*Broker*) holds the interaction between the involved agents and oversees managing and routing the information flows generated and consumed by all system agents. Well-known architecture such as ROS [205] implements a mediation architecture as a message exchange model across the platform agents. However, API programming language was the operational environment constraint that precluded using ROS standards.

Figure 6.2 depicts an overview of the involved agents that compose monitors concerning the driver and environment according to information flow directions, whereas the Broker establishes the interaction pattern with all the agents involved in the driver-centered safety feature.

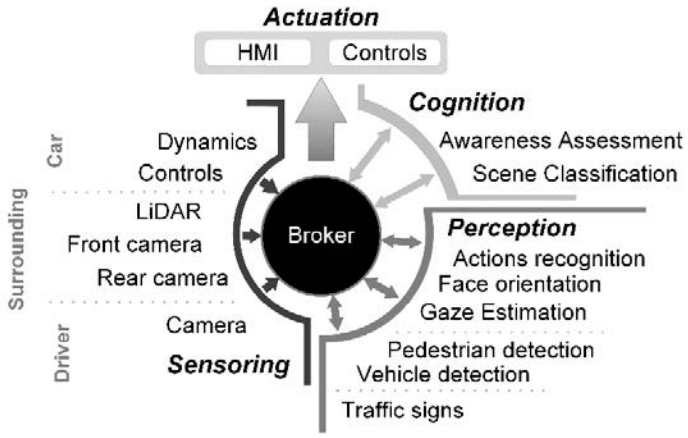


Figure 6.2. Integration model of the distributed multi-agent system based on mediation engine.

The stages which compose autonomous driving states how the agents are classified. The sensor agents gather the raw data from the driver, the car, and the environment. In the case of human-in-the-loop driving simulators, the plugged sensors are a physical and emulated devices mixture. The perception agents enclose the machine learning models that use raw data.

The cognition agents manage knowledge bases that gather the perception information to assess the DDT performance. Finally, the actuation agents oversee reporting safety feedbacks to the driver or actuate the vehicle control in case of DDT fallback.

As Figure 6.2 illustrated above, the *Broker* is a server host that constitutes the central part of the proposed integration model. The *Broker* leads a Publish-Subscribe Messaging System as integration infrastructure. It takes charge of the communication of all agents deployed whose common goal is to provide a safety feature. Therefore, the *Broker* manages the receiving streams from data providers and feeds the subscribers with the information that they require at any moment. Figure 6.3 depicts the publish/subscribe mediation protocol led by the Broker. Moreover, it can observe three agent profiles, producers that only write records to a Broker's specific topic, consumers that only request records from broker topics and, dual agents that both read and write broker records. Producers agents define their topic identifiers, thus grouping their publications under an established name.

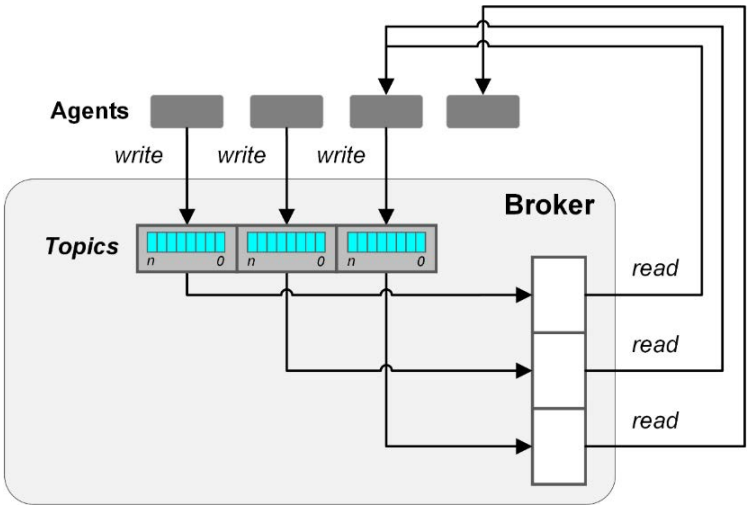


Figure 6.3. Broker architecture integration pattern

The proposed integration model presents several advantages concerning monolithic integration architectures whereby point-to-point connections establish the communication between different subsystems:

- **Flexibility.** The uncoupled model eliminates the code dependencies, permitting the agent deployment regardless of its implementation.
- **Scalability.** The architecture allows new agents to incorporate analysis resources that enrich the driving scene definition and maintain high throughput.
- **Load balancing.** The decentralized and distributed processes along different resources stimulate performance improvement, avoiding system overload.
- **Support for additional tools.** The architecture can also allow collecting data to consolidate data warehouses for training new models and storing logs for result assessment.

Given the above, the integration architecture allows creating pipelines for composing the safety feature stages, creating more complex systems with no dependencies between the deployed systems.

6.3. Knowledge representation

As was explained above, the Broker component has the mission of routing all information along with the infrastructure. However, the correct data formation that both will generate and consumed by the different systems should be known for the Broker. Therefore, the Broker requires a knowledge data model that leads the routing of the information flow appropriately.

Figure 6.4 depicts how the knowledge representation of the driving scene is structured. In this case, the ontology-based model is the paradigm of data representation that leads the semantic information that describes all concerning aspects across all the elements. In this case, the data abstraction representing the driving scene comprises concepts including the car, the driver, and the vehicle context.

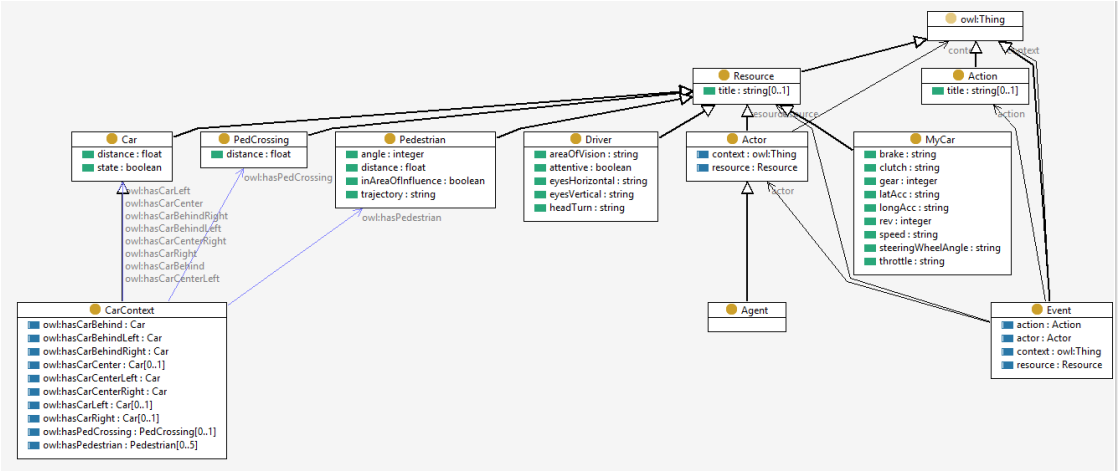


Figure 6.4. Ontology-based knowledge representation model for representing a driving scene.

As Figure 6.4 shows, each entity has features that describe how they behave and intervene in the driving scene. In this integration approach, the driving scene representation was simplified by discretizing ontology attributes. Each ontology instance allows identifying specific driving scenes in which potential hazards can put the safety and integrity of any represented entities at risk. Table 6.1exposes the possible discrete values that can take each entity's properties.

TABLE 6.1. ENTITIES, PROPERTIES, AND POSSIBLE DISCRETE VALUES OF DRIVING SCENE ABSTRACTION

Entity	Property	Possible values
Car	Distance to car	far away, far, normal, close, very close
	State	moving, stopped
Driver	Area of vision	front, front-right, front-left, left, right, behind-left, behind-right, interior rear-view mirror, left rear-view mirror, right rear-view mirror, speedometer, radio / air conditioning / down, roof / up.
	Attentive	yes, no.
	Horizontal direction of the eyes	Left, right
	Vertical direction of the eyes	Up, down
Car Context	Existence of car in front/left/right	yes, no
	Existence of pedestrian/pedestrian crossing:	yes, no
Pedestrian	Angle (relative to driver)	front, front-left, front-right.
	Distance	far away, far, normal, close, very close
	Trajectory	N, S, E, W, NW, NE, SE, SW
Pedestrian crossing	Distance to pedestrian	far away, far, normal, close, very close

6.4. Material and methods

6.4.1. Materials

The STISIM Drive M300WS driving simulator system [183] with VDANL Drive capability was used as the driving testing platform. The computing resources were the desktop computers *Intel Core i7-3770* 16 GB Ram with graphics *NVIDIA GeForce GTX 680* and *Intel Core i7-4790K* 16 GB with graphics *NVIDIA GeForce GTX 960*, both connected to Local Area Network. *Microsoft Kinect v2* sensor was employed as a vision device. Figure 6.5 shows the driving simulation system on execution.

Concerning the software systems, Java Agent Development Environment (JADE) was used for the implementation and deployment of the cognitive layer agents. Moreover, the cognitive layer was deployed in *Apache Tomcat 7* Application Server, thus allowing its connectivity via HTTP protocol. The implementation of agents on driving simulation software was developed using *STISIM Open Module* capability and .NET Framework. The vision software was implemented using Kinect Development Kit and *EmguCV* computer

vision library in C# programming language. The Scenario Definition Language (SDL) provided by STISIM was used for designing the study cases.

Figure 6.5. Driving simulator system

6.4.2. Deployment scheme

The architecture deployment on the driving simulation system has been conducted employing different systems implemented in previous works as part of this Ph.D. thesis. These systems were implemented as agents involved in accomplishing the functionality requirements of the proposed warning system.

Figure 6.6 shows the decentralized deployment scheme based on the proposed integration model.

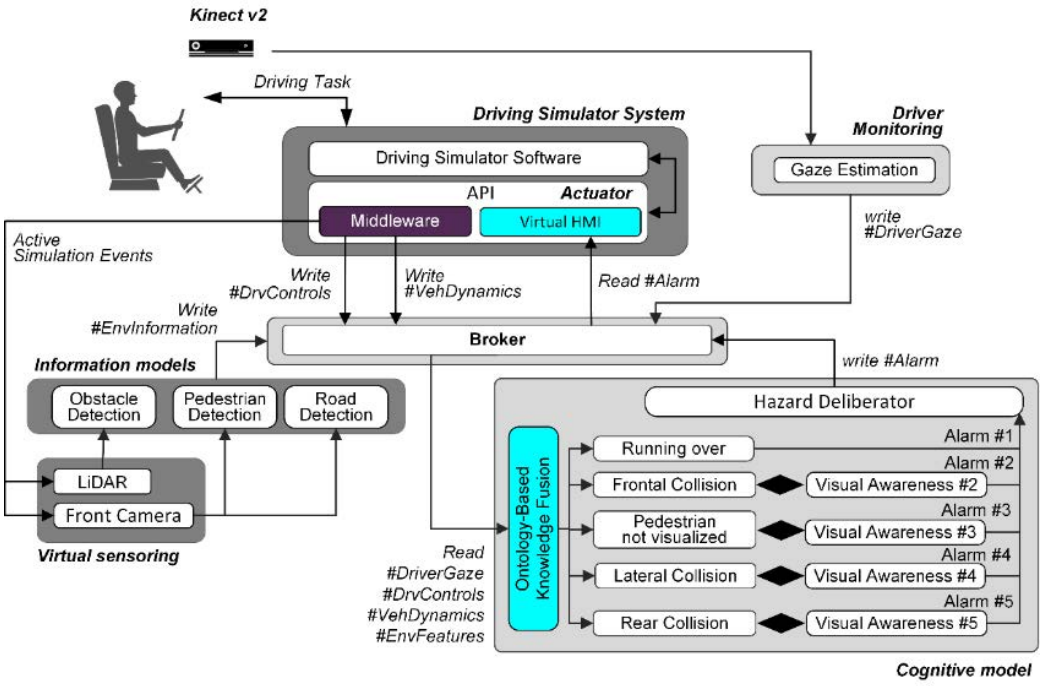


Figure 6.6. Decentralized deployment distributed on simulator system (dark grey) and monitor and reasoning system (light grey).

The support system behind this architecture is integrated by different systems distributed in two computers where each involved system interacts with the Broker to publish or read several topics. The system has as objective to warn the driver in the face of five specific cases of use according to their attention over them. In this case, the analysis of the driver's visual field was employed to assess whether their driver's attention is paired with specific driving situations or not. The integration of each agent inside the architecture deployment is specified as follows.

The Perception and Information agents for the data treatment of the simulated driving environment were implemented. In this work, *LiDAR, Front Camera* was implemented by a class model that extracts the surrounding information according to each technology restriction. As its part, information agents emulate the behavior of algorithms that uses raw data for characterizing specific aspects of the environment (*Obstacle Detection, Pedestrian Detection, and Road Detection*).

Concerning driver monitoring, the computer vision software was developed using a Microsoft Kinect V2 framework. This system acts as an Information Tier agent (*Gaze Estimation*), and it is aimed at obtaining the visual area where the driver's attention is focused. Figure 6.7 shows the results of this system, where the driver's pupil position detection (green circle) and the visual area where the driver's attention is focused (red text) can be observed.

The integrated cognitive model was a rule-based alarm system with a communication model based on web services [206]. In Figure 6.6, the cognitive model is depicted as a substructure of agents where four kinds of agents can be distinguished:

- The *Ontology-Based Knowledge Fusion* is the process led in charge of asking the required information to the Broker and structuring heterogeneous data. This agent provides an ontological representation of all known aspects involved in the performance of the driving task.

- A set of agents that receive the ontology instance aims to identify the specific cases of use that comprise a driving risk situation (*Running over, Frontal Collision, Pedestrian not Visualized, Lateral Collision, Rear Collision*).

- The identification of each use case activates the Visual Awareness agent actuation. This kind of agent relates the use case progress with the driver's visual perception spread by the Gaze Estimation Agent. Each *Visual Awareness agent* is defined as a parametric model that contains information about where the driver's attention must focus on avoiding the incident according to the parameters of the risky driving situation.

Figure 6.7. Eye gaze system results: the pupil position (green circle), calibration point (red circle), and gaze area estimation (red text).

- In the case of the driver's lack of attention, the *Hazard Deliberator agent* receives the signals from the *Visual Awareness agents*. This agent handles a knowledge base about the priority level of each use case integrated into the system.

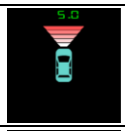
If the driver's inattention occurs on at least two use cases, the *Hazard Deliberator agent* will take the highest priority hazard, writing its corresponding alarm code in the topic named alarm. The weaknesses of actors involved in the driving scene determine this priority level on each alarm. Finally, the actuator is a *virtual HCI* deployed in the visualization system of the driving simulator [206]. This agent integrates different feedbacks, and it reads the topic published by the *Hazard Deliberator* to trigger the appropriate alarm.

The simulation frame rate was set to 60 Hz to facilitate the synchronization between distributed agents and avoid the bottlenecks in the slower systems.

6.4.3. HCI Alarm messages

As Figure 6.6 shows, the hazard deliberator raises the corresponding alarm to the detected hazard. These visual messages are presented when the Hazard deliberator infers a lack of attention against a specific driving situationTable 6.2 overviews the visual messages deployed on the virtual HCI that warn the driver for their attention.

TABLE 6.2 DESCRIPTION OF THE SET OF VISUAL AND SOUND MESSAGES DEPLOYED ON THE VIRTUAL HCI.

Id	Message	Sound	Presentation	Description
#1		Yes	High frequency blinking and warning sound.	It is shown when the risk of running over a pedestrian are critical (i.e., pedestrian is occluded by some environment object such as parked car).
#2		Yes, when reach the closest level	Closeness indicator with distance in meters.	It is shown when the risk of front collision with another car.
#3		No	Mild frequency blinks.	This alarm is shown when some pedestrian is going to cross the street.
#4		No	Animation wherein the environment car (yellow) moves to the right	Signal shown when an environment vehicle performs a lane changing towards the ego-vehicle's current lane.
#5		No	Animation wherein the environment car (yellow) moves to the front	It is shown when an environment vehicle is the blind spot of the ego-vehicle.

The purpose of the designed warnings was to represent each situation with interface metaphors that facilitate their comprehension during the driving. This experiment was not focused on evaluating interface-related aspects. Consequently, the designing of interface messages did not cover all recommendations specified in standards and guidelines such as ISO 11429:1996(en)[207] that promote the establishment of a unified criterion in designing ADAS interface across automotive OEMs.

6.4.4. Driving test design

This subsection will describe the process of performing a driving trial based on case studies as an evaluation method of the proposed architecture. This process is composed of several steps, and it has been made following a continuous improvement methodology. The cycle of life of this methodology is composed of some activities that can be summarized as PDCA (*Plan-Do-Check-Act*).

The *Plan* activity consisted of the specification and the case studies design. As the starting point, it was considered the generic definition of some critical urban traffic situations, such as occluded pedestrians, override risk, or frontal collision risk [126]. However, the case studies specification requires the definition of concrete scenes in which these critical driving situations happen to evaluate the functioning of the warning ADAS system in its operational driving domain.

As a reference for designing the study cases, it was token statistical public domain information relating to accident types and circumstances specifications in urban areas collected by road safety agencies such as The Spanish Directorate-General for Traffic (DGT). According to the last accessible accident records [208], it was produced 52222 traffic crashes in urban areas in 2013.

Figure 6.8 shows the proportion of urban traffic accidents according to their crash circumstances. It did not consider an overturning in the road due to could be a consequence of various collision types. Run over, lateral, frontolateral, and rear collisions are the most common types of accidents, and they can be effortlessly reproducible in a driving simulator. In addition, running-off-road accidents depend on the driver's attention and fit condition.

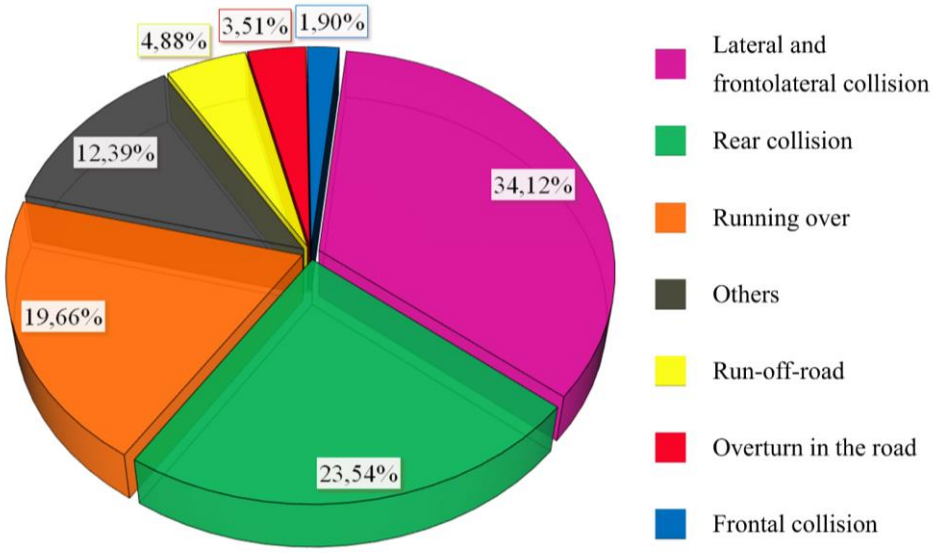
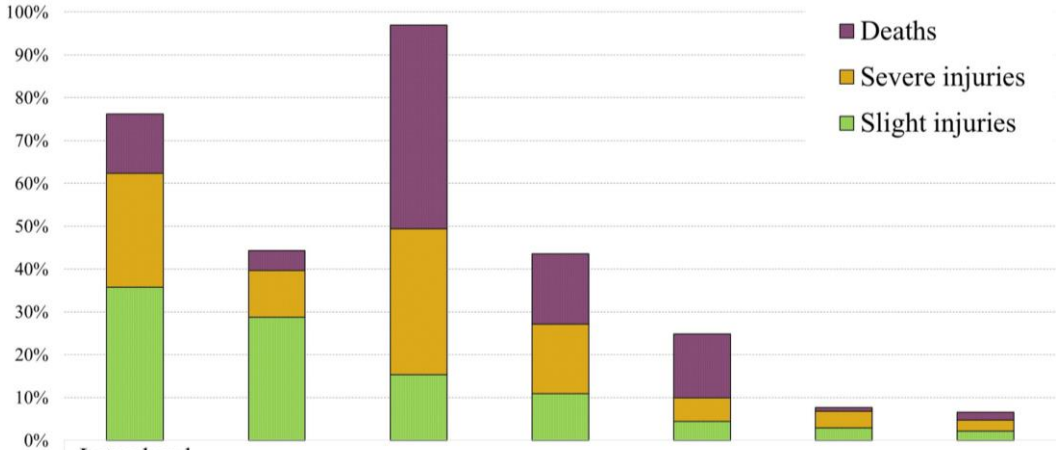


Figure 6.8. DGT statistics about accident types produced in Spanish urban areas in 2013.

Even though the three most relevant accident causes exhibit a similar proportion, damage and mortality rates broadly differ between the distinct accident types; therefore, the scenarios did not have the same repercussion. Figure 6.9 outlines which impact concerns personal harms caused by the accident types previously presented in Figure 6.8.



	Lateral and frontolateral collision	Rear collision	Running over	Others	Run-off-road	Overturn in the road	Frontal collision
Deaths	13,78%	4,67%	47,56%	16,44%	14,89%	0,89%	1,78%
Severe injuries	26,67%	10,89%	34,05%	16,23%	5,61%	3,85%	2,69%
Slight injuries	35,72%	28,72%	15,32%	10,86%	4,37%	2,91%	2,10%

Figure 6.9. Occasioned harms by distinct accident types occurred in Spanish urban areas in the year 2013.

As can be observed, running over is the most frequent mortality cause wherein there is a probability near 50% of provoking mortal victims each time that occurs. Lateral and frontolateral collisions exhibit the second-highest rate of severe injuries just over by running over. Finally, rear collisions capitalize the highest rate of slight injuries proportionally.

This analysis highlights the importance of ensuring the safety of the most vulnerable road user: the pedestrians, because running over is the first cause of mortality in traffic crashes in urban areas. Accordingly, case studies were designed to keep this proportionality.

In *Do* activity was established and tested the case studies parameters. The primary parameter that should be set was distance respect the driver's car at which the case study happens to generate a hazardous but manageable situation. For validating the correct outcome of the designed study cases, each one of them was tested individually on the driving simulation system. Table 6.3 summarizes the specification of four specific cases resulting from an inductive process with other researchers. Each case study is described through its identifier, the location features description, the involved actors, and the actor's action that produces the driving hazard.

TABLE 6.3. DESCRIPTION OF THE DESIGNED DRIVING HAZARDS.

Case studies	Description	Scenography
Case 1	Pedestrian occluded by parked cars crosses the roadway.	Residential areas with narrow streets or/and commercial areas.
Case 2	Occluded pedestrian by a car which is waiting for performing a left turning maneuver crosses the roadway.	Residential area with a lot of intersections.
Case 3	Absent-minded pedestrian crosses the roadway from the left or right sidewalk.	At any urban area location.
Case 4	Parked car merges into traffic improperly.	High volume of traffic (i.e., commercial areas and midtown)

The *Check* activity assessed the designed scenarios on the driving simulator system. To evaluate the driver's behavior in the designed scenarios, some persons had to face the driving challenge. The evaluation with a small group of drivers revealed an aspect of the driver's attention for this experiment. Most of the drivers reflected an excess of concentration when they drove in the simulator system, negatively affecting the experiment's progress. As mentioned previously, the alarm system based on the proposed architecture evaluates the driver's perception towards a specific hazard. Consequently, a drivers' excess of concentration avoids the correct assessment of the system.

At the *Act* activity, a plan for solving the underlying issue of this experiment was devised. As a result of this process, the need arose to introduce a mechanism in the designed scenarios for dispersing the driver's concentration enough to test the alarm system in a divided attention way. Hence, some divided attention events were included in the driving scenarios to correct this issue. The challenge consists in that, during the driving task, the driver must press the corresponding button from the dashboard when a visual and sound notification appears. Moreover, the hazard situations were presented to the drivers critically so that they did not have too much time to face each hazard exposition successfully.

6.4.5. Experimental setup

Ten men and women between 21 to 32 years (24.6 ± 3.13), with driving experience of more than two years and between 6000 and 15000 km/year, performed the driving trial. The experiment consists of two sessions performed on separate days.

At the first session, the dated participant gives information for the study about their driving skills. Then, the participant performs a driving task of around 5 minutes as an adaption period to the driving simulator system. After that, each driver performed the driving task in the designed scenario, including the study cases and the divided attention challenge. Finally, for collecting information about the user experience, the participants answered a brief questionnaire about how they perceived their reaction time.

At the second session, the implemented alarm system was activated. Participants were informed about incorporating this technology, and the distinct information signals were presented with an informative booklet. Each participant is faced with the driving challenge in an alternative designed scenario that contains an equal number of situations to the first session scenario. Although, driving situations were located in different spots and urban scenes to avoid participants being remembered partial or completely the situations faced in the first session. Finally, the participant answered a brief questionnaire about the experience with the alarm system. This questionnaire provides answers in a five-point Likert scale regarding the system usability and how distractive the presented ADAS was.

6.5. Results

Qualitative and quantitative information about its performance was collected in the experimental process to assess the proposed architecture.

The qualitative information was obtained from questionnaires to gather aspects only appreciable by the drivers. Table 6.4 shows a comparison between information about the driver's reaction time appreciation in the study cases 2 and 3, in the absence and presence of the warning system. It can notice that all the drivers could not react with time against these study cases in the absence of the warning system. The study cases 2 and 3 were selected because they are the critical cases in which the driver's inattention can have the most tragic consequences. In the presence of the warning system, 95% of the answers reflect the impression of reacting either just in time or with time against both study cases.

TABLE 6.4. QUALITATIVE METRICS AS A RESULT OF DRIVER QUESTIONNAIRES: REACTION TIME APPRECIATION FOR STUDY CASES 2 AND 3

Study Case	Without System			With System		
	Reaction Time Appreciation			Reaction Time Appreciation		
	No time	***Just in time***	***With time***	***No time***	***Just in time***	***With time***
Case 2	60%	40%	0%	0%	50%	50%
Case 3	30%	70%	0%	10%	50%	40%

Table 6.5 shows the drivers' system evaluation results, having as qualitative measures the system utility and the distraction level generated by the system. This table shows that 80% of the drivers assess the system usability between 4 and 5. About the distraction produced by the system, 40% of drivers felt disturbed by the warning system during the driving task performance. The reason is that the tested system does not consider the use case progress when it is identified and whether the driver has overcome it.

TABLE 6.5. QUALITATIVE METRICS AS A RESULT OF DRIVER QUESTIONNAIRES: OVERALL ACTIVE SAFETY SYSTEM EVALUATION

Qualitative Measure	Value				
	1	**2**	**3**	**4**	**5**
Utility	0%	0%	20%	50%	30%
Distracting	20%	30%	10%	40%	0%

The quantitative information obtained by objective factors in the face of occurrence of all the study cases is explained below. Figure 6.10 shows the comparative bar-plot with the accident rate that happened both in the absence and presence of the warning system, respectively. The subscripts here represent the event occurrences of the same sort of study case. It can be observed the reduction of the number of accidents in the study cases denoted as Case2, $Case3_1$, $Case4_1$. and $Case4_2$. About the study cases in which the number of accidents increased with alarm system ($Case1_1$, $Case1_2$, $Case3_3$), the drivers' mental overload could be affected negatively during the driving in the second scenario because they dealt with the driving task, the shared attention and understanding of the alarm system at the same time. In addition, the positioning of these events at the end of the scenario and the previous factor explained above suggest the apparition of tiredness signs due to the performance of three simultaneous tasks for 5 minutes.

Figure 6.11 shows a comparative Boxplot between the reaction time experimented by the drivers in the absence and presence of the warning system based on the proposed architecture. The reaction time was measured from the event that was activated until the driver reacted with a specific action on the driving controls, pressing the brake pedal most of the time. This plot can notice the reduction of the reaction time at most times in terms of average and dispersion. The case study denoted as $Case3_1$ can observe an increase of the measures, although the number of accidents for this case is mainly mitigated. Therefore, this result reflects an early warning report by the system before the event has been triggered.

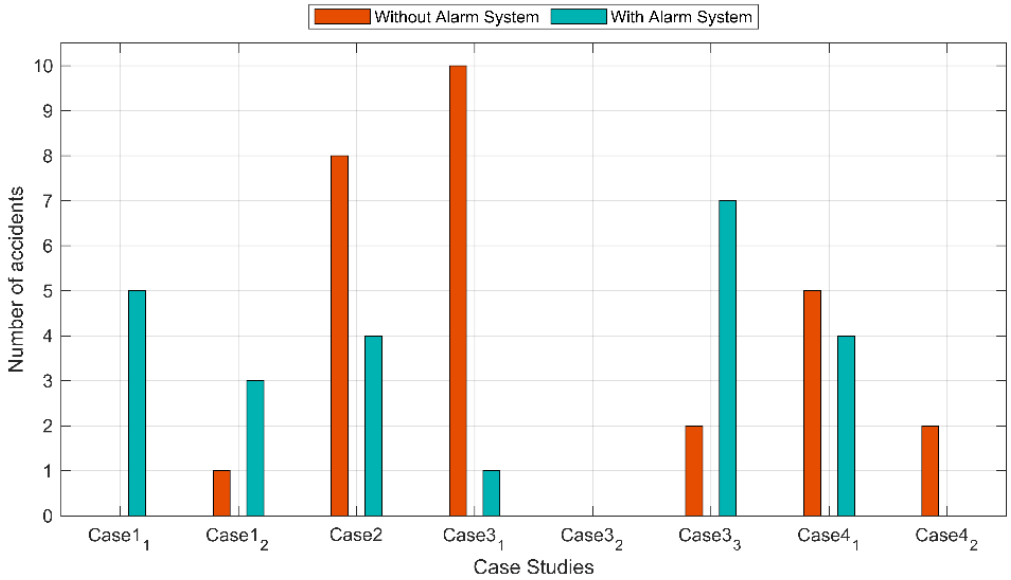


Figure 6.10. Global results in terms of the number of accidents.

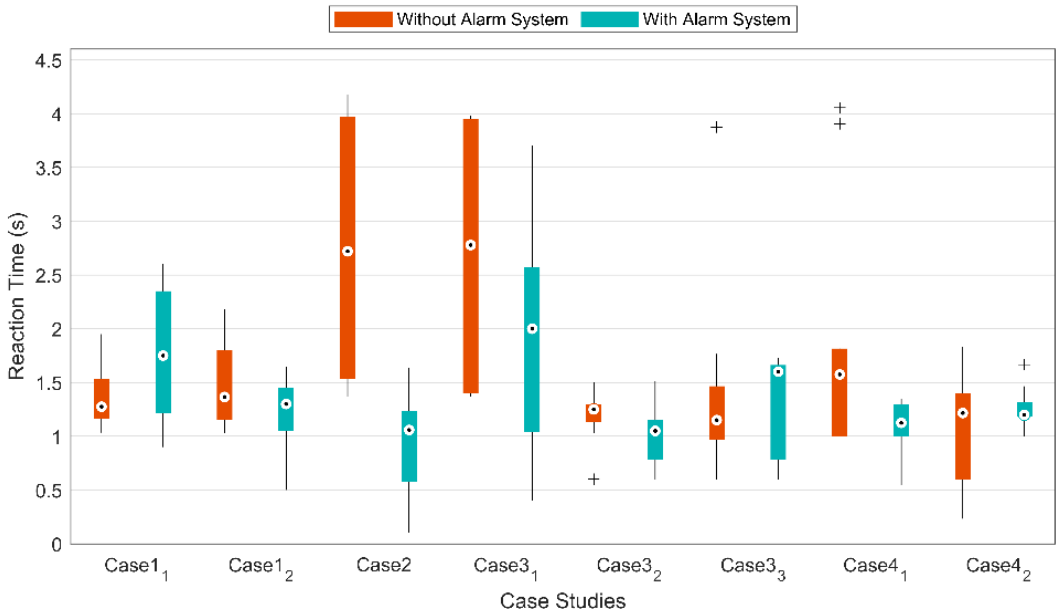


Figure 6.11. Comparison of reaction times measured in the study cases.

The design of the driving trial was evaluated by analyzing the dependency between the consecutive events. Thus, Figure 6.12 shows in general terms how the presentation of

consecutive driving hazards affects the driver in terms of the occurrence of incidents. As it can be observed, there is a slight difference between the different possibilities.

Figure 6.13 shows the elapsed time between consecutive events regarding the casuistry about what happened. In general, the driving scenario with the alarm system has a lower elapsed time in most of the cases and, for the cases without consecutive incidents, even though it has more significant elapsed time, the global statistics show a higher number of incidents than the driving scenario conducted without an alarm system.

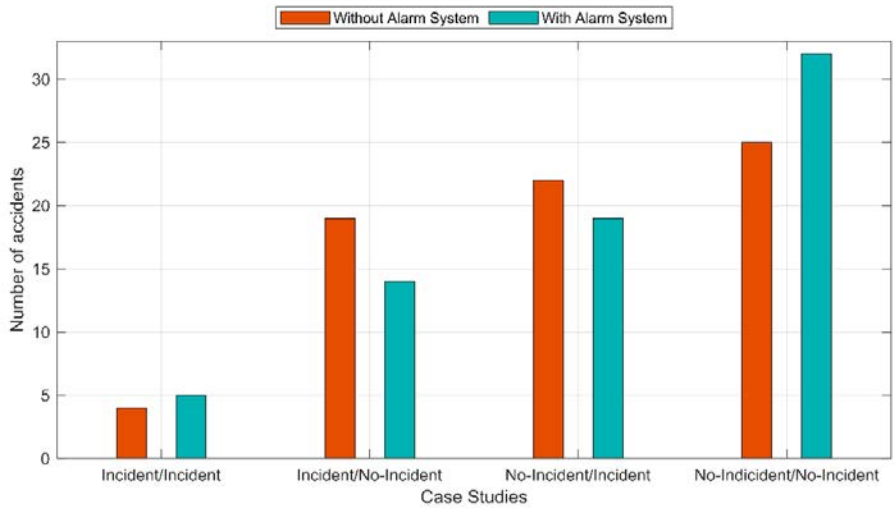


Figure 6.12. Comparative global accidents rate regarding its precedent incident case.

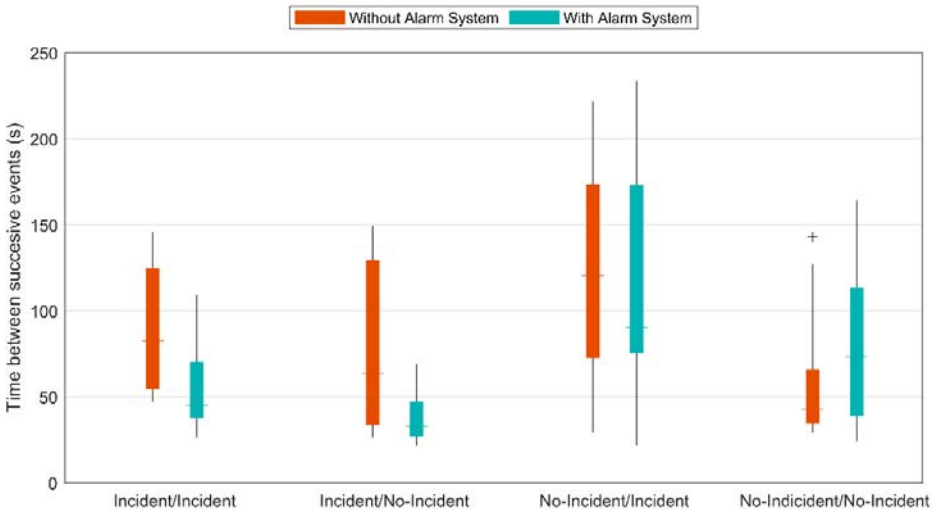


Figure 6.13. Comparative Blox-plot of obtained time between the consecutive events regarding its precedent event.

6.6. Summary

Data fusion is a crucial aspect in designing active safety systems, and it joins the present and future perspectives of human-centered approaches. However, in driving simulator environments, deploying a model based on fusion data is an arduous task.

In this work has been designed a decoupled architecture oriented to data fusion in which intervene multi-agent systems and Publish-Subscribe Messaging System. As proof of concept, an active safety system whose application gathers information from the driver monitor and the environment to determine the driver's attention was deployed on a decentralized basis. A driving trial was conducted to measure the driver's reaction time and driving experience, both with and without the deployed support system.

The integration model presents a low latency on the communication across the different deployed agents. The result of our study evidences the attenuation of the average and standard deviation of the drivers' reaction time in the presented cases of use, justifying both the usefulness of the tested warning system and the proposed architecture in this work. Regarding the results obtained from the number of accidents, the accident rate remains elevated in both driving sessions in some circumstances. Two aspects could cause this particular finding. First, the use cases immediacy makes the driver's timely reaction impossible. Second, the maintenance of constant vigilance over different areas of the screen for completing the divided attention task while coping with challenging driving could have originated tiredness and cognitive load on the drivers, being this the most plausible cause for this particular finding present in both sessions.

As a limitation of the presented work, the first integration architecture approach was performed using systems with continuous improvement cycles of life. For this reason, the experimental process was bound in some specific cases of use, and promising results were obtained. On the other hand, although the experimentation was conducted in a suitable features simulation environment, the parallelism with a videogame cannot be avoided. Besides, this study merged a set of extreme driving situations in a brief period that possibly some drivers will not deal with them in their whole life long.

Future works will include different analysis models such as stress detector and maneuver detection to increase the driver modeling perspective. In addition, it will study other methods to achieve an improved driving simulator experience in terms of visual distraction. Furthermore, the integration model will be tested in another driving simulator system to validate the proposed approach across different platforms.

7. Conclusions and future works

7.1. Contributions

This Ph. D thesis contributes to understanding the insights of analyzing human factor facets in driving, approaching the recognition of activities and cognitive states at different levels, and the design patterns required for integrating them in testing platforms. Even though proposed methods were not deployed in real-world driving tests, most non-functional requirements were exposed to addressing this thesis's presented issues, thus elevating the difficulty of achieving solutions. As a most remarkable contribution, the stress detection framework exhaustively tackles physiological responses variability issues, achieving a highly accurate model for stress pattern identification using only one wearable and non-intrusive physiological sensor. Moreover, the validation with data of well-accepted stress induction tests confirms the reliability of the proposed methodology. Another highlighted contribution of this thesis was the integration model based on mediation architectures led with driving ontology deployed in the driving simulator, providing a framework for designing and testing new approaches to improve the safety and ergonomics of driving safety applications.

7.2. Dissemination

During the predoctoral period, several topics concerning driving safety technologies have been published in JCR-indexed journals for presenting this Ph.D. thesis's results as well as in a collaborative way with other researchers.

Moreover, the participation in driving specialized conferences was conducted which oral presentations, both in poster sessions and as a lecturer.

Further, the issues and approaches tackled in this thesis allowed publishing the chapter "Integration Model of Multi-Agent Architectures for Data Fusion-Based Active Driving System" in the book "Human Factors in Intelligent Vehicles".

7.3. Future works

This thesis states the basis for upcoming projects in the CAOS research group related to driver health monitoring systems. Among the general challenges to tackle, the real-time stress detection model can be deployed using a wrist-worn wearable device, thus incorporating this information in the driver's state characterization. After that, the stress effects on the driver can be considered in the design of new safety applications at different levels of automation. For instance, in the paradigm of conditional automation driving, where

the driver must be ready to intervene in case of system failure, the driver's readiness after a take-over request depends on several subjective factors in which stress is involved.

References

[1] "Country-Level Commuting Mode Share," *Wonder*, 2020. [Online]. Available: https://askwonder.com/research/country-level-commuting-mode-share-giaipw40e. [Accessed: 09-Jul-2021].

[2] Eurostat, "Europeans travel mostly by car," *Products Eurostat News*, 2017. [Online]. Available: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20170630-1. [Accessed: 20-Dec-2020].

[3] World Health Organization, "Global status report on road safety 2018," 2018. [Online]. Available: https://www.who.int/publications/i/item/9789241565684. [Accessed: 21-Dec-2020].

[4] L. H. Donaldson, K. Brooke, and S. G. Faux, "Orthopaedic trauma from road crashes: is enough being done?," *Aust. Heal. Rev.*, vol. 33, no. 1, pp. 72–83, 2009.

[5] European Commission, "SEC/2010/0903 final - Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the regions towards a European road safety area: policy orientations on road safety 2011-2020," 2010.

[6] European Commission, "Horizon Europe - Investing to shape our future," 2021. [Online]. Available: https://ec.europa.eu/info/files/horizon-europe-investing-shape-our-future_en. [Accessed: 05-Jan-2022].

[7] European Commission, "Horizon Europe Work Programme 2021-2022. 8. Climate, Energy and Mobility (European Commission Decision C(2021)9128 of 15 December 2021)," pp. 419–422, 2021.

[8] PIARC, "Road Safety Manual, a guide for practitioners," in *World Road Association*, 2nd ed., Paris, 2004.

[9] J. J. Rolison, S. Regev, S. Moutari, and A. Feeney, "What are the factors that contribute to road accidents? An assessment of law enforcement views, ordinary drivers' opinions, and road accident records," *Accid. Anal. Prev.*, vol. 115, pp. 11–24, Jun. 2018.

[10] European Commision, "Road Safety in the European Union: Trends, Statistics and Main Challenges: Trends, statistics and main challenges," 2018.

[11] A. Chapanis, "Some Reflections on Progress," *Proc. Hum. Factors Soc. Annu. Meet.*, vol. 29, no. 1, pp. 1–8, Oct. 1985.

[12] V. Ahlstrom and K. Longo, "Human Factors Design Guide Update (Report Number DOT/FAA/CT-96/01): A Revision to Chapter 8 – Computer Human Interface Guidelines." .

[13] European Automobile Manufacturers' Association (ACEA), "Code of Practice for the Design and Evaluation of ADAS," 2009. [Online]. Available: https://www.acea.be/publications/article/code-of-practice-for-the-design-and-evaluation-of-adas. [Accessed: 23-Dec-2020].

[14] K. Bengler, K. Dietmayer, B. Farber, M. Maurer, C. Stiller, and H. Winner, "Three Decades of Driver Assistance Systems: Review and Future Perspectives," *IEEE Intell. Transp. Syst. Mag.*, vol. 6, no. 4, pp. 6–22, 2014.

[15] Volvo, "Driver Alert Control." [Online]. Available:

https://www.volvocars.com/py/support/manuals/xc40/2018w17/apoyo-del-conductor/driver-alert-control/driver-alert-control. [Accessed: 25-Dec-2020].

[16] J. A. Michon, "A Critical View of Driver Behavior Models: What Do We Know, What Should We Do?," in *Human Behavior and Traffic Safety*, Boston, MA: Springer US, 1985, pp. 485–524.

[17] M. R. Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 37, no. 1, pp. 32–64, Mar. 1995.

[18] L. K. Ainsworth and B. Kirwan, *A Guide To Task Analysis*. CRC Press, 1992.

[19] P. C. Cacciabue, *Modelling driver behaviour in automotive environments: critical issues in driver interactions with intelligent transport systems*. Springer, 2007.

[20] N. A. Stanton, M. S. Young, and G. H. Walker, "The psychology of driving automation: a discussion with Professor Don Norman," 2007.

[21] M. A. Recarte and L. M. Nunes, "Driver Distractions," in *Human Factors of Visual and Cognitive Performance in Driving*, CRC Press, 2008, pp. 75–88.

[22] SAE International, "J3016. Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems," *SAE Standard J3016*. pp. 1–16, 2018.

[23] E. Grandjean, "Fatigue in industry.," *Occup. Environ. Med.*, vol. 36, no. 3, pp. 175–186, 1979.

[24] S. K. L. Lal and A. Craig, "A critical review of the psychophysiology of driver fatigue," *Biol. Psychol.*, vol. 55, no. 3, pp. 173–194, 2001.

[25] F. Steinberger, R. Schroeter, and C. N. Watling, "From road distraction to safe driving: Evaluating the effects of boredom and gamification on driving behaviour, physiological arousal, and subjective experience," *Comput. Human Behav.*, vol. 75, pp. 714–726, 2017.

[26] D. P. Wyon, "Studies of Children under Imposed Noise and Heat Stress," *Ergonomics*, vol. 13, no. 5, pp. 598–612, Sep. 1970.

[27] W. G. Moons, N. I. Eisenberger, and S. E. Taylor, "Anger and fear responses to stress have different biological profiles," *Brain. Behav. Immun.*, vol. 24, no. 2, pp. 215–219, Feb. 2010.

[28] D. Carr and D. Umberson, "The Social Psychology of Stress, Health, and Coping," in *Handbook of Social Psychology*, J. DeLamater and A. Ward, Eds. Dordrecht: Springer Netherlands, 2013, pp. 465–487.

[29] M. R. Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 37, no. 1, pp. 32–64, 1995.

[30] J. L. Deffenbacher, E. R. Oetting, and R. S. Lynch, "Development of a driving anger scale.," *Psychol. Rep.*, vol. 74, no. 1, pp. 83–91, 1994.

[31] J. L. Deffenbacher, R. S. Lynch, E. R. Oetting, and R. C. Swaim, "The Driving Anger Expression Inventory: a measure of how people express their anger on the road," *Behav. Res. Ther.*, vol. 40, no. 6, pp. 717–737, Jun. 2002.

[32] U. Trutschel, B. Sirois, D. Sommer, M. Golz, and D. Edwards, "PERCLOS: An Alertness Measure of the Past," in *Proceedings of the 6th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design : driving assessment 2011*, 2011, pp. 172–179.

[33] P. Viola and M. Jones, "Robust real-time face detection," in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 2, pp. 747–747.

[34] V. Bazarevsky, Y. Kartynnik, A. Vakunov, K. Raveendran, and M. Grundmann, "BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs," Jul. 2019.

[35] T. Albrecht, M. Lüthi, and T. Vetter, "Deformable Models," in *Encyclopedia of Biometrics*, S. Z. Li and A. Jain, Eds. Boston, MA: Springer US, 2009, pp. 210–215.

[36] L. Gu and T. Kanade, "A generative shape regularization model for robust face alignment," in *european conference on computer vision*, 2008, pp. 413–426.

[37] C. Lugaresi *et al.*, "MediaPipe: A Framework for Building Perception Pipelines," Jun. 2019.

[38] M.-H. Sigari, M.-R. Pourshahabi, M. Soryani, and M. Fathy, "A review on driver face monitoring systems for fatigue and distraction detection," 2014.

[39] G. Sikander and S. Anwar, "Driver Fatigue Detection Systems: A Review," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2339–2352, 2019.

[40] G. Marquart, C. Cabrall, and J. de Winter, "Review of eye-related measures of drivers' mental workload," *Procedia Manuf.*, vol. 3, pp. 2854–2861, 2015.

[41] D. F. Dinges and R. Grace, "PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance," *US Dep. Transp. Fed. Highw. Adm. Publ. Number FHWA-MCRT-98-006*, 1998.

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

[43] S. Darshana, D. Fernando, S. Jayawardena, S. Wickramanayake, and C. DeSilva, "Efficient PERCLOS and gaze measurement methodologies to estimate driver attention in real time," in *2014 5th International Conference on Intelligent Systems, Modelling and Simulation*, 2014, pp. 289–294.

[44] L. Lin *et al.*, "Driver fatigue detection based on eye state," *Technol. Heal. care*, vol. 23, no. s2, pp. S453--S463, 2015.

[45] F. Zhang, J. Su, L. Geng, and Z. Xiao, "Driver Fatigue Detection Based on Eye State Recognition," in *2017 International Conference on Machine Vision and Information Technology (CMVIT)*, 2017, pp. 105–110.

[46] M. C. Catalbas, T. Cegovnik, J. Sodnik, and A. Gulten, "Driver fatigue detection based on saccadic eye movements," in *2017 10th International Conference on Electrical and Electronics Engineering (ELECO)*, 2017, pp. 913–917.

[47] M. Flores, J. Armingol, and A. de la Escalera, "Driver Drowsiness Warning System Using Visual Information for Both Diurnal and Nocturnal Illumination Conditions," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, p. 438205, Dec. 2010.

[48] J. Xu, J. Min, and J. Hu, "Real-time eye tracking for the assessment of driver fatigue," *Healthc. Technol. Lett.*, vol. 5, no. 2, pp. 54–58, 2018.

[49] M. Saradadevi and P. Bajaj, "Driver fatigue detection using mouth and yawning analysis," *Int. J. Comput. Sci. Netw. Secur.*, vol. 8, no. 6, pp. 183–188, 2008.

[50] S. Abtahi, B. Hariri, and S. Shirmohammadi, "Driver drowsiness monitoring based on yawning detection," in *2011 IEEE International Instrumentation and Measurement Technology Conference*, 2011, pp. 1–4.

[51] X. Fan, B. Yin, and Y. Sun, "Yawning Detection for Monitoring Driver Fatigue," in *2007 International Conference on Machine Learning and Cybernetics*, 2007, vol. 2, pp. 664–668.

[52] S. Abtahi, M. Omidyeganeh, S. Shirmohammadi, and B. Hariri, "YawDD: A yawning detection dataset," in *Proceedings of the 5th ACM multimedia systems conference*, 2014, pp. 24–28.

[53] R. Ekman, *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997.

[54] E. Vural, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett, and J. Movellan, "Drowsy driver detection through facial movement analysis," in *International Workshop on Human-Computer Interaction*, 2007, pp. 6–18.

[55] T. Nakamura, A. Maejima, and S. Morishima, "Detection of driver's drowsy facial expression," in *2013 2nd IAPR Asian Conference on Pattern Recognition*, 2013, pp. 749–753.

[56] M. P. S. Lorente, E. M. Lopez, L. A. Florez, A. L. Espino, J. A. I. Martínez, and A. S. de Miguel, "Explaining Deep Learning-Based Driver Models," *Appl. Sci.*, vol. 11, no. 8, p. 3321, Apr. 2021.

[57] R. Theagarajan, B. Bhanu, A. Cruz, B. Le, and A. Tambo, "Novel representation for driver emotion recognition in motor vehicle videos," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 810–814.

[58] A. Benoit and A. Caplier, "Hypovigilence analysis: open or closed eye or mouth? Blinking or yawning frequency?," in *IEEE Conference on Advanced Video and Signal Based Surveillance, 2005.*, 2005, pp. 207–212.

[59] M. A. Corbett, "Science & Technology Watch: A Drowsiness Detection System for Pilots: Optalert®," *Aviat. Space. Environ. Med.*, vol. 80, no. 2, p. 149, 2009.

[60] Cipia Vision Ltd., "Cipia Announces General Availability of Fleet Sense - Driver Monitoring Device," *prnewswire.com*, 2020. [Online]. Available: https://www.prnewswire.com/il/news-releases/cipia-announces-general-availability-of-fleet-sense---driver-monitoring-device-301194855.html. [Accessed: 15-Jan-2022].

[61] Cipia Vision Ltd., "Driver Sense - Cipia." [Online]. Available: https://cipia.com/driver-sense/. [Accessed: 15-Jan-2022].

[62] GSMA, "Quectel Launches Proactive Security Solution for Intelligent Driving to Enhance Driving Safety - Membership," 2019. [Online]. Available: https://www.gsma.com/membership/resources/quectel-launches-proactive-security-solution-for-intelligent-driving-to-enhance-driving-safety/. [Accessed: 15-Jan-2022].

[63] G. Peláez, F. García, A. de la Escalera, and J. M. Armingol, "Obtaining a 3D Model from a Facial Recognition in 2D," 2013, pp. 33–38.

[64] C. Castro, *Human Factors of Visual and Cognitive Performance in Driving*. CRC Press, 2008.

[65] "APA: 3 Types of Stress," 2018. [Online]. Available: https://crmhs.org/apa-3-types-of-stress/.

[66] S. E. Taylor, L. C. Klein, B. P. Lewis, T. L. Gruenewald, R. A. R. Gurung, and J. A. Updegraff, "Biobehavioral responses to stress in females: Tend-and-befriend, not fight-or-flight.," *Psychol. Rev.*, vol. 107, no. 3, pp. 411–429, 2000.

[67] M. S. Goligorsky, "The concept of cellular 'fight-or-flight' reaction to stress," *Am. J. Physiol. Physiol.*, vol. 280, no. 4, pp. F551–F561, Apr. 2001.

[68] H. S. Bracha, "Freeze, Flight, Fight, Fright, Faint: Adaptationist Perspectives on the Acute Stress Response Spectrum," *CNS Spectr.*, vol. 9, no. 9, pp. 679–685, Sep. 2004.

[69] A. Feder, E. J. Nestler, M. Westphal, and D. S. Charney, "Psychobiological mechanisms of resilience to stress," *Handb. adult Resil.*, pp. 35–54, 2010.

[70] B. J. Lukey and V. Tepe, *Biobehavioral resilience to stress*. CRC Press, 2008.

[71] R. W. Picard, *Affective computing*. MIT press, 2000.

[72] M. Feidakis, "A review of emotion-aware systems for e-learning in virtual environments," *Form. assessment, Learn. data Anal. gamification*, pp. 217–242, 2016.

[73] F. Shaffer, Z. M. Meehan, and C. L. Zerr, "A Critical Review of Ultra-Short-Term Heart Rate Variability Norms Research," *Front. Neurosci.*, vol. 14, Nov. 2020.

[74] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement*, vol. 28, no. 3. Physiol Meas, 01-Mar-2007.

[75] E. D. Chan, M. M. Chan, and M. M. Chan, "Pulse oximetry: Understanding its basic principles facilitates appreciation of its limitations," *Respiratory Medicine*, vol. 107, no. 6. W.B. Saunders, pp. 789–799, 01-Jun-2013.

[76] F. Shaffer and J. P. Ginsberg, "An Overview of Heart Rate Variability Metrics and Norms," *Front. Public Heal.*, vol. 5, no. September, pp. 1–17, 2017.

[77] A. Schäfer and J. Vagedes, "How accurate is pulse rate variability as an estimate of heart rate variability?," *Int. J. Cardiol.*, vol. 166, no. 1, pp. 15–29, Jun. 2013.

[78] E. Yuda *et al.*, "Pulse rate variability: a new biomarker, not a surrogate for heart rate variability," *J. Physiol. Anthropol.*, vol. 39, no. 1, p. 21, Dec. 2020.

[79] E. Mejía-Mejía, K. Budidha, T. Y. Abay, J. M. May, and P. A. Kyriacou, "Heart Rate Variability (HRV) and Pulse Rate Variability (PRV) for the Assessment of Autonomic Responses," *Front. Physiol.*, vol. 11, Jul. 2020.

[80] U. R. Acharya, K. P. Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: a review," *Med. Biol. Eng. Comput.*, vol. 44, no. 12, pp. 1031–1051, 2006.

[81] R. P. Sloan, P. A. Shapiro, E. Bagiella, J. T. Bigger, E. S. Lo, and J. M. Gorman, "Relationships between circulating catecholamines and low frequency heart period variability as indices of cardiac sympathetic activity during mental stress," *Psychosom. Med.*, vol. 58, no. 1, pp. 25–31, 1996.

[82] M. S. Roy, R. Gupta, J. K. Chandra, K. Das Sharma, and A. Talukdar, "Improving photoplethysmographic measurements under motion artifacts using artificial neural network for personal healthcare," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 12, pp. 2820–2829, 2018.

[83] T. Tamura, Y. Maeda, M. Sekine, and M. Yoshida, "Wearable Photoplethysmographic Sensors—Past and Present," *Electronics*, vol. 3, no. 2, pp. 282–302, Apr. 2014.

[84] D. Nabil and F. Bereksi Reguig, "Ectopic beats detection and correction methods: A review," *Biomed. Signal Process. Control*, vol. 18, pp. 228–244, Apr. 2015.

[85] G. G. Berntson *et al.*, "Heart rate variability: origins, methods, and interpretive caveats," *Psychophysiology*, vol. 18, no. 6, pp. 623–648, 1997.

[86] F. Wen and F. T. He, "An efficient method of addressing ectopic beats: New insight

into data preprocessing of heart rate variability analysis," *J. Zhejiang Univ. Sci. B*, vol. 12, no. 12, pp. 976–982, Dec. 2011.

[87] K. Solem, P. Laguna, and L. Sornmo, "An Efficient Method for Handling Ectopic Beats Using the Heart Timing Signal," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 1, pp. 13–20, Jan. 2006.

[88] J. Mateo and P. Laguna, "Analysis of heart rate variability in the presence of ectopic beats using the heart timing signal," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 3, pp. 334–343, Mar. 2003.

[89] D. Nabil and F. Bereksi Reguig, "Ectopic beats detection and correction methods: A review," *Biomed. Signal Process. Control*, vol. 18, pp. 228–244, Apr. 2015.

[90] M. Benedek and C. Kaernbach, "A continuous measure of phasic electrodermal activity," *J. Neurosci. Methods*, vol. 190, no. 1, pp. 80–91, Jun. 2010.

[91] W. Boucsein, *Electrodermal activity*. Springer Science & Business Media, 2012.

[92] M. E. Dawson, A. M. Schell, and D. L. Filion, "The electrodermal system.," 2017.

[93] K. Saga, "Structure and function of human sweat glands studied with histochemistry and cytochemistry," *Prog. Histochem. Cytochem.*, vol. 37, no. 4, pp. 323–386, Jan. 2002.

[94] M. van Dooren, J. J. G. G. J. de Vries, and J. H. Janssen, "Emotional sweating across the body: Comparing 16 different skin conductance measurement locations," *Physiol. Behav.*, vol. 106, no. 2, pp. 298–304, May 2012.

[95] A. Greco, G. Valenza, A. Lanata, E. Scilingo, and L. Citi, "cvxEDA: a Convex Optimization Approach to Electrodermal Activity Processing," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 1–1, 2016.

[96] S. Prabhakaran and N. K. Ayyamperumal, "A modified compressed sensing-based decomposition approach to analyze electrodermal activity signals using matrix-free convex optimization," *Int. J. Intell. Eng. Syst.*, vol. 13, no. 1, pp. 46–56, 2020.

[97] A. S. Anusha, J. Jose, S. P. Preejith, J. Jayaraj, and S. Mohanasankar, "Physiological signal based work stress detection using unobtrusive sensors," *Biomed. Phys. Eng. Express*, vol. 4, no. 6, 2018.

[98] D. Leiner, A. Fahr, and H. Früh, "EDA Positive Change: A Simple Algorithm for Electrodermal Activity to Measure General Audience Arousal During Media Exposure," *Commun. Methods Meas.*, vol. 6, no. 4, pp. 237–250, Oct. 2012.

[99] M. Gjoreski, M. Luštrek, M. Gams, and H. Gjoreski, "Monitoring stress with a wrist device using context," *J. Biomed. Inform.*, vol. 73, pp. 159–170, Sep. 2017.

[100] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Troster, and U. Ehlert, "Discriminating Stress From Cognitive Load Using a Wearable EDA Device," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 410–417, Mar. 2010.

[101] A. S. Anusha *et al.*, "Electrodermal Activity Based Pre-surgery Stress Detection Using a Wrist Wearable," *IEEE J. Biomed. Heal. Informatics*, vol. 24, no. 1, pp. 1–1, 2019.

[102] M. Kołodziej, P. Tarnowski, A. Majkowski, and R. J. Rak, "Electrodermal activity measurements for detection of emotional arousal," *Bull. Polish Acad. Sci. Tech. Sci.*, vol. 67, no. 4, pp. 813–826, 2019.

[103] A. S. Anusha, J. Jose, S. P. Preejith, J. Jayaraj, and S. Mohanasankar, "Physiological signal based work stress detection using unobtrusive sensors," *Biomed. Phys. Eng. Express*, vol. 4, no. 6, p. 65001, Sep. 2018.

[104] J. J. Braithwaite and D. G. Watson, "Issues Surrounding the Normalization and Standardisation of Skin Conductance Responses (SCRs).," University of Birmingham, 2015.

[105] S. Cohen, T. Kamarck, and R. Mermelstein, "A Global Measure of Perceived Stress," *J. Health Soc. Behav.*, vol. 24, no. 4, p. 385, Dec. 1983.

[106] T. H. Holmes and R. H. Rahe, "The social readjustment rating scale," *J. Psychosom. Res.*, vol. 11, no. 2, pp. 213–218, Aug. 1967.

[107] R. Castaldo, L. Montesinos, P. Melillo, C. James, and L. Pecchia, "Ultra-short term HRV features as surrogates of short term HRV: A case study on mental stress detection in real life," *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, pp. 1–13, 2019.

[108] M. Lohani, B. R. Payne, and D. L. Strayer, "A Review of Psychophysiological Measures to Assess Cognitive States in Real-World Driving," *Front. Hum. Neurosci.*, vol. 13, Mar. 2019.

[109] J. A. Healey, "Wearable and automotive systems for affect recognition from physiology," Massachusetts Institute of Technology, 2000.

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

[111] M. N. Rastgoo, B. Nakisa, A. Rakotonirainy, V. Chandran, and D. Tjondronegoro, "A Critical Review of Proactive Detection of Driver Stress Levels Based on Multimodal Measurements," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 1–35, Sep. 2019.

[112] M. Subramaniyam *et al.*, "Recent developments on driver's health monitoring and comfort enhancement through IoT," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 402, p. 012064, Sep. 2018.

[113] E. Strickland, "3 Ways Ford Cars Could Monitor Your Health - IEEE Spectrum," *https://spectrum.ieee.org/*, 2017. [Online]. Available: https://spectrum.ieee.org/3-ways-ford-cars-could-monitor-your-health. [Accessed: 13-Jan-2022].

[114] Toyota Collaborative Safety Research Center, "Making Roads Safer by Detecting Driver Heart Anomalies - Toyota USA Newsroom," *https://pressroom.toyota.com/*, 2020. [Online]. Available: https://pressroom.toyota.com/making-roads-safer-by-detecting-driver-heart-anomalies/. [Accessed: 13-Jan-2022].

[115] M. Jabon, J. Bailenson, E. Pontikakis, L. Takayama, and C. Nass, "Facial expression analysis for predicting unsafe driving behavior," *IEEE Pervasive Comput.*, vol. 10, no. 4, pp. 84–95, 2011.

[116] L. Fletcher and A. Zelinsky, "Driver inattention detection based on eye gaze--road event correlation," *Int. J. Rob. Res.*, vol. 28, no. 6, pp. 774–801, Jun. 2009.

[117] A. Allamehzadeh and C. Olaverri-Monreal, "Automatic and manual driving paradigms: Cost-efficient mobile application for the assessment of driver inattentiveness and detection of road conditions," in *IEEE Intelligent Vehicles Symposium, Proceedings*, 2016, vol. 2016-Augus, pp. 26–31.

[118] V. Leonhardt, T. Pech, and G. Wanielik, "Data fusion and assessment for maneuver prediction including driving situation and driver behavior," *FUSION 2016 - 19th Int. Conf. Inf. Fusion, Proc.*, pp. 1702–1708, 2016.

[119] B. G. Lee and W. Y. Chung, "A smartphone-based driver safety monitoring system using data fusion," *Sensors (Switzerland)*, vol. 12, no. 12, pp. 17536–17552, 2012.

[120] G. Rigas, Y. Goletsis, and D. I. Fotiadis, "Real-Time Driver's Stress Event Detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 1, pp. 221–234, Mar. 2012.

[121] C. Gold, D. Damböck, L. Lorenz, and K. Bengler, ""Take over!' How long does it take to get the driver back into the loop?," *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 57, no. 1, pp. 1938–1942, Sep. 2013.

[122] Z. Lu, B. Zhang, A. Feldhütter, R. Happee, M. Martens, and J. C. F. De Winter, "Beyond mere take-over requests: The effects of monitoring requests on driver attention, take-over performance, and acceptance," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 63, pp. 22–37, May 2019.

[123] X. Li, R. Schroeter, A. Rakotonirainy, J. Kuo, and M. G. Lenné, "Effects of different non-driving-related-task display modes on drivers' eye-movement patterns during take-over in an automated vehicle," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 70, pp. 135–148, Apr. 2020.

[124] Z. Xiong, V. V. Dixit, and S. T. Waller, "The development of an Ontology for driving Context Modelling and reasoning," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, 2016, pp. 13–18.

[125] A. Armand, D. Filliat, and J. Ibanez-Guzman, "Ontology-based context awareness for driving assistance systems," *IEEE Intell. Veh. Symp. Proc.*, no. Iv, pp. 227–233, 2014.

[126] J.-P. Barthès, P. Bonnifait, and P. B. Multi, "Multi-Agent Active Interaction with Driving Assistance Systems," 2010.

[127] I. Colwell, B. Phan, S. Saleem, R. Salay, and K. Czarnecki, "An automated vehicle safety concept based on runtime restriction of the operational design domain," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018, pp. 1910–1917.

[128] P. C. Cacciabue, B. W. OFF, M. Z. OFF, E. V. SUP, and D. B. SUP, "Ontology of the driving task," 2008.

[129] Member States Initiatives, "Regulation (EU) 2019/2144 of the European Parliament and the Council of 27 November 2019," *Off. J. Eur. Union*, vol. 62, no. November, 2019.

[130] R. Christ *et al.*, "GADGET: Guarding Automobile Drivers through Guidance Education and Technology. Final Report, Investigations on Influences upon Driver Behaviour - Safety Approaches in Comparison and Combination," Vienna, 1999.

[131] M. Panou, E. Bekiaris, and V. Papakostopoulos, "Modelling Driver Behaviour in European Union and International Projects," in *Modelling Driver Behaviour in Automotive Environments*, London: Springer London, 2007, pp. 3–25.

[132] N. AbuAli and H. Abou-zeid, "Driver behavior modeling: Developments and future directions," *Int. J. Veh. Technol.*, vol. 2016, 2016.

[133] W. Wang, J. Xi, and D. Zhao, "Learning and Inferring a Driver's Braking Action in Car-Following Scenarios," Jan. 2018.

[134] A. Doshi and M. M. Trivedi, "Tactical driver behavior prediction and intent inference: A review," in *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2011, pp. 1892–1897.

[135] J. Dai, B. Yang, C. Guo, and Z. Ding, "Personalized route recommendation using big trajectory data," in *2015 IEEE 31st International Conference on Data Engineering*, 2015, pp. 543–554.

[136] S.-Y. Wang, Y.-W. Cheng, C.-C. Lin, W.-J. Hong, and T.-W. He, "A Vehicle Collision Warning System Employing Vehicle-To-Infrastructure Communications," in *2008 IEEE Wireless Communications and Networking Conference*, 2008, pp. 3075–3080.

[137] J.-H. Hong, B. Margines, and A. K. Dey, "A smartphone-based sensing platform to model aggressive driving behaviors," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, pp. 4047–4056.

[138] F. Zheng, J. Li, H. van Zuylen, and C. Lu, "Influence of driver characteristics on emissions and fuel consumption," *Transp. Res. Procedia*, vol. 27, pp. 624–631, 2017.

[139] B. Varga, A. Sagoian, and F. Mariasiu, "Prediction of Electric Vehicle Range: A Comprehensive Review of Current Issues and Challenges," *Energies*, vol. 12, no. 5, p. 946, Mar. 2019.

[140] I. Škrjanc, G. Andonovski, A. Ledezma, O. Sipele, J. A. Iglesias, and A. Sanchis, "Evolving cloud-based system for the recognition of drivers' actions," *Expert Syst. Appl.*, vol. 99, pp. 231–238, Jun. 2018.

[141] G. Andonovski, O. Sipele, J. A. Iglesias, A. Sanchis, E. Lughofer, and I. Skrjanc, "Detection of driver maneuvers using evolving fuzzy cloud-based system," in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2020, pp. 700–706.

[142] STISIM, "M300WS driving simulation system." [Online]. Available: http://stisimdrive.com/wp-content/uploads/2016/08/new-M300WS-datasheet.pdf.

[143] P. Angelov and R. Yager, "Simplified fuzzy rule-based systems using non-parametric antecedents and relative data density," in *2011 IEEE Workshop on Evolving and Adaptive Intelligent Systems (EAIS)*, 2011, pp. 62–69.

[144] T. Pereira, P. R. Almeida, J. P. S. Cunha, and A. Aguiar, "Heart rate variability metrics for fine-grained stress level assessment," *Comput. Methods Programs Biomed.*, vol. 148, pp. 71–80, Sep. 2017.

[145] I. Labuschagne, C. Grace, P. Rendell, G. Terrett, and M. Heinrichs, "An introductory guide to conducting the Trier Social Stress Test," 2019.

[146] H.-C. Leung, "An Event-related Functional MRI Study of the Stroop Color Word Interference Task," *Cereb. Cortex*, vol. 10, no. 6, pp. 552–560, Jun. 2000.

[147] P. Satish, K. Muralikrishnan, K. Balasubramanian, and Shanmugapriya, "Heart rate variability changes during stroop color and word test among genders.," *Indian J. Physiol. Pharmacol.*, vol. 59, no. 1, pp. 9–15.

[148] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: The PANAS scales.," *J. Pers. Soc. Psychol.*, vol. 54, no. 6, pp. 1063–1070, 1988.

[149] EMPATICA S.R.L, "Empatica E4," p. 4, 2014.

[150] L. Vanitha, G. Suresh, M. Chandrasekar, and P. Punita, "Development of four stress levels in group stroop colour word test using HRV analysis," Allied Academies.

[151] J. Karim, R. Weisz, and S. U. Rehman, "International positive and negative affect schedule short-form (I-PANAS-SF): Testing for factorial invariance across cultures," *Procedia - Soc. Behav. Sci.*, vol. 15, pp. 2016–2022, 2011.

[152] J. W. Roberti, L. N. Harrington, and E. A. Storch, "Further Psychometric Support for the 10-Item Version of the Perceived Stress Scale," *J. Coll. Couns.*, vol. 9, no. 2, pp.

135–147, Sep. 2006.

[153] K. M. Connor and J. R. T. Davidson, "Development of a new resilience scale: The Connor-Davidson Resilience Scale (CD-RISC)," *Depress. Anxiety*, vol. 18, no. 2, pp. 76–82, Sep. 2003.

[154] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biol. Psychol.*, vol. 84, no. 3, pp. 394–421, Jul. 2010.

[155] N. Schneiderman, G. Ironson, and S. D. Siegel, "Stress and health: Psychological, behavioral, and biological determinants," *Annual Review of Clinical Psychology*, vol. 1. NIH Public Access, pp. 607–628, 2005.

[156] J. Wijsman, B. Grundlehner, Hao Liu, H. Hermens, and J. Penders, "Towards mental stress detection using wearable physiological sensors," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 1798–1801.

[157] B. W. Nelson, C. A. Low, N. Jacobson, P. Areán, J. Torous, and N. B. Allen, "Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research," *npj Digit. Med.*, vol. 3, no. 1, p. 90, Dec. 2020.

[158] Y. Maeda, M. Sekine, T. Tamura, A. Moriya, T. Suzuki, and K. Kameyama, "Comparison of reflected green light and infrared photoplethysmography," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008, pp. 2270–2272.

[159] C. S. Carver and M. F. Scheier, *Perspectives on personality*. Pearson education, 2012.

[160] G. Stemmler and J. Wacker, "Personality, emotion, and individual differences in physiological responses," *Biol. Psychol.*, vol. 84, no. 3, pp. 541–551, Jul. 2010.

[161] M. A. Birkett, "The Trier Social Stress Test protocol for inducing psychological stress.," *J. Vis. Exp.*, no. 56, Oct. 2011.

[162] A. Mukherjee, "Recovering Skin Conductance Responses in Under-sampled Electrodermal Activity Data from Wearables," The University of Toledo, 2019.

[163] H. F. Posada-Quintero, J. P. Florian, A. D. Orjuela-Cañón, T. Aljama-Corrales, S. Charleston-Villalobos, and K. H. Chon, "Power Spectral Density Analysis of Electrodermal Activity for Sympathetic Function Assessment," *Ann. Biomed. Eng.*, vol. 44, no. 10, pp. 3124–3135, Oct. 2016.

[164] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection," in *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, 2018, pp. 400–408.

[165] A. C. Samson, S. D. Kreibig, B. Soderstrom, A. A. Wade, and J. J. Gross, "Eliciting positive, negative and mixed emotional states: A film library for affective scientists," *Cogn. Emot.*, vol. 30, no. 5, pp. 827–856, Jul. 2016.

[166] U. R. Hülsheger, H. J. E. M. Alberts, A. Feinholdt, and J. W. B. Lang, "Benefits of mindfulness at work: The role of mindfulness in emotion regulation, emotional exhaustion, and job satisfaction.," *J. Appl. Psychol.*, vol. 98, no. 2, pp. 310–325, Mar. 2013.

[167] T. A. B. Snijders and R. J. Bosker, *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. sage, 2011.

[168] E. Vildjiounaite *et al.*, "Unobtrusive stress detection on the basis of smartphone usage

data," *Pers. Ubiquitous Comput.*, vol. 22, no. 4, pp. 671–688, Aug. 2018.

[169] R. Wirth and J. Hipp, "CRISP-DM: Towards a standard process model for data mining," in *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, 2000, vol. 1.

[170] A. A. Neath and J. E. Cavanaugh, "The Bayesian information criterion: background, derivation, and applications," *WIREs Comput. Stat.*, vol. 4, no. 2, pp. 199–203, Mar. 2012.

[171] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu, "Understanding of internal clustering validation measures," in *2010 IEEE international conference on data mining*, 2010, pp. 911–916.

[172] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987.

[173] T. M. Kodinariya and P. R. Makwana, "Review on determining number of Cluster in K-Means Clustering," *Int. J.*, vol. 1, no. 6, pp. 90–95, 2013.

[174] E. Vildjiounaite, J. Kallio, J. Mäntyjärvi, V. Kyllönen, M. Lindholm, and G. Gimel'farb, "Unsupervised Stress Detection Algorithm and Experiments with Real Life Data," 2017, pp. 95–107.

[175] J. Tervonen *et al.*, "Personalized mental stress detection with self-organizing map: From laboratory to the field," *Comput. Biol. Med.*, vol. 124, p. 103935, Sep. 2020.

[176] D. Das, S. Datta, T. Bhattacharjee, A. D. Choudhury, and A. Pal, "Eliminating Individual Bias to Improve Stress Detection from Multimodal Physiological Data," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 5753–5758.

[177] D. Reynolds, "Gaussian Mixture Models," in *Encyclopedia of Biometrics*, Boston, MA: Springer US, 2009, pp. 659–663.

[178] I.-V. Bornoiu and O. Grigore, "A study about feature extraction for stress detection," in *2013 8Th International Symposium on Advanced Topics In Electrical Engineering (ATEE)*, 2013, pp. 1–4.

[179] J. J. Batzel and F. Kappel, "Time delay in physiological systems: Analyzing and modeling its impact," *Math. Biosci.*, vol. 234, no. 2, pp. 61–74, Dec. 2011.

[180] M. Löning, A. Bagnall, S. Ganesh, V. Kazakov, J. Lines, and F. J. Király, "sktime: A unified interface for machine learning with time series," *arXiv Prepr. arXiv1909.07872*, 2019.

[181] P. Virtanen *et al.*, "{SciPy} 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nat. Methods*, vol. 17, pp. 261–272, 2020.

[182] D. Makowski *et al.*, "{NeuroKit}2: A Python toolbox for neurophysiological signal processing," *Behav. Res. Methods*, vol. 53, no. 4, pp. 1689–1696, Feb. 2021.

[183] J. Shukla, M. Barreda-Angeles, J. Oliver, G. C. Nandi, and D. Puig, "Feature Extraction and Selection for Emotion Recognition from Electrodermal Activity," *IEEE Trans. Affect. Comput.*, vol. 12, no. 4, pp. 857–869, Oct. 2021.

[184] H. Feng, H. M. Golshan, and M. H. Mahoor, "A wavelet-based approach to emotion classification using EDA signals," *Expert Syst. Appl.*, vol. 112, pp. 77–86, Dec. 2018.

[185] S. Taylor, N. Jaques, Weixuan Chen, S. Fedor, A. Sano, and R. Picard, "Automatic identification of artifacts in electrodermal activity data," in *2015 37th Annual*

International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015, pp. 1934–1937.

[186] C.-P. Hsieh, Y.-T. Chen, W.-K. Beh, and A.-Y. A. Wu, "Feature Selection Framework for XGBoost Based on Electrodermal Activity in Stress Detection," in *2019 IEEE International Workshop on Signal Processing Systems (SiPS)*, 2019, pp. 330–335.

[187] P. S. Addison, *The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance*. CRC press, 2017.

[188] R. Polikar and others, "The wavelet tutorial." 1996.

[189] J. Shlens, "A Tutorial on Principal Component Analysis," Apr. 2014.

[190] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2005.

[191] V. Satopaa, J. Albrecht, D. Irwin, and B. Raghavan, "Finding a 'Kneedle' in a Haystack: Detecting Knee Points in System Behavior," in *2011 31st International Conference on Distributed Computing Systems Workshops*, 2011, pp. 166–171.

[192] G. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt, and A. O'Leary, "PyWavelets: A Python package for wavelet analysis," *J. Open Source Softw.*, vol. 4, no. 36, p. 1237, Apr. 2019.

[193] L. Buitinck *et al.*, "{API} design for machine learning software: experiences from the scikit-learn project," in *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, 2013, pp. 108–122.

[194] J. Vilo, "Discovering frequent patterns from strings," *Dep. Comput. Sci. Univ. Helsinki*, 1998.

[195] N. Jambigi, T. Chanda, V. Unnikrishnan, and M. Spiliopoulou, "Assessing the Difficulty of Labelling an Instance in Crowdworking," in *ECML PKDD 2020 Workshops*, 2020, pp. 363–373.

[196] P. Schäfer, "The BOSS is concerned with time series classification in the presence of noise," *Data Min. Knowl. Discov.*, vol. 29, no. 6, pp. 1505–1530, 2015.

[197] P. Schäfer and U. Leser, "Fast and Accurate Time Series Classification with WEASEL," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 637–646.

[198] R. Henriques, A. Paiva, and C. Antunes, "Accessing Emotion Patterns from Affective Interactions Using Electrodermal Activity," in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, 2013, pp. 43–48.

[199] J. Lin, E. Keogh, L. Wei, and S. Lonardi, "Experiencing SAX: a novel symbolic representation of time series," *Data Min. Knowl. Discov.*, vol. 15, no. 2, pp. 107–144, 2007.

[200] J. A. Iglesias, A. Ledezma, and A. Sanchis, "Sequence Classification Using Statistical Pattern Recognition," in *Advances in Intelligent Data Analysis VII*, vol. 4723 LNCS, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 207–218.

[201] R. Murugappan, J. J. Bosco, K. Eswaran, P. Vijay, and V. Vijayaraghavan, "User Independent Human Stress Detection," in *2020 IEEE 10th International Conference on Intelligent Systems (IS)*, 2020, pp. 490–497.

[202] S. A. H. Aqajari, E. K. Naeini, M. A. Mehrabadi, S. Labbaf, A. M. Rahmani, and N. Dutt, "Gsr analysis for stress: Development and validation of an open source tool for noisy naturalistic gsr data," *arXiv Prepr. arXiv2005.01834*, 2020.

[203] S. Föll *et al.*, "FLIRT: A feature generation toolkit for wearable data," *Comput. Methods Programs Biomed.*, vol. 212, p. 106461, Nov. 2021.

[204] I. Oppenheim and D. Shinar, "Human factors and ergonomics," in *Handbook of traffic psychology*, Elsevier, 2011, pp. 193–211.

[205] Stanford Artificial Intelligence Laboratory, "Robotic Operating System." .

[206] V. Zamora, O. Sipele, A. Ledezma, and A. Sanchis, "Intelligent Agents for Supporting Driving Tasks: An Ontology-based Alarms System," in *3rd International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS)*, 2017, pp. 165–172.

[207] International Stardard Orgatization, "ISO 11429:1996 - Ergonomics — System of auditory and visual danger and information signals," 2014. [Online]. Available: https://www.iso.org/standard/19369.html. [Accessed: 20-Jan-2022].

[208] Dirección General de Tráfico, "Portal estadístico," 2021. [Online]. Available: https://sedeapl.dgt.gob.es/WEB_IEST_CONSULTA/informePredefinidoCaptcha.faces. [Accessed: 20-Jan-2022].

Annex A. Self-assessment questionnaire

Information that can identify the individuals involved in this study will not be stored. The data collected will only have investigative purposes without any profit. Please, choose your answers as honestly as possible; otherwise, the study results could be affected.

Sociodemographic information

Date *(DD/MM/YYYY)* _____

Age _____

Sex *Female* ☐ *Male* ☐

Education *School* ☐ 1 *High-school* ☐ 2 *Labour education* ☐ 3 *College* ☐ 4

Occupation *Student* ☐ 1 *Partial working* ☐ 2 *Fulltime working* ☐ 3 *Unemployed* ☐ 4 *Retired* ☐ 5

Driving skills

Length of driving license _____

Driving experience (Kms/year) *Lesser 1000* ☐ 1 *Between 1000 and 5000* ☐ 2 *Between 5000 and 10000* ☐ 3 *More than 10000* ☐ 4

Affective State

The questions in this scale ask your mood state. Indicate the extent you have felt this way over the past week.

	Very slightly or not at all	**A little**	**Moderately**	**Quite a bit**	**Extremely**
1. Active	☐*1*	☐*2*	☐*3*	☐*4*	☐*5*
2. Afraid	☐*1*	☐*2*	☐*3*	☐*4*	☐*5*
3. Alert	☐*1*	☐*2*	☐*3*	☐*4*	☐*5*
4. Ashamed	☐*1*	☐*2*	☐*3*	☐*4*	☐*5*

	Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
5. Attentive	☐ *1*	☐ *2*	☐ *3*	☐ *4*	☐ *5*
6. Hostile	☐ *1*	☐ *2*	☐ *3*	☐ *4*	☐ *5*
7. Determined	☐ *1*	☐ *2*	☐ *3*	☐ *4*	☐ *5*
8. Nervous	☐ *1*	☐ *2*	☐ *3*	☐ *4*	☐ *5*
9. Inspired	☐ *1*	☐ *2*	☐ *3*	☐ *4*	☐ *5*
10. Upset	☐ *1*	☐ *2*	☐ *3*	☐ *4*	☐ *5*

Stress perceived

The questions in this scale ask you about your feelings and thoughts during the last month. In each case, you will be asked to indicate by marking how often you felt or thought a certain way.

Answer values:

0 = Never 1 = Almost Never 2 = Sometimes 3 = Fairly Often 4 = Very Often

Question	Answers				
	0	1	2	3	4
1. In the last month, how often have you been upset because of something that happened unexpectedly?					
2. In the last month, how often have you felt that you were unable to control the important things in your life?					
3. In the last month, how often have you felt nervous and “stressed”?					
4. In the last month, how often have you felt confident about your ability to handle your personal problems?					
5. In the last month, how often have you felt that things were going your way?					
6. In the last month, how often have you found that you could not cope with all the things that you had to do?					
7. In the last month, how often have you been able to control irritations in your life?					
8. In the last month, how often have you felt that you were on top of things?					
9. In the last month, how often have you been angered because of things that were outside of your control?					
10. In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?					

Driving session

The questions as follows are related to the stress that you have perceived during the driving session.

1. In general, assess the stress you have experienced during the experiment.

Non-stress	☐ *1*	☐ *2*	☐ *3*	☐ *4*	☐ *5*	*Much stress*

2. At the beginning of the experiment, you were driving through a quiet area. Considering this as normal driving, did you feel any stress in that part?

Yes ☐ *No* ☐

3. Assess the stress you have experienced in the last part of the experiment, in which you had to press the dashboard buttons.

Non-stress	☐ *1*	☐ *2*	☐ *3*	☐ *4*	☐ *5*	*Much stress*

4. In your opinion, which was the most hazardous event?

☐ a) Pedestrian halfway across
☐ b) The boy who crosses to catch the ball
☐ c) The car that merges improperly
☐ d) Crossing dog
☐ e) The police car with emergency lights

5. Indicates the degree of stress caused by the situation: a) Pedestrian halfway across:

Non-stress	☐ *1*	☐ *2*	☐ *3*	☐ *4*	☐ *5*	*Much stress*

6. Indicates the degree of stress caused by the situation: b) Boy who crosses to catch the ball:

Non-stress	☐ *1*	☐ *2*	☐ *3*	☐ *4*	☐ *5*	*Much stress*

7. Indicates the degree of stress caused by the situation: c) The car that merges improperly:

Non-stress	☐ *1*	☐ *2*	☐ *3*	☐ *4*	☐ *5*	*Much stress*

8. Indicates the degree of stress caused by the situation: d) Crossing dog:

Non-stress	☐ *1*	☐ *2*	☐ *3*	☐ *4*	☐ *5*	*Much stress*

9. Indicates the degree of stress caused by the situation: e) Police car with emergency lights:

Non-stress	☐ *1*	☐ *2*	☐ *3*	☐ *4*	☐ *5*	*Much stress*

Annex B. Data use agreement

Identifying data

Surname, Name:
ID:
email:

Background

URBANITA-SAFE[1] y CAV[2] projects address monitoring the driver by driving controls, cameras, and wearable sensors such as smartwatches. The physiological information will be employed to determine cognitive states such as cognitive load and stress. Moreover, the information from driving controls will allow identifying specific driving patterns and the performing of driving-related activities as well.

Data limited

According to General Data Protection Regulation, the collected data only holds the necessary data for the stated purpose. Thus, collected data will exclude direct identifiers (name). However, that may include age, city, state, elements of date, and other numbers and characteristics.

Data use agreement

I hereby give my explicit consent to transfer my personal data to develop and execute the research project called URBANITA-SAFE and CAV executed by the Universidad Carlos III de Madrid.

I declare that I have received sufficient information on the content of the research project and that all my doubts and questions have been resolved. My participation in this study is absolutely voluntary. In addition, I have been informed of the possibility of revoking my consent at any time.

Mark in your case with an X

☐ I expressly authorize my personal data to be published on the occasion of the dissemination of the project results, for which it will be sufficient to inform me thereof, through a communication to the email provided.

☐ I expressly authorize that my data be used for the execution and development of other similar projects, for which it will be enough to inform me about it, through communication to the email provided.

Leganés, ___ (day) ____(month) of 20__ Signature _____

[1] Sistema de Ayuda para una Conducción Urbana más Segura (TRA2015-63708-R
[2] Integración de sistemas cooperativos para vehículos autónomos en trafico compartido: Análisis del entorno de conducción (TRA2016-78886-C3-1-R)

Annex C. Results of stress classification models

This annex presents the experimental results of binary stress detection using the WESAD dataset, distinguishing the general and semi-personalized approaches.

TABLE C.1. SCORE METRICS FOR GENERAL STRESS-VS-NON-STRESS CLASSIFICATION MODELS.

Experiment	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC AUC (x100)
Only PCA	Default	77.83±0.04	0.00±0.00	0.00±0.00	0.00±0.00	50.00±0.00
	LinearSVC	87.14±0.06	82.10±0.78	53.69±0.77	64.91±0.35	90.02±0.05
	LDA	86.23±0.14	85.83±0.25	45.39±0.70	59.38±0.59	90.01±0.06
	GradientBoosting	92.37±0.14	82.76±0.53	82.83±0.69	82.80±0.33	96.80±0.12
	LogisticRegression	87.08±0.05	80.56±0.52	54.98±0.68	65.35±0.32	89.95±0.06
	SGDClassifier	87.05±0.10	81.28±1.34	54.09±1.76	64.91±0.87	89.78±0.07
	AdaBoost	88.95±0.11	77.03±0.48	71.46±0.90	74.14±0.38	95.36±0.07
	RandomForest	92.85±0.11	83.38±0.54	84.63±0.84	83.99±0.27	97.70±0.05
PCA + UP	Default	77.83±0.05	0.00±0.00	0.00±0.00	0.00±0.00	50.00±0.00
	LinearSVC	88.03±0.10	85.38±0.66	55.51±0.73	67.27±0.42	89.18±0.13
	LDA	86.45±0.09	86.12±0.34	46.37±0.47	60.28±0.37	89.95±0.10
	GradientBoosting	95.40±0.12	89.57±0.45	89.69±0.55	89.63±0.29	98.54±0.11
	LogisticRegression	88.14±0.10	84.35±0.79	57.12±0.63	68.11±0.34	89.15±0.15
	SGDClassifier	88.73±0.36	85.36±0.94	59.39±2.65	70.00±1.59	87.67±0.27
	AdaBoost	90.31±0.14	80.09±0.41	74.92±0.76	77.42±0.38	96.06±0.06
	RandomForest	95.63±0.09	89.17±0.41	91.39±0.49	90.27±0.21	99.07±0.03
PCA + EDA-States	Default	77.83±0.06	0.00±0.00	0.00±0.00	0.00±0.00	50.00±0.00
	LinearSVC	87.68±0.06	75.38±0.58	65.98±1.24	70.36±0.47	89.88±0.08
	LDA	87.94±0.05	73.11±0.36	72.15±0.59	72.63±0.17	89.53±0.09
	GradientBoosting	93.27±0.14	84.86±0.42	84.75±0.63	84.80±0.33	97.45±0.12
	LogisticRegression	87.46±0.07	75.82±0.46	63.77±0.98	69.27±0.41	90.33±0.07
	SGDClassifier	87.96±0.09	72.76±0.37	73.04±1.19	72.89±0.49	87.64±1.02
	AdaBoost	89.71±0.11	79.15±0.42	72.75±0.67	75.81±0.32	95.87±0.06
	RandomForest	93.82±0.11	86.22±0.43	85.82±0.70	86.02±0.28	98.30±0.05
PCA + EDA-States + UP	Default	77.83±0.06	0.00±0.00	0.00±0.00	0.00±0.00	50.00±0.00
	LinearSVC	88.13±0.14	75.53±0.97	68.74±1.22	71.96±0.38	88.96±0.10
	LDA	87.98±0.04	72.96±0.18	72.77±0.22	72.86±0.09	88.71±0.13
	GradientBoosting	95.39±0.11	89.62±0.48	89.61±0.60	89.61±0.25	98.54±0.12
	LogisticRegression	88.55±0.19	77.97±1.04	67.43±1.03	72.30±0.43	89.66±0.09
	SGDClassifier	87.98±0.05	72.76±0.15	73.16±0.35	72.96±0.17	85.38±0.62
	AdaBoost	90.55±0.13	80.80±0.40	75.26±0.77	77.93±0.38	96.27±0.06
	RandomForest	95.67±0.09	89.91±0.40	90.64±0.70	90.27±0.24	99.09±0.03

TABLE C.2. PROFILE A: SEMI-PERSONALIZED STRESS-VS-NON-STRESS MODELS

Experiment	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC AUC (x100)
Only PCA	Default	78.18±0.01	0.0±0.0	0.0±0.0	0.0±0.0	50.0±0.0
	AdaBoost	92.01±0.62	82.48±1.45	80.5±2.24	81.46±1.55	97.44±0.32
	GradientBoosting	96.88±0.38	93.09±0.64	92.58±1.5	92.83±0.92	99.04±0.23
	LDA	86.34±0.45	84.2±1.62	46.06±1.74	59.53±1.64	83.43±0.84
	LinearSVC	86.07±0.37	84.59±1.88	44.25±1.39	58.09±1.31	83.38±0.87
	LogisticRegression	87.16±0.47	83.7±1.56	51.1±1.7	63.44±1.57	83.1±0.88
	RandomForest	98.54±0.28	96.62±0.68	96.69±0.8	96.65±0.65	99.84±0.06
	SGDClassifier	85.49±0.34	90.82±2.93	37.3±1.58	52.85±1.48	82.75±0.89
PCA + EDA-States	Default	78.18±0.01	0.0±0.0	0.0±0.0	0.0±0.0	50.0±0.0
	AdaBoost	91.92±0.41	82.54±1.05	79.89±1.26	81.18±0.97	97.41±0.22
	GradientBoosting	96.95±0.31	93.56±0.81	92.38±0.91	92.96±0.72	98.97±0.33
	LDA	86.65±0.37	83.15±2.49	48.72±0.6	61.43±0.74	86.89±0.47
	LinearSVC	87.53±0.39	85.78±2.15	51.37±0.69	64.25±0.96	87.99±0.44
	LogisticRegression	87.57±0.41	85.63±2.17	51.73±0.74	64.49±1.01	87.99±0.48
	RandomForest	98.6±0.22	96.58±0.51	97.02±0.67	96.8±0.51	99.85±0.06
	SGDClassifier	85.16±0.36	82.23±2.36	40.89±1.44	54.59±1.27	86.14±0.78

TABLE C.3. PROFILE B: SEMI-PERSONALIZED STRESS-VS-NON-STRESS MODELS

Experiment	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC AUC (x100)
Only PCA	Default	78.36±0.02	0.0±0.0	0.0±0.0	0.0±0.0	50.0±0.0
	AdaBoost	99.37±0.14	98.24±0.68	98.88±0.35	98.56±0.31	99.98±0.01
	GradientBoosting	99.8±0.15	99.3±0.47	99.77±0.28	99.53±0.36	99.87±0.15
	LDA	91.66±0.45	96.05±0.93	64.11±2.53	76.86±1.63	99.42±0.13
	LinearSVC	97.46±0.18	95.82±0.83	92.31±1.0	94.02±0.45	99.56±0.11
	LogisticRegression	97.49±0.22	95.59±0.89	92.7±0.93	94.12±0.53	99.47±0.12
	RandomForest	99.78±0.13	99.09±0.57	99.89±0.17	99.49±0.29	100.0±0.0
	SGDClassifier	97.52±0.25	94.89±1.04	93.58±0.95	94.23±0.57	99.34±0.15
PCA + EDA-States	Default	78.36±0.02	0.0±0.0	0.0±0.0	0.0±0.0	50.0±0.0
	AdaBoost	99.38±0.11	98.26±0.37	98.88±0.31	98.57±0.26	99.98±0.01
	GradientBoosting	99.8±0.11	99.33±0.48	99.73±0.25	99.53±0.25	99.95±0.06
	LDA	97.42±0.38	94.4±1.26	93.62±0.93	94.01±0.87	99.64±0.09
	LinearSVC	97.76±0.35	95.36±1.05	94.26±0.94	94.81±0.81	99.72±0.08
	LogisticRegression	97.54±0.36	94.56±1.2	94.05±0.9	94.3±0.83	99.65±0.11
	RandomForest	99.83±0.08	99.4±0.26	99.8±0.34	99.6±0.17	100.0±0.0
	SGDClassifier	97.39±0.39	94.38±1.24	93.53±0.99	93.95±0.9	99.59±0.13

TABLE C.4. PROFILE D: SEMI-PERSONALIZED STRESS-VS-NON-STRESS MODELS

Experiment	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC AUC (x100)
Only PCA	Default	77.1±0.02	0.0±0.0	0.0±0.0	0.0±0.0	50.0±0.0
	AdaBoost	92.9±0.52	83.92±1.6	85.41±0.74	84.65±1.02	97.66±0.26
	GradientBoosting	95.15±5.4	88.38±12.48	94.03±1.05	90.66±8.15	96.55±6.72
	LDA	83.61±0.44	77.12±1.96	40.47±1.43	53.07±1.47	92.06±0.48
	LinearSVC	83.69±0.4	77.9±1.61	40.2±1.72	53.01±1.59	91.68±0.5
	LogisticRegression	84.6±0.43	77.38±1.66	46.3±1.66	57.92±1.45	91.88±0.53
	RandomForest	98.57±0.19	95.57±0.81	98.33±0.32	96.93±0.41	99.88±0.03
	SGDClassifier	83.81±0.38	77.29±1.55	41.52±1.91	53.99±1.63	91.78±0.47
PCA + EDA-States	Default	77.1±0.02	0.0±0.0	0.0±0.0	0.0±0.0	50.0±0.0
	AdaBoost	92.85±0.36	83.77±0.87	85.33±1.18	84.54±0.81	97.67±0.15
	GradientBoosting	97.35±0.23	93.69±0.75	94.82±0.56	94.25±0.5	98.96±0.15
	LDA	86.04±0.49	69.47±1.33	69.72±1.7	69.58±1.07	92.57±0.44
	LinearSVC	86.1±0.52	70.05±1.44	68.72±1.82	69.36±1.18	92.43±0.46
	LogisticRegression	86.08±0.57	70.03±1.53	68.66±1.86	69.32±1.28	92.5±0.48
	RandomForest	98.75±0.25	96.23±0.93	98.42±0.45	97.31±0.54	99.89±0.04
	SGDClassifier	86.11±0.55	69.32±1.3	70.59±1.93	69.94±1.27	90.77±3.54