

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

# AUTOMOTIVE: A case study on AUTOMATIC multiMODal drowsiness detection for smart VEHICLES

TELMA ESTEVES<sup>1,2,3</sup>, JOÃO RIBEIRO PINTO<sup>1,4</sup>, PEDRO M. FERREIRA<sup>1,4</sup>, PEDRO AMARO COSTA<sup>3</sup>, LOURENÇO ABRUNHOSA RODRIGUES<sup>3</sup>, INÉS ANTUNES<sup>1,4</sup>, GABRIEL LOPES<sup>1,4</sup>, PEDRO GAMITO<sup>5</sup>, ARNALDO J. ABRANTES<sup>6</sup>, PEDRO M. JORGE<sup>6</sup>, ANDRÉ LOURENÇO<sup>3,6</sup>, ANA F. SEQUEIRA<sup>1</sup>, JAIME S. CARDOSO<sup>1,4</sup>, AND ANA REBELO<sup>1</sup>

<sup>1</sup>Centre for Telecommunications and Multimedia, INESC TEC, Porto, Portugal

<sup>2</sup>Universidade Nova de Lisboa, Lisbon, Portugal

<sup>3</sup>CardioID Technologies, Lisbon, Portugal

<sup>4</sup>Faculdade de Engenharia, Universidade do Porto, Porto, Portugal

<sup>5</sup>Universidade Lusófona de Humanidades e Tecnologias (ULHT), HEI-Lab, Lisbon, Portugal

<sup>6</sup>Instituto Superior de Engenharia de Lisboa, Lisbon, Portugal

Corresponding authors: Ana F. Sequeira (e-mail: ana.f.sequeira@inesctec.pt) and André Lourenço (e-mail: andre.lourenco@isel.pt).

**ABSTRACT** As technology and artificial intelligence conquer a place under the spotlight in the automotive world, driver drowsiness monitoring systems have sparked much interest as a way to increase safety and avoid sleepiness-related accidents. Such technologies, however, stumble upon the observation that each driver presents a distinct set of behavioral and physiological manifestations of drowsiness, thus rendering its objective assessment a non-trivial process. The AUTOMOTIVE project studied the application of signal processing and machine learning techniques for driver-specific drowsiness detection in smart vehicles, enabled by immersive driving simulators. More broadly, comprehensive research on biometrics using the electrocardiogram (ECG) and face enables the continuous learning of subject-specific models of drowsiness for more efficient monitoring. This paper aims to offer a holistic but comprehensive view of the research and development work conducted for the AUTOMOTIVE project across the various addressed topics and how it ultimately brings us closer to the target of improved driver drowsiness monitoring.

**INDEX TERMS** Biometrics, biosignals, computer vision, data, driver, drowsiness, simulator, vehicles.

## I. INTRODUCTION

**D**RIVER status monitoring (DSM) systems have emerged as an innovative solution to prevent drowsiness and fatigue-related traffic accidents. Some major automotive manufacturers, such as Ford, Toyota, BMW, and Nissan, have been developing DSM systems since the 2000s [1], making use of visual, vehicle, and physiological measurements to monitor and evaluate the state of the driver and triggering automatic alerts if drowsiness is detected. By using systems with a half-second warning time, one can expect an estimated 60% decrease in the number of accidents, whereas an extra second can prevent up to 90% of collisions [2].

To develop such systems, since real on-road drowsiness acquisitions are generally unsafe, researchers often rely on simulated environments. However, drivers' behavior in such scenarios is unrealistic since they do not perceive risk in a similar way, resulting in discrepancies between the observed

behavioral patterns and the expected real-world observations.

Emulating on-road experiences more faithfully during simulation-based acquisitions has resulted in improved data. The collection of biological signals has had a similar effect [3], [4]. Several simulation-based data collection projects have been recently conducted, including naturalistic driving studies and field operational tests [5]. However, even when considering these advances, the disparities between naturalistic data and simulator-based acquisitions are still a relevant challenge.

One other significant hurdle to the reliability of developed products in real-life conditions is the high variability of behavioral patterns across drivers. A drowsiness monitoring system can present acceptable average accuracy levels for a given set of drivers and, simultaneously, be inadequate at recognizing the specific sleepiness patterns of other drivers. This could be improved by taking advantage of biometric

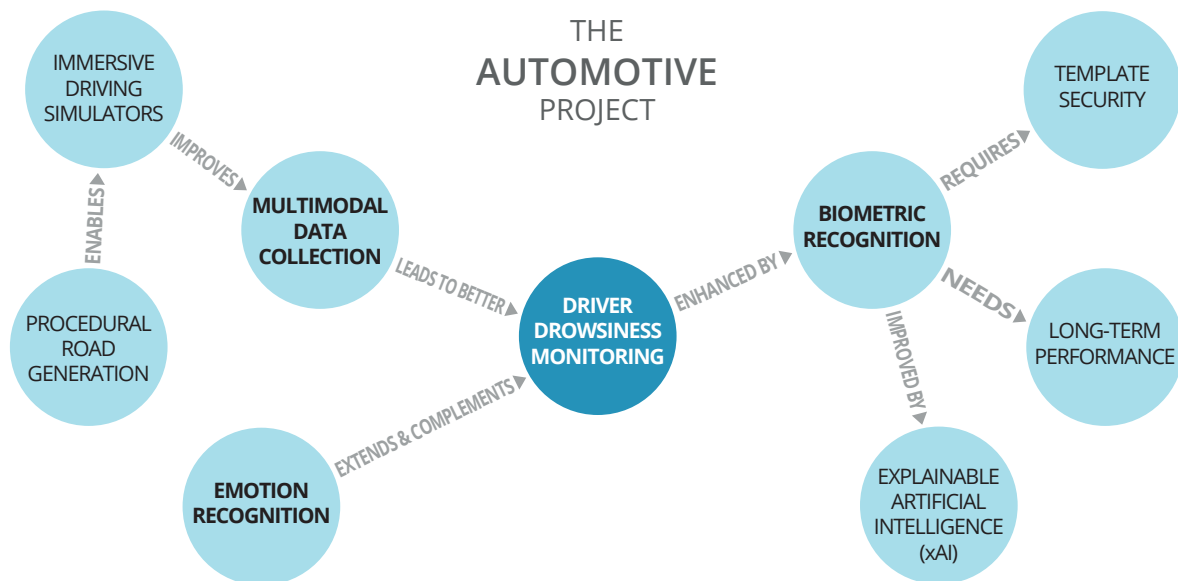


FIGURE 1: Holistic view of the AUTOMOTIVE Project, showing the links between the explored topics and how they ultimately contribute for the central goal of improved driver drowsiness monitoring.

recognition solutions which, by delivering continuous identity predictions, would enable the continuous learning of user-specific drowsiness models using the data acquired from each driver. Although major automotive brands have been introducing biometric recognition technology solutions, the use of biometrics for continuous learning of driver-specific models is still an open topic.

For both drowsiness monitoring and biometric recognition in driving scenarios (as well as several related research topics, *e. g.*, emotion recognition) two human characteristics stand out: the face and the electrocardiogram (ECG). The face is the focus of most research, since it can be acquired with inexpensive cameras, carries drowsiness information, and is universally accepted as one of the most effective and intuitive approaches for biometric recognition. However, its performance is known to suffer significantly when the quality of the acquisition is compromised by varying illumination, occlusions, movement, among other factors [6], which should be expected when monitoring a driver inside a moving vehicle.

On the other hand, the ECG is generally robust against illumination and visibility factors that would affect face acquisitions. Research on ECG biometrics has quickly been gaining traction due to its universality, ease of acquisition and processing, difficult counterfeiting, and reliable distinctiveness [7]. As a physiological signal, the ECG also varies significantly with psychophysiological states, thus carrying information on wellbeing factors such as sleepiness, stress, or emotions. The need for continuous contact is a considerable inconvenience, but this can be effectively minimized through modern off-the-person acquisition techniques (such as the CardioWheel [8] or the Nyami Band [9]).

With this, the complementary nature of the ECG and face

data for driving scenarios seems obvious. A multimodal solution would combine the benefits observed in recent years with deep learning approaches for face recognition in highly unconstrained scenarios [10], [11] with the meaningful strides verified in ECG biometrics [12]. Its integration in a reliable, robust, and personalized alert system combining multimodal biometric recognition and drowsiness detection represents a sizable challenge. The main hurdle concerns the processing and classification of real data by algorithms developed on simulated data. Nevertheless, this is still the most promising path towards the next generation of intelligent recognition of driver drowsiness.

This was the central goal of the AUTOMOTIVE project. To advance the state-of-the-art algorithms regarding behavioral variability patterns and acquisition quality, while dealing with the domain mismatch between simulated scenarios and real-world setups. The AUTOMOTIVE project was conducted in Portugal and led by the Institute for Systems and Computer Engineering, Technology and Science (INESC TEC), with the participation of the company CardioID Technologies, the Instituto Superior de Engenharia de Lisboa (ISEL), and the Universidade Lusófona de Humanidades e Tecnologias (ULHT).

This paper presents the main achievements and highlights of the AUTOMOTIVE project. This entails the description of the meaningful strides achieved in its diverse research topics (see Fig. 1) and how they contribute to the central goal of advancing the state-of-the-art in driver drowsiness monitoring. Beyond this introduction, this paper covers the project's work on driving simulators and data collection, in section II; the pattern recognition methodologies for biometrics, emotion recognition, and drowsiness monitoring, in section III; and the general conclusions drawn from the

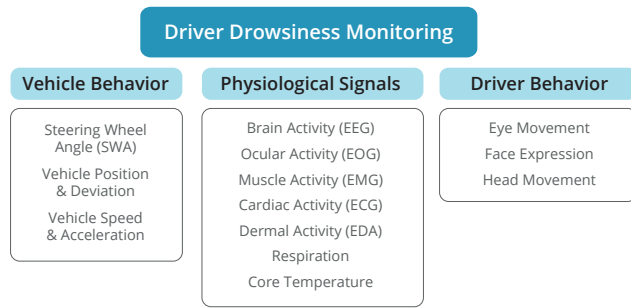


FIGURE 2: Summary of the diverse measurement approaches used for driver drowsiness detection (adapted from [13]).

AUTOMOTIVE project, in section IV.

## II. DATA ACQUISITION

Driving research is a broad field with numerous projects typically carried either in simulated or naturalistic environments. Virtual driving environments stand out due to the minimal risk to the volunteers during the experiments and the possibility of fully controlling vehicle physics and traffic complexity. As such, whether by collecting large-scale data and training deep neural networks without any logistic restraints or by emulating exact traffic circumstances to validate trained models in extreme scenarios, using simulator-based development has been a key contributor to the great recent strides in autonomous driving.

This section presents some driving simulation environments currently publicly available or provided upon request for research purposes:

- **Microsoft AirSim**<sup>1</sup> [14] was specifically designed as a testing platform for artificial intelligence (AI) experiments. This platform offers an application programming interface (API) for acquiring sensor, telemetry, and image data from large-scale urban and natural environments. Its cameras have photo-realistic lighting, real-time depth view, and object segmentation. The control of the vehicle is based on the Unreal physics engine;
- **CARLA**<sup>1</sup> [15] enables complete programmatic control over the simulation. This platform includes an autonomous driving sensor suite with configurable sensors and ground-truth data, emulating a virtual city with dynamic pedestrians and vehicles. This simulator extends a C++/Python API and manages remote procedure call (RPC) communication through an event-based server-client model;
- **Deepdrive Voyage**<sup>2</sup> offers a TensorFlow engine for deep reinforcement learning while providing extensive hours of prerecorded training from a self-driving competition between users performing complex maneuvers;

<sup>1</sup>Among the simulators, Microsoft AirSim and CARLA are open-source.

<sup>2</sup>GitHub: deepdrive/deepdrive, available on <https://github.com/deepdrive/deepdrive>, more info at <https://deepdrive.io/index.html>.

TABLE 1: Main databases currently available for research in biometrics and drowsiness monitoring with ECG and face video (those used in the AUTOMOTIVE project are highlighted in *italics*).

Database	Data Type	Access
<b>DROZY, The ULG Multimodality Drowsiness Database [18]</b>	<b>KSS, ECG, Face Video</b>	<b>Public</b>
<b>UTA Real-Life Drowsiness Dataset [19]</b>	<b>KSS, Face Video</b>	<b>Upon Request</b>
<i>Swedish National Road and Transport Research Institute [20]</i>	<i>KSS, ECG</i>	<i>Upon Request</i>
<b>Stress Recognition in Automobile Drivers [21]</b>	<b>ECG</b>	<b>Public</b>
<i>UofTDB [22]</i>	<i>ECG</i>	<i>Upon Request</i>
<i>PTB [23]</i>	<i>ECG</i>	<i>Public</i>
<i>CYBHi [24]</i>	<i>ECG</i>	<i>Public</i>
<i>E-HOL [25]</i>	<i>ECG</i>	<i>Public</i>
<i>Youtube Faces Database [26]</i>	<i>Face Video</i>	<i>Public</i>
<b>NTHU Driver Drowsiness Detection Dataset [27]</b>	<b>Face Video</b>	<b>Upon Request</b>

- **NVIDIA's**<sup>3</sup> platform recreates actual locations with very high levels of detail, realism, and complexity, thus massively extending the development and validation of deep learning models [16]. It has a first server that renders the environment and generates simulated sensor data and a second with the prototype self-driving hardware engine that processes data as if deployed in-vehicle.

Regarding the data that can be acquired in driving research, either through simulated or naturalistic environments, the most commonly used in-vehicle measurement sensors for detecting driver drowsiness can be categorized into vehicle behavior, driver behavior, and physiological signals (according to [17]). Fig. 2 organizes the types of information acquired through these three forms of sensing. Some data is publicly available (or provided upon request for research purposes) as part of the databases presented in Table 1, where those used in the current project are highlighted in bold.

One of AUTOMOTIVE's objectives was to develop a platform to simulate a real driving experience and effectively collect multimodal data in real-time. Hence, two simulator platforms were developed, where the main observable difference between the two relies on the immersiveness of the environment. The first simulator, AUTOMOTIVE DD (Desk-Driver) Simulator, runs on a simple desktop monitor with the driver seated on a chair. The second, AUTOMOTIVE

<sup>3</sup>NVIDIA DRIVE Sim. Available on <https://developer.nvidia.com/drive/drive-sim>.

IHC (In-Half-Car) Simulator, includes a more realistic virtual driving environment, projected onto wide screens surrounding the driver. For added realism, the driver is behind the wheel of a real (parked) car cut in half.

Both simulator platforms integrate different sensory information from different sources, namely:

- The electrocardiogram (ECG), acquired with the CardioWheel, a machine learning solution that has sensors embedded in the steering wheel and recognizes the driver's identity based on ECG [8]. Several heart metrics features are automatically extracted, to be used in the current work;
- Face video, acquired with the Intel RealSense™ SR300 camera, which simultaneously captures RGB, depth, and infrared video channels. This camera's software development kit automatically detects 78 face landmarks [28];
- Driving style events, *i. e.* dynamic signals such as acceleration and braking, are obtained directly from the simulator [28];
- Performance telemetry in simulated urban and highway environments, *i. e.* information such as speed control, lane following, acceleration, and braking inputs, are obtained using the Mobileye™ collision-avoidance system [28].

The designed experimental protocol for the acquisition is the same for both simulators and is presented below.

### A. EXPERIMENTAL PROTOCOL

The data acquisition protocol was influenced by the work of Naurois *et al.* [29]. Its design addresses the need for robust and complete data. Hence, we collect data regarding vehicle movement (steering wheel angle, vehicle speed, acceleration, and braking), the driver's behavior (facial expressions and features), physiology (their electrocardiogram), and their sleepiness state, as well as additional information that can be useful to further understand the subjects and their driving experience.

Before the acquisition, we check if the volunteer has a valid driver's license, is not diagnosed with epilepsy, and is not susceptible to kinocytosis (the short form of the Motion Sickness Susceptibility Questionnaire [30] is applied for this purpose). Furthermore, their susceptibility to sleepiness is assessed by applying the Epworth Sleepiness Scale<sup>®</sup> [31], and the Horne and Ostberg morningness-eveningness questionnaire is filled to determine the volunteer's circadian rhythm relation to the different times of the day. Other information is collected, including the subject's age, sex, self-assessed sleep quality, coffee consumption, comfort with the driving task, yearly driving frequency, general daily schedule, diagnosed sleep conditions, and use of medication or devices that may influence their natural heart rate.

On the day of the acquisition, the subject is asked to not drink caffeinated nor alcoholic beverages and to sleep six to nine hours the previous night. If possible, the acquisition

is scheduled after lunchtime, since the probability of falling asleep in this period increases three-fold [32]. Just before the start of the data collection, the volunteer has a short test drive for as long as they need, to acclimate to the particular driving experience in the simulator. This initial preparation period aims to minimize the influence of the adaptation to the simulator in the collected data.

After this initial period, the subject drives a car out of an urban area and into a highway, taking approximately 2 minutes. They then drive on the aforementioned highway for approximately sixty minutes before arriving at their destination, another urban area where the drive ends. During this whole process, the driver is periodically asked (every 5 to 15 minutes) for their self-assessed score on the Karolinska Sleepiness Scale (KSS), on a range from 1 (the highest level of alertness) to 9 (the highest level of sleepiness, where the subject is combating falling asleep). After the acquisition, the general well-being of the subject is assessed (looking for signs of kinocytosis), and they are asked to fill a final questionnaire including the System Usability Scale.

### B. AUTOMOTIVE DD SIMULATOR

#### 1) Setup

This simulator's graphics and physics are supported by Unity 3D, a rendering and physics engine that has gained widespread adoption in gaming and research areas, and is featured in several state-of-the-art driving simulators. It is a general-purpose platform that enables fast development and prototyping of simulated environments while allowing for flexible scripting control of all virtual environment elements with C# or Javascript.

The designed virtual environment is composed of two urban areas connected by a highway. The urban areas provide a set of complex road geometries and interaction opportunities with a variety of agents (*e. g.*, pedestrians, traffic signs, crosswalks, intersections, buildings). Meanwhile, the highway gives this environment a flexible component, enabling experiments of customizable length and duration, as well as varying road monotony. The procedural generation of this highway road is what creates this flexibility. It was implemented by adapting the MicroGSD RoadArchitect<sup>4</sup> module, which generates a static road geometry from a spline defined by user-placed nodes. The adaptation eliminated direct user inputs on road geometry and instead parametrized node location as a random distribution of 3D points. Such parameters are:

- Segment size, the longitudinal distance between consecutive nodes;
- Number of nodes, which together define the total highway length;
- Number of lanes, that can range from single to three-lane;

<sup>4</sup>MicroGSD: Road Architect for Unity. Available on <https://github.com/MicroGSD/RoadArchitect>.





FIGURE 3: Running simulation in the AUTOMOTIVE DD Simulator: view from inside the vehicle.

- Transverse position and height range: from these, a coordinate is randomly chosen, hence defining curve and elevation variations along the highway.

By defining these parameters, a highway is generated for every session, allowing us to control the experiment duration and the complexity of the driving task and ensuring that infinite road geometries with equivalent difficulties are available. Furthermore, having the spline definition of road trajectory makes it possible to define a set of non-playable character (NPC) vehicles that travel the highway alongside the user. Generating cars that follow the defined splines and whose speeds are drawn from random distributions makes it possible to simulate and adjust traffic dynamics for a more realistic experience.

An example of the running simulation can be seen in Fig. 3. The multimodal nature of this setup requires it to communicate with several sensors, each collecting data at different rates. The setup incorporates state-of-the-art sensor devices such as the CardioWheel, acquiring ECG at 1 kHz [8], the PulseOn wrist-band, acquiring optical heart rate signals at 25 Hz [33], [34], the MoveSense<sup>5</sup> chest-band, acquiring clinical-grade heart rate at 512 Hz, the Mobileye Connect<sup>6</sup> collision-avoidance system, and the Intel RealSense<sup>7</sup> camera, performing RGB-depth facial analysis at 50 Hz. Furthermore, the simulated environment processes the data provided by these sensors and is also capable of producing performance telemetry as control of speed limits, lane following angle, obstacles, collision events, traffic violations, and acceleration/braking inputs.

However, managing multiple high-throughput low-latency data sources with precise sampling rates requires a performance-focused architecture. This type of system is usually set with a central queuing structure, where inputs

<sup>5</sup>MOVESENSE medical sensor. Available on <https://www.movesense.com/wp-content/uploads/2020/12/Movesense-Medical-Spec-Sheet-2020-12-v1.0.pdf>.

<sup>6</sup>MOBILEYE: Technical Specification Sheet. Available on <https://fleetsafe.com.au/wp-content/uploads/2019/06/Mobileye-6-Technical-Spec-v0.2-1.pdf>.

<sup>7</sup>Intel® RealSense™ Computer Vision: Depth and Tracking cameras. Available on <https://www.intelrealsense.com/>.

and outputs are integrated - a classic example of the producer/consumer model, typically implemented using Message Queuing Telemetry Transport (MQTT) or Advanced Message Queuing Protocol (AMQP).

Specifically designed for handling multiple real-time intensive data streams, Apache Kafka is the proposed solution as an integration middleware platform. The core simplicity of this system favors performance and allows for massive scalability: a server (zookeeper) receives data from several producers and allows it to be consumed from a transaction log distributed into multiple nodes (brokers). By replicating messages through all brokers, Apache Kafka guarantees performance and redundancy: if a node fails, consumption is resumed at the next available broker. Typical use cases are streaming services or telemetry analytics.

In terms of performance, changing the batch queue size enables reaching a balance between transmission latency and overall throughput, as higher throughput also implies higher latency between production and consumption due to batch buffering (see Table 2). With the Unity driving simulator as API for production/consumption, and reducing latency to 20 ms, the platform still manages approximately five thousand messages per second, *i. e.*, sample rates of 5 kHz in simultaneous parallel streams directly feeding the rendering simulation. If larger messages are transmitted, both the average latency and throughput drop significantly (see Table 3), though the average bit rates can rise to 1 Gbit/s depending on batch queue size.

To test the capabilities of this middleware platform, a benchmarking test was performed at localhost with an i7-8565U CPU, 16 GB DDR4 SDRAM, and a 256GB PCI NVM SSD. Two sample messages of 1 kB and 1 MB were used to evaluate performance at different batch sizes on a two-broker configuration by sending three thousand identical messages in each run. Timestamps were registered at production and consumption for each message, then averaged at test completion for average latency. Elapsed times for each test offer average throughput (in messages per second). The average bit rate converts average throughput from messages per second to bits per second. Linger time for Apache Kafka broker is set at 1 ms.

Apache Kafka's low latency is paramount for a real-time streaming platform, specifically for biometric systems or obstacle detection alerts, both of which depend on millisecond-level response times to be effective. Benchmark results (Table 2 and Table 3) confirm that this platform can process several thousand messages per second at millisecond transmission latency. Overall, these are significantly faster than regular socket communication and hence fulfill the requirements of a driving environment handling real-time analysis of multiple sources with variable sampling intervals.

As a robust but loosely coupled integration architecture, this platform can also integrate several different producing and consuming modules, including adapting to an actual driving environment (which can replace this simulator with no further changes required).

TABLE 2: Apache Kafka metrics for 1 kB sample messages.

Batch Queue (messages)	Average Latency (s)	Average Throughput (message/s)	Average Bitrate (Mbit/s)
1	0.011	1157	9.256
10	0.021	5114	40.912
100	1.965	12851	102.808
1000	2.544	19991	159.928

TABLE 3: Apache Kafka metrics for 1 MB sample messages.

Batch Queue (messages)	Average Latency (s)	Average Throughput (message/s)	Average Bitrate (Mbit/s)
1	9.239	78	624
10	9.724	84	672
100	11.836	95	760
1000	22.629	145	1160

The resulting simulation is displayed on an ASUS Monitor (model VS197DE TN 18.5" 16:9 60Hz FWXGA), which stands in front of the driver. The steering wheel is a Logitech G Dual-Motor Feedback Driving Force G29 Gaming Racing Wheel with Responsive Pedals. A photograph of the platform being used is shown in Fig. 4.

## 2) Collected Data

A different experimental protocol than the one previously presented was designed to compare the heart rate variability (HRV) features derived from the cardiac data collected using the Cardiwheel, PulseOn, and MoveSense sensors. Results confirming the equivalency of such features during driving tasks would promote the introduction of physiological insight in drowsiness detection systems, as that would imply that off-the-person in-vehicle systems (CardioWheel) [8] are effectively interchangeable with on-the-person systems (MoveSense and PulseOn).

This particular experiment involved thirteen volunteers, each participating in two half-hour drives on the AUTOMOTIVE DD Simulator. Participants would rate their drowsiness level every five minutes using the KSS. Session scheduling intended to cover a variety of alertness states, with each participant having a session in the middle of the morning and another in the late afternoon. For each of these, circadian settings in the simulator were tuned to potentiate alertness (using daylight in morning sessions) or drowsiness (night-light in late afternoon sessions).

## 3) Data Applications

Using the collected data, the correlation between inter-beat intervals (IBIs) from each source was computed, as well as the similarity of the resulting HRV, evaluating time, frequency, and non-linear domain features. The cardiac signals retrieved from each sensor were synchronized in each session. From R-peak locations detected using the Pan-Tompkins algorithm, series of IBIs were computed and afterward corrected using the algorithm described in [35].

Direct comparison of IBI sequences showed a satisfactory level of equivalency between all data sources, although

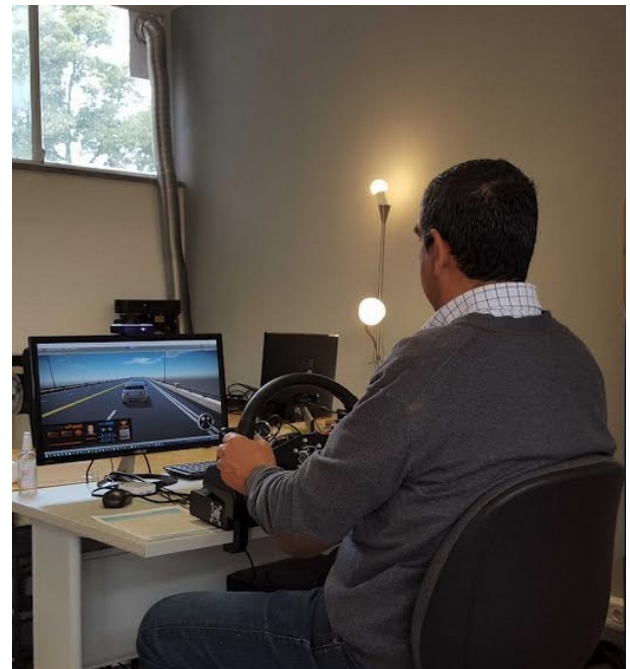


FIGURE 4: Experimental setup of the AUTOMOTIVE DD Simulator during use. In the monitor, the simulation is being displayed in the view from outside the vehicle.

the PPG-based IBIs differ more consistently from the other sources. This disparity is mitigated when IBI sequences are transformed into sets of HRV features. This convergence is justified since compressing a sequence of IBIs to a single HRV feature loses the vascular modulation of the rhythm measured by the wrist PPG sensor, but maintains the significant contribution of heart compression measured with the cardiac sensors.

These results suggest the feasibility of flexible driver drowsiness detection systems, using any type of cardiac rhythm sensor to assess the driver's state. The possibility of such a flexible system can accelerate the implementation of driver monitoring systems based on cardiac features and increase the functionality of advanced driver-assistance systems (ADAS) by endowing them with prevention and adjustment capabilities.

## C. AUTOMOTIVE IHC SIMULATOR

### 1) Setup

Based on the Apache Kafka platform described in AUTOMOTIVE DD Simulator, the functionalities of acquisition and storage of the physiologic data, as well as the biometric recording, user identification, and authentication, are available for the integration of multiple simulation environments. As such, a simulator developed externally to the current project by the ULHT partner was integrated to create a more immersive simulation.

The ULHT simulator offers several driving scenarios for greater diversity during the driving sessions. These include

highways, suburban roads, and urban scenarios. Moreover, it includes the procedural generation of traffic events for the analysis of the driver's response. Accessing the database, the event manager can provide event timestamps, allowing the extraction of the physiological data that describes the driver's reaction to the corresponding events and, ultimately, their driving performance.

The order of event occurrences must ensure some novelty across different sessions for the same user. As the route mandates a logical sequence in the highway and suburban road environments, these driving sessions are expected to include all planned events. On the other hand, to ensure a consistent experience across all users, the vehicle has a GPS available in the driving interface, indicating the route to follow in the urban environment. This environment also stores several invisible checkpoints in particular spots, such as road intersections or avenues, which trigger events according to the drivers' proximity.

Some implemented suburban event scenarios are:

- Moving emergency vehicles;
- Stopped vehicles (including after accidents);
- Vehicles breaking road laws;
- Weather such as rain and fog.

Some urban event scenarios available in the simulator are:

- Pedestrians crossing the road (on crosswalks or not);
- Cyclists on the road;
- Traffic light intersections.

Summing up, with the integration of the ULHT simulator, an automated event generator enables the simulation of more realistic driving sessions (for example, commuting between urban and suburban environments with regular traffic events). An example of the running simulation can be seen in Fig. 5. The resulting simulation was designed to be projected onto an angled display surrounding the driver, who is sitting on a real (parked) car with an integrated steering wheel and pedals. An illustration of the planned setup can be seen in Fig. 6.

### III. DEVELOPED ALGORITHMS

The following subsections describe several proposed methods and comparison studies carried out over the last years of the AUTOMOTIVE project. Some of the challenges and current state-of-the-art related work were already discussed in [12]. This reference was the guideline for our published research in electrocardiogram (ECG) presented below.

#### A. ECG BIOMETRICS

A biometric system aims to either identify or verify the identity of a person based on a measurement of one or multiple biometric traits [12]. Usually, it is composed of the following modules: acquisition, quality assessment, feature extraction, storage, and decision (see Fig. 7).

The biometric algorithm uses the data from the acquisition and storage modules and performs quality assessment, feature extraction, and decision. The data is commonly fingerprint, iris, palmprint, or face images, but can also be medical



(a) Running simulation on the Highway scenario in the AUTOMOTIVE IHC Simulator. Adverse weather conditions are active (rain).



(b) Running simulation on the Suburban Road scenario in the AUTOMOTIVE IHC Simulator.



(c) Running simulation on the Urban scenario in the AUTOMOTIVE IHC Simulator. Traffic lights, a pawn on the crosswalk, and a GPS indicating the route to follow can be seen.

FIGURE 5: Running simulation on different scenarios in the AUTOMOTIVE IHC Simulator. Views from inside the vehicle.

biometric measurements, such as the electroencephalogram (EEG) or the electrocardiogram (ECG). The use of the ECG as a biometric trait is relatively recent but has been gaining momentum, as researchers explore deep learning methodologies applied to this signal and novel ECG acquisition setups enable comfortable signal collection from daily objects [12], [36].

In normal conditions, an ECG signal is a cyclic repetition of five easily recognizable deflections: the P, Q, R, S, and T waves (see Fig. 8). The ECG can be used to discriminate between subjects in biometric recognition since it varies according to the following inter-subject factors:

- Heart Geometry such as heart size or cardiac muscle thickness affects the depolarization of the heart;



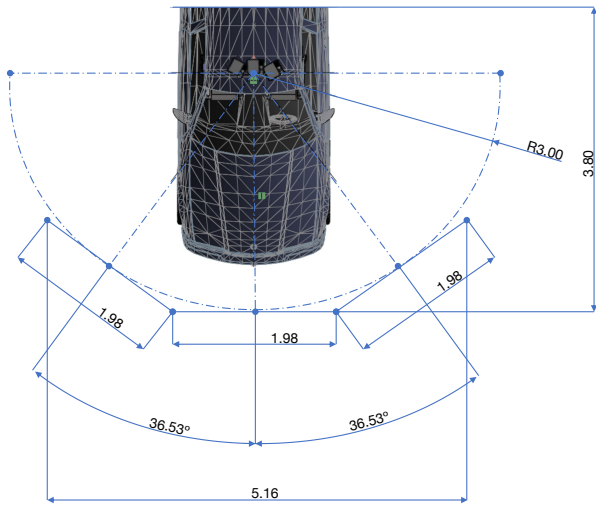


FIGURE 6: Design of the experimental setup of AUTOMOTIVE IHC Simulator. 3 projectors are displayed on top of the car. The throw distances of these projectors are represented.

- Individual Attributes such as age, weight, or pregnancy, shifts the orientation of the electrical current conduction vectors across the heart.

Nevertheless, intra-subject variability factors, such as physical exercise or meditation, cardiac conditions, posture, emotions, fatigue, and electrode characteristics and placement, should not be overlooked since they may undermine the process of biometric recognition [12].

The literature survey in [12] contributed towards the AUTOMOTIVE project as follows:

- **History of ECG biometrics:** a deep survey with the evolution and current landscape of ECG-based biometric recognition based on the review of ninety-three state-of-the-art publications;
- **Fundamental background knowledge:** a solid overview of fundamental concepts (such as anatomy, physiology, and intra-subject and inter-subject variability), providing a comprehensive guide to new and current researchers;
- **Future research paths:** a discussion of the most relevant challenges and the most promising future possibilities regarding research and development in each part of ECG biometric systems, from acquisition to decision.

For the research topic of drowsiness monitoring, knowledge from ECG biometrics is extremely relevant. One of the major data sources for driver drowsiness recognition is the ECG signal, and the topic of ECG biometrics has been cultivating deeper insights that could benefit AUTOMOTIVE’s target task. Beyond this, ECG biometric algorithms can offer identity information required for continuous user-tuning drowsiness monitoring algorithms and, ultimately, improve their performance.

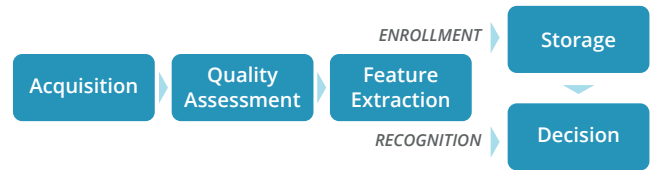


FIGURE 7: General structure of a biometric system (adapted from [12]).

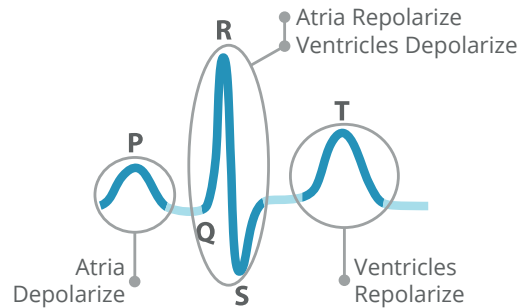


FIGURE 8: The heartbeat, its characteristic waveforms, and their relationship with cardiac cycle polarization events on the heart’s atria and ventricles (adapted from [12]).

The following sections delve deeper into the research work conducted on ECG biometrics for the AUTOMOTIVE project, inspired by the conclusions drawn by the literature survey presented in [12].

1) End-to-end deep learning models

Addressing the lack of end-to-end deep learning approaches for ECG biometrics, a simple method was proposed for both identification [37] and identity verification [38] tasks. The goal was to fully take advantage of the potential and flexibility of deep learning to integrate the whole ECG biometrics pipeline into a single model. Such a model receives all the information carried by the raw signals and is optimized as a whole to freely choose which data is most useful for accurate and robust decisions.

**Proposed model:** Inspired by the typical structure of a convolutional neural network, the model is composed of two parts: one for feature extraction followed by one for decision (see Fig. 9). The feature extraction part is composed of four convolutional layers (with, respectively 24, 24, 36, and 36  $1 \times 5$  filters), interleaved with  $1 \times 5$  max-pooling layers. The decision part of the model is composed of a fully connected layer. For identification (in [37]), this last layer has one neuron for each identity and softmax activation. It is trained to offer probabilities for each identity. For identity verification (in [38]), the  $n$ -dimensional embeddings output by the layer (with ReLU activation) are processed representations of the input samples. The task of finding identity matches is performed through their similarity to other embeddings.

**Identification Experiments and Results:** The work conducted on [37] and [38] used mainly data from the University



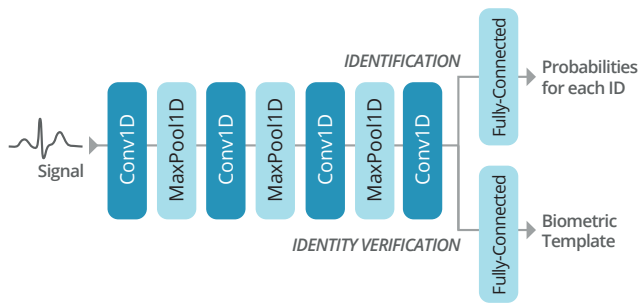


FIGURE 9: Architecture of the proposed model for ECG identification and identity verification.

TABLE 4: Identification accuracy results obtained for the explored experimental setups (\* denotes results with data augmentation based on random permutations).

Experimental Setup	Identification Accuracy (%)
A (traditional pipeline)	90.6
B	93.1
C	92.0 (94.2*)
D (fully end-to-end)	92.7 (96.1*)

of Toronto ECG database (UofTDB) [22]. As the current major off-the-person database, with recordings from 1019 subjects over multiple sessions and postures, it offers an optimal setup to develop and evaluate robust ECG biometric algorithms.

Firstly, for identification [37], a study was conducted on the successive integration of processing stages into the deep learning model. Let's consider the traditional ECG biometric algorithm pipeline divided into four stages: (1) denoising; (2) preparation; (3) feature extraction; and (4) decision. This study encompassed four experimental setups: (A) stages 1 + 2 + 3 + deep model; (B) stages 1 + 2 + deep model; (C) stage 1 + deep model; and (D) only with the deep model. This amounted to a progressive evolution from the traditional ECG pipeline to a fully end-to-end model and enabled the assessment of the benefits of using the latter.

The results (see Table 4) show that the traditional pipeline results in the poorest accuracy. Setup B, with the deep model performing just feature extraction and decision, corresponds to the best results, followed closely by the fully end-to-end model. Nevertheless, setups C and D can be improved with data augmentation strategies. Data augmentation is generally essential for avoiding overfitting deep learning models [39]. However, current strategies are highly specific for image-based tasks. As such, data augmentation strategies were specifically designed for ECG biometrics, aiming to mimic common noise and distortions verified in realistic ECG signals. Out of seven types of data augmentation, four offered improved performance: random permutations, magnitude scaling, baseline wander, and flip. Random permutations offered the largest improvements, raising the accuracy in setups C and D to 94.2% and 96.1%, respectively.

**Identity Verification Experiments and Results:** For

TABLE 5: Identity verification equal error rate (EER) results (%) for the methodology proposed in [38], when compared with state-of-the-art alternatives

Method	Enrollment duration			
	5 s	10 s	15 s	30 s
Identification training	13.70	10.92	9.52	7.86
Triplet loss training	13.93	11.89	10.90	9.94
AC/LDA [41]	30.27	17.90	16.55	15.82
Autoencoder [42]	21.82	19.68	18.84	17.09
DCT [37], [43]	23.05	20.41	18.55	17.38

identity verification [38], two training strategies were explored: using the aforementioned identification training, and using the triplet loss [40]. Several experiments were conducted for a thorough and realistic evaluation of the performance of the methods: using separate sets of subjects for training and testing, shorter enrollment durations, and cross-database tests on the PTB [23] and CYBHi [24] databases (with less and more noisy signals, respectively).

The verification results (see Table 5) show that identification training surpasses the triplet loss training, especially when enrollment data is scarcer. Overall, the proposed method was able to achieve equal error rates (EER) as low as 7.86% for 30 seconds of enrollment. The proposed method performed significantly better than alternative state-of-the-art approaches based on autocorrelation (AC/LDA) [41], autoencoders [42], and discrete cosine transforms (DCT) [37], [43] when evaluated in the same conditions. Moreover, observing the results on PTB and CYBHi (see [38]), one can conclude the proposed method is especially promising for more realistic off-the-person data. Overall, combining all typical pipeline processes into a single end-to-end model seems to be the key for robust ECG biometric models in real applications.

The main contributions of this work for the AUTOMOTIVE project can be summarized as follows:

- **End-to-end model:** An end-to-end architecture was developed to perform biometric identification with ECG signals. Despite being based on deep learning, its relatively simple and lightweight structure, requiring no pre-processing stages beyond signal normalization, paves the way towards future deployment into embedded systems in real scenarios, as foreseen by the AUTOMOTIVE project;
- **Data augmentation for ECG signals:** The proposed data augmentation strategies, specifically tailored for ECG signals, enable us to take full advantage of available data for more accurate and robust algorithms;
- **Realistic performance evaluation setup:** The evaluation with off-the-person data and the careful division of training, enrollment, and testing data results in a challenging and thoroughly realistic evaluation setup that not only shows the improvements *vs.* the state-of-the-art but also illustrates faithfully how the proposed method would behave in real AUTOMOTIVE applications.

The end-to-end, robust, and lightweight model developed in this work also provide a strong scaffolding for the future

development of personalized drowsiness monitoring. The availability of identity labels for new ECG data is essential to teach general drowsiness monitoring models to recognize subject-specific markers of drowsiness.

## 2) Long-term performance

The performance of biometric systems is known to decay over time, eventually rendering them ineffective. Based on previous studies on long-term ECG permanence [44] and the prior knowledge of ECG variability [45], it is expected that long-term signal variations will have a large effect on real ECG biometric applications. The work described in this section, presented in further detail in [46], aimed to study ECG variability over time and its real impact on the performance of state-of-the-art algorithms. Then, template and model update strategies were implemented to bridge the observed performance gap.

**Identification Algorithms:** Four state-of-the-art ECG-based identification methods were implemented to serve as foundation for the long-term performance tests. These were: (1) the approach based on autocorrelation and discrete cosine transform (DCT) proposed by Plataniotis *et al.* [47]; (2) the methodology proposed by Tawfik *et al.* [48] using DCT coefficients from average QRS; (3) an approach based on discrete wavelet transform (DWT) proposed by Belgacem *et al.* [49]; and (4) the deep autoencoder proposed by Eduardo *et al.* [42].

**Template Update Strategies:** Two strategies of template update were implemented to ensure the models are up-to-date on the identity information of the enrolled subjects. The first is *FIFO* (first-in-first-out): the database is updated using new samples whose similarity to the current templates is above a defined threshold (obtained empirically using the training data), replacing the oldest template of the same identity. The second is *Fixation*, where certain templates are fixed, allowing only the remaining stored samples to be updated: this ensures some initial labeled identity information remains on the system over time.

**Experiments:** Experiments used the E-HOL-03-0202-003 database (commonly called E-HOL 24h). It consists of three-lead Holter recordings of 202 healthy subjects. The training set consisted of the last 30 seconds of the first 60 minutes, to mimic short enrollment times and avoid the initial resting period. To study performance over time, testing was performed over seven time points: one immediately after enrollment, another after one hour, and regularly until the end of the records.

**Results:** The performance results at each test hour (see Fig. 10) show that performance decays significantly, even over relatively short periods. The template update techniques were successful in reducing the performance decay over time. The FIFO technique resulted in improvements of 8 – 9% accuracy, on average, whereas the Fixation strategy resulted in an average accuracy increase of 10%. Overall, the results show long-term identification performance in ECG biometrics is generally weak, and template update techniques should

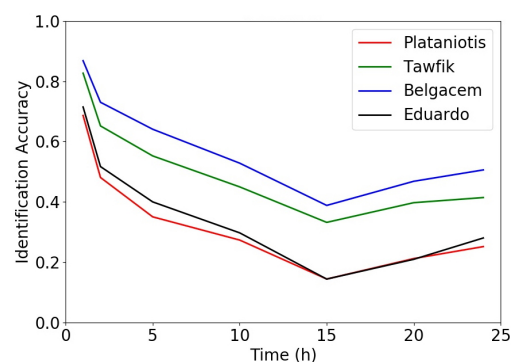


FIGURE 10: Evolution of the state-of-the-art ECG-based identification accuracy over time [46].

be studied further for enhancing the long-term performance of state-of-the-art methods.

The main contributions of this work for the AUTOMOTIVE project can be summarized as follows:

- **Deeper knowledge of long-term performance decay:** through this comprehensive study of the impact of ECG variability over time on the performance of state-of-the-art algorithms, the AUTOMOTIVE project now enjoys deeper knowledge of the hurdles that should be overcome to achieve suitable real applications;
- **Comparison of update strategies:** template update is shown to be a must for real ECG biometric applications, and the comparison of update strategies performed in this work paves the way for more robust algorithms in realistic scenarios.

For AUTOMOTIVE’s central topic of drowsiness detection, the knowledge and methods built in this work are essential for robust long-term ECG biometrics. In turn, these are paramount for the aforementioned continuous user-tuning of drowsiness monitoring models.

## B. xAI IN BIOMETRICS

In the absence of a mathematical definition, interpretability is the degree to which a human can “understand the cause of a decision” [50] or, in Machine Learning context, “consistently predict the model’s result” [51]. Thus, a model is more interpretable when it is easier for a person to identify why it took a certain decision. Moreover, a model is more interpretable than another if the former’s decisions are easier to understand [50]. For biometrics, this challenge has only just started to be unveiled, with researchers now beginning to use interpretations to improve their biometric models [52], [53].

The work proposed by us in [54] was the first to use interpretability to understand how ECG signals carry identity information, how that information changes with more realistic acquisition settings, and discuss the possibility of using such insights during training for improved robustness to signal noise and variability. In addition, our work in [55] was a pioneer at incorporating interpretability in face Pre-

sensation Attack Detection (PAD) methods regarding a wide range of attacks. We believe that the research in both of those problems, inserted in the AUTOMOTIVE project, can remarkably impact the outcomes of the next generation of biometric systems: more accurate, more robust, and more transparent.

### 1) xAI for ECG Biometrics

Research in ECG biometrics has been steadily evolving from smaller databases of high-quality signals acquired in medical settings (*on-the-person* settings) towards larger databases of noisier signals acquired in more realistic scenarios (*off-the-person* settings). One defining characteristic of this evolution is the movement away from fiducial-based methodologies in favor of holistic approaches. Initially, several approaches successfully used the QRS complex (the key defining feature of the ECG) or its amplitude or time measures for identification, as it was considered more stable over time and across variable conditions. However, using only QRS complexes is increasingly uncommon as research evolves towards off-the-person signals and larger databases. This denotes the QRS may not be enough for identity recognition in large populations and more realistic scenarios.

This subsection describes a study [54] conducted to understand how a state-of-the-art end-to-end model uses signal information in diverse scenarios, leveraging the recent tools developed for model interpretability. The structure of the proposed methodology had the following main steps:

**Biometric Identification Model:** The model followed the architecture proposed in [37], [38] (see section III-A1): an end-to-end 1D convolutional neural network (CNN) with four convolutional layers. Neighboring convolutional layers were separated by max-pooling layers. The last convolutional layer was followed by two fully-connected layers, of which the first has 100 neurons and ReLU activation, and the last has  $N$  neurons and softmax activation ( $N$  corresponds to the number of considered identities). This model was trained with either the PTB [23] or the UofTDB [22] databases, on increasingly larger subsets of identities.

**Interpretability Method:** Interpretability is a quickly growing topic that is contributing to the more complete understanding of the often elusive behavior of deep learning models [56]. To better understand the trained models for ECG biometrics, the interpretability tools used in this work were Occlusion [57], Saliency [58], Gradient-SHAP [59], and DeepLIFT [60], as implemented in the Captum [61] library for PyTorch.

**Results:** Analysing the explanations obtained (see an example overview in Fig. 11), one can verify a trend from smaller to larger identity subsets where the relevance of QRS, initially dominant, is increasingly shared with other parts of the signal. A similar dynamic is verified when comparing explanations with on-the-person *vs.* off-the-person signals: the focus is mostly on the QRS complex for the former, but relevance is more evenly shared with other waveforms when considering the latter.

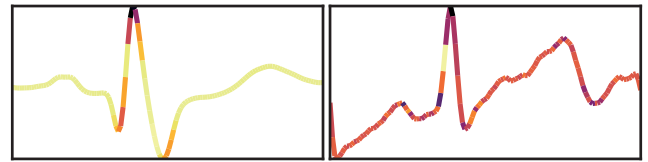


FIGURE 11: Example average heartbeat with the relevance of each portion (darker colour relates to higher relevance; left: on-the-person with small identity set; right: off-the-person with large identity set).

Overall, while the QRS seems to be the most important part of the ECG signal for biometrics, it can only be reliably used alone for on-the-person scenarios with smaller populations. When considering larger sets of identities and more realistically noisy signals, the information carried by other ECG waveforms is important for robust and accurate decisions.

The main contribution of this work for the AUTOMOTIVE project can be summarized as follows:

- **Better understanding of ECG biometrics:** the AUTOMOTIVE project moves forward with new evidence of the relative relevance of the ECG waveforms on diverse conditions of signal quality and population size. This generated new knowledge that steers research towards the right path to more robust and accurate ECG biometric algorithms;
- **Application impact:** this work was the first one in interpretability for ECG biometrics, establishing a guide for researchers in this field with several results, suggestions for future work, and an intuitive way to visualize interpretations for unidimensional signals.

For the topic of ECG biometrics, the exploration of transparency-related topics through the use of interpretability is important as research quickly moves towards deep learning models. This leads to improved ECG biometrics which, by itself, already benefits drowsiness monitoring as discussed before. However, exploring interpretability and explainability for 1D signals opens new doors to understand what information is indeed useful for drowsiness monitoring, where it is located in the physiological signals, and how to best capture it.

### 2) xAI for Face PAD

In this subsection, a study on interpretability tools applied on deep neural networks trained for PAD in face biometrics is reported [55]. The structure of the proposed methodology had the following main steps:

**Presentation Attack Detection Network:** The deep neural network model was a simple end-to-end convolutional neural network composed of four convolutional layers, with three max-pooling layers interposed between them, and three fully connected layers. The four convolutional layers were composed of 32, 32, 64, and 64 filters, respectively, with size  $3 \times 3$ , unit stride, and padding. The max-pooling was



performed in  $2 \times 2$  regions with stride 2. The dense layers were composed of 100, 100, and 2 neurons, respectively. All convolutional and fully connected layers were followed by rectified linear unit (ReLU) activations, except for the last dense layer, which was followed by softmax activation.

**Interpretability Method:** The Gradient-weighted Class Activation Mapping (GradCAM) tool was applied to the last convolutional layer of the model, allowing for a) different importance values to each neuron for a particular decision of interest, b) explanations for any layer of the network, and c) analysis of the model predictions at the class level.

#### Evaluation Frameworks:

- **Mix-Attack:** the model was trained and tested with *bona fide* samples and all the varieties of attacks available;
- **One-Attack:**
  - The model was trained and tested with *bona fide* samples and one single type of attack, which was already seen by the network during the training step;
  - The *bona fide* samples of one random subject were present in the training set on one evaluation iteration, and were then swapped to the test set for a second evaluation run;
- **Unseen-Attack:** the model was trained with all but one type of attack and tested with only this type, besides the *bona fide* samples in train and test steps.

The main contribution of this work for the AUTOMOTIVE project can be summarized as follows:

- **Identification of desirable properties for generalization of deep neural models to unseen data and attacks:** the AUTOMOTIVE project moves forward by using interpretability tools to state the following properties of the models: 1) explanations for the same sample should be similar whether or not it is seen during training; 2) explanations for the same sample should be similar whether or not the model is trained to detect that specific attack; 3) explanations should be similar for different samples with the same predicted label; and 4) explanations should be meaningful;
- **Application Impact:** this work was the first one in interpretability for face biometrics, establishing a guide for researchers in this field with several results and suggestions for future work.

By exploring interpretability for face images, this work also paves the way for more transparent driver drowsiness monitoring based on face video. The knowledge acquired with this study not only leads to improved biometric recognition but also illustrates new frameworks to understand what facial features are the most informative for recognizing drowsiness in vehicle drivers.

### C. GENERAL BIOMETRIC APPLICATIONS

#### 1) Template security on end-to-end models

The task of recognizing identity requires the storage of highly sensitive personal information. As such, three template se-

TABLE 6: Summary of the Secure Triplet Loss results for scenario A.

Method	$EER$ (%)	$FNMR @ FMRV = 0.1\%$	$FMR_C @ EER$	$D_{\leftrightarrow}^{sys}$
Triplet Loss	12.56	0.9033	-	-
STL w/KLD	13.58	0.8700	0.0	0.005
STL w/SL	13.33	0.9458	0.0	0.004
BF [62]	15.76	0.9242	0.0075	0.234
HE [63]	12.49	0.9573	0.0806	0.002

curity properties should be verified by any good biometric system: cancelability, unlinkability, and irreversibility. These are typically achieved through tailored feature extraction, encryption, or bihashing schemes before storage and matching. The most prominent of such schemes are Bloom Filters (BF) [62] and Homomorphic Encryption (HE) [63].

However, the state-of-the-art in biometric recognition is increasingly dominated by deep learning approaches, and adding separate processes of protection and matching is sub-optimal and creates additional hurdles that may limit achievable performance. All of this calls for an integration of template protection within deep learning models, and since these are so flexible and have been able to learn so many sophisticated tasks, why not have them learn template protection as well?

**Proposed Methodology:** The Secure Triplet Loss (STL) was proposed [64], [65] to achieve the aforementioned goals. Having a model which receives a biometric sample and a cancelability key and outputs a template, the training objective function includes a component for identity and cancelability and a component for linkability (further details in the original publication [65]). The first component adapts the original triplet loss to not only push away templates with different identities but also templates with different keys, clustering only templates that agree on both identity and key and thus promoting cancelability. The second loss component measures the linkability of the templates in a batch, either through the Kullback-Leibler divergence (*STL w/KLD*) or using the differences in mean and standard deviation (*STL w/SL*).

**Experiments:** The proposed approach was evaluated for identity verification in two scenarios: (A) training a network “from scratch”, and (B) adapting and fine-tuning an existing model to make it output protected templates. Scenario A was explored for ECG biometrics, using the end-to-end approach in [37], [38], [54], [66] and the UofTDB database [22]. Scenario B was explored for face biometrics, using the Inception-ResNet-V1 [67], pretrained on VGGFace2, with the YouTube Faces database [26].

**Results:** The results on scenarios A and B are presented in Table 6 and Table 7, respectively. Results are presented for performance (equal error rate -  $EER$ ); and false non-match rate at 0.1% false match rate -  $FMR @ FNMR = 0.1$ ), cancelability (false cancelability match rate at the  $EER$  point -  $FMR_C$ ), and linkability ( $D_{\leftrightarrow}^{sys}$  as proposed in [62]).

In scenario A, the results show the proposed method largely avoids performance losses vs. the unprotected triplet

TABLE 7: Summary of the Secure Triplet Loss results for scenario B.

Method	EER (%)	$FNMR @ FMR_V = 0.1\%$	$FMR_C @ EER$	$D_{\leftrightarrow}^{sys}$
Triplet Loss	13.99	0.8496	-	-
STL w/KLD	15.93	0.8586	0.0089	0.132
STL w/SL	15.15	0.8771	0.0182	0.070
BF [62]	17.07	0.9103	0.0396	0.245
HE [63]	15.06	0.8312	0.0371	0.001

loss, offering better performance than with BF. The STL also offers the best cancelability results and unlinkability levels near those offered by HE. In scenario B, the STL verifies some performance gap *vs.* the baseline, but is nevertheless aligned with the state-of-the-art HE and considerably better than BF. STL once again offered the best cancelability and acceptable linkability results.

The main contributions of this work for the AUTOMOTIVE project can be summarized as follows:

- **A novel method for biometric data protection:** STL is a simple and competitive alternative to state-of-the-art template protection schemes. Through a tailored objective function, it teaches cancelability and unlinkability to the biometric model, avoiding any separate protection process. Thus, STL can be successfully applied for any biometric trait and most trainable architectures, and not only for new models but also to transform existing ones and lead them to deliver protected templates;
- **Security without performance losses:** Conducted experiments show that STL is able to match the state-of-the-art in template protection (HE) in unlinkability. Most importantly, STL attains the best template cancelability results while avoiding considerable performance decay;
- **Lightweight and flexible:** As it only requires minor architecture changes and no additional processes, STL allows models to retain the original size and average inference time while offering good security levels. Moreover, STL can be applied to new or existing models of varying complexity, paving the way for the lightweight in-vehicle biometric systems foreseen by AUTOMOTIVE.

The lightweight, flexible strategy for biometric template protection proposed in [64], [65] is also of paramount relevance for the new generation of driver drowsiness monitoring systems. To continuously learn subject-specific patterns of drowsiness, such models require biometric systems, which will need to securely store driver identity information. Through the Secure Triplet Loss, this can be done effortlessly and with minimal processing requirements, in a way that is suitable for embedded systems in in-vehicle scenarios.

#### D. EMOTION RECOGNITION THROUGH ECG AND FACE ANALYSIS

Automatic facial expression recognition (FER) has been one of the key problems in the human-computer interaction field,

with growing application areas including neuromarketing, crowd analytics, biometrics, or clinical monitoring [68]. Expression recognition is a task that human beings perform daily and effortlessly, but it is not yet easily performed by computers. Although recent methods, particularly those using deep learning, have demonstrated remarkable performances in highly controlled environments, the automatic FER in real-world scenarios is still a very challenging task [68]. In addition, the performance of deep models is still below its full potential as training high capacity models in small datasets, such as the ones available in the FER field, usually result in overfitting.

To work around the problem of training high-capacity classifiers on small datasets, previous FER works have mainly resorted to (i) transfer learning [69], where a CNN is typically pre-trained in some domain-related dataset before being fine-tuned to the target dataset; and (ii) classifier ensembles [70], in which an ensemble of CNNs is created to combine their decisions and, hence, reduce the model's variance. However, their benefits are tightly coupled with the source-target domain similarity.

In terms of motivation, the work of Liu *et al.* [71] is probably the most related to the proposed methodology in the AUTOMOTIVE project, as they also explore the psychological theory that facial expressions are the result of the motions of facial muscles.

However, some remain skeptical about emotion recognition based on facial expressions, as people are capable of counterfeiting these to convey fake emotions [72]. It is known that emotional states influence the autonomic nervous system and, consequently, the morphology of the physiological signals. This explains the current efforts towards affective computing based on physiological signals (such as the ECG), that cannot be voluntarily altered to fake emotions.

Despite several studies addressing emotion recognition using physiological signals, it is still uncertain how emotion variations translate into actual pattern alterations in each physiological signal. Moreover, for the specific case of ECG-based emotion recognition, despite the generally encouraging results reported in the literature [73], there are a plethora of problems to be addressed, mostly related to the scarcity of data, their limited variety, and the subjectivity of corresponding emotion labels.

These current challenges in ECG-based emotion recognition reflect on the performance of the algorithms. These often offer severely inferior accuracy when evaluated under realistic scenarios, failing to live up to the expectations set by the results reported in the literature. For example, the methodologies proposed by [74]–[77] all offer relatively high accuracy, but only when evaluated on random data splits. One should expect performance to decay sharply once the methods are evaluated on disjoint sets of recordings (signal-independent) and subjects (subject-independent), as they would in real applications.

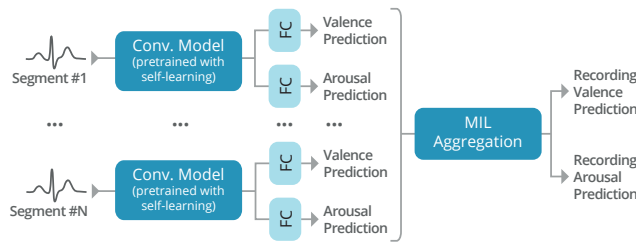


FIGURE 12: Schema of the proposed methodology for ECG-based emotion recognition as proposed in [73]. MIL aggregates individual predictions for consecutive segments into a single recording prediction.

### 1) ECG Analysis

The work described in this section (and in [73] in further detail) focused on the development of a methodology for emotion recognition that takes advantage of the continuous nature of the ECG signal for improved accuracy. Special efforts were devoted to ensuring the evaluation settings adequately mimic realistic signal-independent and subject-independent settings, thus offering accurate performance estimates that are more likely to be verified in real applications.

**Methodology:** The proposed methodology is built upon the approach proposed in [77], taking advantage of the self-learning pre-training that offered higher robustness despite data scarcity. The approach (see Fig. 12) is based on a convolutional neural network that receives short ECG segments. Initially, the model is prepared and trained to recognize a set of eight transformations applied to the input signals (self-learning): noise addition, scaling, negation, temporal inversion, permutation, time-warping, baseline wander, and magnitude warping. After this pre-training stage, the convolutional layers are frozen and the top of the model is adapted to provide valence and arousal predictions. Individual predictions from consecutive segments are combined into long-time predictions using multiple instance learning (MIL) through heuristic methods (maximum, mean, and median), a multilayer perceptron (MLP), a long short-term memory (LSTM) network, or a bi-directional LSTM (BiLSTM).

**Experiments:** Experiments used ECG signals from the DREAMER [78], AMIGOS [79], and MAHNOB-HCI [80] databases. For the individual predictions, 10-second segments were considered. All segments from each recording were considered for the corresponding aggregated MIL predictions. Data were either divided randomly between training and testing (random division), divided by recording (signal-independent setting), divided by subject (subject-independent setting), or trained and tested in different databases (cross-database setting).

**Results:** As expected, results show a sharp decay of performance when we move from random data division settings (the most common in the literature) to signal-independent settings (more realistic). For the former, the methodology achieves 75–79% accuracy on individual predictions vs. 54–56% accuracy for the latter. In cross-database experiments,

trained on the DREAMER database, the method achieves 46 – 53% accuracy on the AMIGOS database and 55 – 61% for the MAHNOB-HCI database. However, results appear to improve once MIL is applied, especially when using heuristic methods, which illustrates the benefit of considering the continuity of the ECG for emotion recognition.

The main contributions of this work for the AUTOMOTIVE project can be summarized as follows:

- **A realistic perspective over the state-of-the-art:** This work identified critical flaws in the way ECG-based emotion recognition algorithms are evaluated. With this, it was possible to restructure the evaluation settings into more challenging and realistic scenarios for more accurate results that are more likely to be verified in real applications;
- **An improved methodology for ECG-based affective computing:** Taking advantage of the continuity of the ECG signal, the base methodology was successfully adapted with MIL techniques for improved performance. This brings us closer to the target real applications foreseen by the AUTOMOTIVE project.

### 2) Face Analysis

This subsection presents a proposed end-to-end deep learning approach for emotion recognition using prior knowledge on facial expressions [81]. The novel deep learning network architecture along with a well-designed loss function explicitly models both informative local facial regions and expression recognition. The intuitive idea is to learn the most relevant facial regions for expression recognition, such as facial components (*i. e.*, eyes, eyebrows, nose, mouth) and expression wrinkles. To accomplish this purpose, the proposed neural network is composed by three main components, namely (i) the *facial-parts component*, (ii) the *representation component*, and (iii) the *classification component* (see Fig. 13).

**I - Facial-Parts Component:** The facial-parts component learns an encoding-decoding function  $E(x)$  that maps an input image  $x$  to a relevance map  $\hat{x}$ , representing the probability of each pixel being relevant for recognition. The loss function, supported by the physiology knowledge that facial expressions are decomposed into several action units of facial muscles, is defined to enforce sparsity and spatial contiguity on the activations of  $\hat{x}$ . Thus, three regularization strategies for regression of  $\hat{x}$  were proposed.

**II - Representation Component:** The representation component learns an embedding function  $F(x, \hat{x})$  that maps an input image  $x$  and its relevance map  $\hat{x}$  to a hidden representation  $h$ . The relevance map  $\hat{x}$  from the facial-parts component is used to filter the learned representation  $h$ , leading it to only activate strongly to the most relevant facial parts.



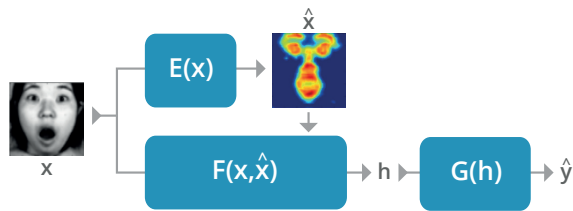


FIGURE 13: Architecture of the proposed model for emotion recognition based on facial expressions (adapted from [81]).

**III - Classification Component:** The classification component consists of a sequence of fully connected layers followed by a softmax output layer. The largest probability output is chosen as the class prediction.

The main contribution of this work for the AUTOMOTIVE project can be summarized as follows:

- **Facial Expression Recognition Model:** the AUTOMOTIVE project moves forward with a model of an end-to-end deep neural network along with a loss function defined to regularize the entire learning process so that the proposed method can explicitly learn expression-specific features; the approach is based on the strong prior knowledge that facial expressions are the result of the motions of some facial muscles and components;
- **Application Impact:** facial expressions represent an important component for emotion recognition and person identification with strong applications in autonomous driving or mobility solutions. The model can be explored for emotion categorization, action recognition, or well-being monitoring of vehicle occupants.

Emotion recognition can complement and extend drowsiness detection and, hence, help pave the way for an all-in-one and more robust well-being monitoring system. Furthermore, these techniques can help promote road safety, particularly if used in the detection of negative emotional states in the driver, such as anger or stress [82].

#### E. DROWSINESS DETECTION THROUGH ECG, PPG AND EOG ANALYSIS

Normal human sleep is composed of a Rapid-Eye-Movement (REM) stage and four non-REM stages. These stages have well-defined characteristics and alternate cyclically, with a standard human adult cycle lasting approximately 90 minutes. This cycle starts with a non-REM stage, and finishing in REM sleep [83]. According to Keenan *et al.* [84], “sleep is a reversible behavioral state of perceptual disengagement from and unresponsiveness to the environment.”

The term *drowsiness*, often used interchangeably with the terms *fatigue* and *sleepiness* in the literature, relates to a physiological need to sleep, as it is the intermediate state between wakefulness and sleep [85]. This state comes with the impairment of visual perception, of higher cognitive functions, the inability to maintain visually focused attention, among other undesirable consequences for safe driving [86].

Since physiological needs cannot be continuously avoided, a drowsy state always precedes a sleeping episode [85]. As such, if a drowsy state is detected then the body is already struggling to remain awake, possibly indulging in microsleeps in the process [87]. Microsleeps have been associated with poor driving performance [88] and, during these episodes, attention gaps can impair the driver’s ability to respond to events [89].

Studies report great inter-subject variability in how drowsiness affects drivers’ performance [90]. Even for a given self-declared drowsiness level, indicators such as eye blink duration vary considerably [29] and, thus, objectively and non-intrusively measuring drowsiness has been a constant challenge in research.

Some subjective scales are used to score drowsiness based on the subjects’ responses concerning standardized sleep symptoms. Some examples include the Epworth Sleepiness Scale (ESS) [31], the Stanford Sleepiness Scale (SSS) [91], a Visual Analog Scale (VAS) [92], and the Karolinska Sleepiness Scale (KSS) [93]. The Objective Sleepiness Scale (OSS), which combines features of the electroencephalogram (EEG) and eye movements, presents some disadvantages, including the insufficient number of sleep stage categories [13] and its intrusiveness.

The KSS is a nine-point scale that defines verbal reference for each stage (1 - very alert; 3 - alert; 5 - neither alert nor sleepy; 7 - sleepy but not fighting sleep; 9 - very sleepy and fighting sleep [93]). The ground truth for the drowsiness level is assessed by having the subject periodically report their perceived score. Although not an objective measurement and capable of influencing the driver’s state [13], [29], the KSS is easy-to-apply, non-invasive, and widely used in research [93]).

While in a state of drowsiness, the cardiac system significantly alters its behavior and, as such, the study of these modifications can retrieve information regarding the alertness level of a driver. When a subject is under drowsiness, their heart rate becomes slower, more irregular, and blood pressure drops [94]. Under this state, the parasympathetic nervous system becomes more active and the sympathetic nervous system activity decreases, while the opposite remains true when the subject is under a wakeful state [94]–[98]. A reliable measure for the autonomic nervous system is the heart rate variability (HRV), which measures the variation of the cardiac cycle duration and is usually obtained through an R-peak detection of the ECG and a frequency-domain analysis [95].

Following this, a higher frequency (HF) band of the HRV relates to the parasympathetic activity, while a lower frequency (LF) band provides information about the sympathetic activity. Hence, the state of an individual can be characterized by calculating the ratio of (LF/HF) [99]. A more profound division of frequency bands is presented in Table 8.

Classification techniques based on the HRV, the ECG, electrooculogram (EOG), face video, or other data types are

TABLE 8: Frequency bands of HRV. The HF band relates to the parasympathetic activity, predominant in sleepy states.

Component	Frequency Range (Hz)
Ultra Low Frequency	$\leq 0.003$
Very Low Frequency	0.003 - 0.04
Low Frequency	0.04 - 0.15
High Frequency	0.15 - 0.4

commonly used to distinguish between drowsiness and alertness. With methods such as Bayesian [100]–[103], Linear Discriminant Analysis (LDA) [94], [100], [103]–[106], k-Nearest-Neighbors (kNN) [106], [107], Support Vector Machines (SVM) [102], [106], [108], and Artificial Neural Networks (ANN) [109], [110] being harnessed for this subject.

Further concurring with the severity of the context of the AUTOMOTIVE project, automotive manufacturers and other companies have developed solutions targeting the detection of drowsiness and attention during driving tasks. Some commercial solutions (for a complete present-day review, see [13]) include:

- **Steer Device by STEER Inc.**<sup>8</sup>, a bracelet that detects the level of drowsiness through the heart rate and the electrodermal activity (EDA);
- **IR-LED by Siemens**<sup>9</sup>, a camera system equipped with an infrared light-emitting diode and an infrared light sensor to detect microsleep events;
- **FaceLAB Driver Safety by Seeing Machines** [13], which detects driver drowsiness in real-time through eye blinks and percentage of eye closure (PERCLOS) analysis, recording face video to track the head, eyes, eyelids, and gaze;
- **Driver Attention Warning by Saab**<sup>10</sup>, which uses an infrared camera to monitor eye blinks and track gaze and head orientation;
- **Driver Alert Control by Volvo**<sup>11</sup>, which uses lane departure as a detection feature, monitoring the car movements concerning the road markings;
- **Active Safety by Volkswagen**<sup>12</sup>, which uses the Steering Wheel Angle (SWA) and lane departure as detection features and continuously evaluates traffic signals;
- **Attention Assist by Mercedes-Benz**<sup>13</sup>, which uses the SWA as a detection feature (it learns the driver steering

<sup>8</sup>Kickstarter: STEER: Wearable Device That Will Not Let You Fall Asleep by Creative Mode. Available on <https://www.kickstarter.com/projects/creativemode/steer-you-will-never-fall-asleep-while-driving?lang=fr>.

<sup>9</sup>Photonics.com: IR-LED Detects Drivers in Microsleep. Available on [https://www.photonics.com/Articles/IR-LED\\_Detects\\_Drivers\\_in\\_Microsleep/a44727](https://www.photonics.com/Articles/IR-LED_Detects_Drivers_in_Microsleep/a44727).

<sup>10</sup>Saab Driver Attention Warning System. Available on <https://www.saabnet.com/tsn/press/071102.html>.

<sup>11</sup>Volvo Support: Driver Alert Control (DAC). Available on <https://www.volvocars.com/en-th/support/manuals/v40/2017w17/driver-support/driver-alert-system/driver-alert-control-dac>.

<sup>12</sup>Volkswagen UK: Driver alert system. Available on <https://www.volkswagen.co.uk/technology/car-safety/driver-alert-system>.

<sup>13</sup>No Doze: Mercedes E-Class alerts drowsy drivers. Available on <https://www.autoweek.com/news/a2032716/no-doze-mercedes-e-class-alerts-drowsy-drivers/>.

TABLE 9: Summary of the open issues to be considered in drowsiness detection regarding three types of information that can be acquired from the driver (adapted from [13]).

Issues	Vehicle Behavior	Physiological Signal	Driver Behavior
Use of standard objective measures	X	X	X
Environment conditions influence	X		X
Difficulty to extract symptoms			X
Circadian and wake phases	X	X	X
Accuracy with fewer sensors		X	
Intrusive nature of sensors		X	
Impact of artifacts		X	
Real-time detection support		X	X

pattern at the beginning of a driving session), as well as braking and acceleration events, the duration of the session, and road conditions;

- **Driver Attention Monitor by Lexus**<sup>14</sup>, which tracks eye and head movements to detect whether the driver is looking forward or not, and includes an obstacle detection feature.

After a comprehensive analysis of both current research and commercial solutions for driver drowsiness detection systems, Doudou *et al.* [13] discussed the open technological issues in this field regarding three types of information that can be acquired from the driver, as presented in Table 9.

In addition to these limitations, one frequent conclusion in this field of research is that generalizing to different individuals is extremely difficult, as models perform significantly worse on data from new subjects [111]. Simultaneously, large amounts of high-quality data are needed to train and validate drowsiness detection methods, which is still a challenge due to the cost of deploying real-traffic monitoring systems and the concerns regarding safety and data privacy.

The following subsections present the work conducted for the AUTOMOTIVE project on the topic of drowsiness detection. These endeavors aimed to study the aforementioned limitations of the current state-of-the-art and advance the field by proposing new and improved algorithms.

#### 1) Intrusive and non-intrusive signal acquisition study

In this subsection, a combination of different methods of intrusive and non-intrusive signals to achieve the best possible performance for driver drowsiness detection is reported [112].

The method follows a standard workflow for a supervised machine learning classification problem: 1) extracting features, 2) randomly splitting data into training and testing datasets, 3) choosing the best hyperparameters with ten-fold

<sup>14</sup>Lexus Safety Technology. Available on <https://www.lexus.com/safety>.

cross-validation, and 4) evaluating classifiers using accuracy and F1-score.

The ECG, EOG, and video measures such as the eyelid distance, gaze angles, head pose, and pupil diameter in non-overlapping two-minute windows were used as features. The available KSS scale was considered to categorize sleepiness. For the comparison study, five machine learning classifiers were explored: Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Networks (ANN), Gradient Boosting Tree (GBT), and K-Nearest Neighbors (kNN).

In a summary of the obtained results, the work states that it is preferable to follow a hybrid approach including ECG features in addition to EOG or video features. Also, no classifier seems to be significantly superior to the others, and an imbalance between the alert and drowsy classes has a substantial impact on classification accuracy.

The main contribution of this work for the AUTOMOTIVE project can be summarized as follows:

- **A reference scheme for researchers in driver handover strategies:** the AUTOMOTIVE project moves forward with a baseline reference scheme that compares intrusive and non-intrusive signal acquisition methods to detect driver drowsiness;
- **Application Impact:** As presented in section II, the main approaches used for detecting driver drowsiness in real-time can be divided into three categories: vehicle-based measures, behavioral measures, and physiological measures. While each method has its strengths and weaknesses, being important to assess the combination of different measures to achieve the best possible performance, real-world drowsiness detection solutions need to use non-intrusive acquisition methods. With this in mind, one approach for drowsiness detection is to consider a non-intrusive camera-based method with a physiological method. However, which combination is the best? How much performance gains can one attain by fusing multiple data sources? This study discussed these questions, whose answers have wide applicability to other topics related to wellbeing monitoring.

## 2) Subject-dependent for driver fatigue classification

This subsection presents the first in-depth study on the use of ECG and EOG for subject-dependent classification in driver sleepiness/fatigue under realistic driving conditions [113]. The proposed methodology starts with a preprocessing stage, specific to each signal, followed by a feature extraction step. The classification procedure was composed of three different tests: multimodality, subject-independent classification, and imbalanced class distributions. KSS ratings collected every fifth minute during the experiments were used as labels. ANNs, RF, SVM, and GBT were used and, based on accuracy values and ten-fold cross-validation, the best combination was selected. Next, the proposed methodology is summarized.

**Preprocessing and feature extraction:** The preprocessing of ECG consisted of a bandpass Butterworth filter with

cut-off frequencies of 4 and 50 Hz, and the feature extraction of ECG consisted in the following calculations: a) for R-peak detection, the signal was divided into two-minute windows; the R-peaks selected are the ones that contain typical heart-beat values and the lowest RR standard deviation (RRSD) value; b) 8 time-domain statistical features were calculated using HRV time series; c) 8 frequency-domain features were calculated from the power spectrum density (PSD) of the HRV signal. The preprocessing of EOG consisted of a band-pass Butterworth filter with cut-off frequencies of 0.1 and 30 Hz and convolution with a Hamming window. The feature extraction of EOG consisted of several procedures to extract blink events and eye saccades.

**Classification:** Multimodality scenario: the data were randomly split between a training set (70%) and a test set (30%). Subject-dependent scenario: a) Training and testing sets were split randomly; b) Training set: data from  $n - 1$  subjects and 30% of the data from the  $n$ -th subject. Testing set: 70% data from the  $n$ th subject; c) Training set: data from  $n-1$  subjects and 10% of the data from the  $n$ th subject. Testing set: 90% data from the  $n$ th subject. Subject-independent scenario: Training set: data from  $n-1$  subjects and 0% of the data from the  $n$ th subject. Testing set: 100% data from the  $n$ th subject. Imbalanced class distributions: A ratio between the number of samples in each class is considered for balancing misclassification costs and, therefore, balancing data for each class.

Based on the obtained results, this work finds that the accuracy improves when a combination of ECG and EOG features is used. Moreover, results show significantly worse performance in subject-independent classification, especially for the sleepy class. In conclusion, applying methods for imbalanced distributions can be a promising approach.

The main contribution of this work for the AUTOMOTIVE project can be summarized as follows:

- **A reference scheme for researchers in driver handover strategies:** the AUTOMOTIVE project moves forward with a baseline study on the use of ECG and EOG for subject-dependent classification in driver sleepiness/fatigue under realistic driving conditions. Several crucial conclusions that help further developments were possible to state under this work. For instance, individual differences are present in the physiological signals and the labels, the reliability of the sleepiness ground truth has an impact on the design and the optimization of the classifiers, combinations of different measures (e.g., lane deviations, mathematical models of sleepiness) can improve the results in real driving conditions;
- **Application Impact:** New lines of research arose from this work, such as the use of biometrics to develop subject-personalized models for drowsiness detection with lifelong learning.



### 3) Peripheral cardiac signal acquisition and use in drowsiness classification

This subsection presents a study of the feasibility of producing a drowsiness detection system based on peripheral cardiac signal [35]. This study encompassed three main stages:

**Signal collection and conversion into streams of inter-beat intervals (IBIs):** To collect cardiac rhythm information, the chest strap Movesense, the capacitive steering wheel CardioWheel, and the wrist photoplethysmograph (PPG) sensor PulseOn were used. The CardioWheel (ECG-based) directly provided the IBIs and the Pan-Tompkins algorithm was used to detect the R-peaks in the Movesense electrocardiographic signals. For the PPG sensor, an online filter that mimics recursive moving average removal was applied to one-second windows, and an adaptive threshold peak detection algorithm was implemented to locate peaks on the PPG signal. An IBI corrector system was then created to treat the outliers that resulted from moments of poor electrical contact and artifacts. This system was shown to be capable of reconstructing the sequence of IBIs from signals corrupted with 10% missed detections and additional 10% false peaks with less than 7.5 ms of mean absolute deviation from the true signal. This system is relevant to ensure that all collected information is used to calculate the HRV.

**Development of a model based on IBIs values to detect drowsiness:** For this stage, only data collected through a chest ECG was used. An initial set of time and frequency domain features was used to compare 4 decision models (SMV, one-class SVM, GBT, and ANN) and it was realized that, independently of the architecture, the model performed poorly in arbitrary individuals. Personalized models were then investigated with great improvements for the twelve selected individuals of the Swedish National Road and Transport Research Institute dataset with a trustworthy self-report and balanced experience of alert and drowsy states. The SVM model was selected as the best fitted to binarily classify the personalized state of drowsiness and, by revising features and tuning hyper-parameters (defining an SVM with linear kernel and C parameter 0.3), a mean performance of  $0.63 \pm 0.03$  MCC (Matthews Correlation Coefficient) was attained.

**Model comparison with IBIs measured from peripheral signal:** Following the previous stages, the capability of the model to detect drowsiness with IBIs measured from a peripheral signal was tested. Per the experiment described in sections II-B2 and II-B3, chest ECG, hands ECG, and wrist PPG were collected, converted into IBIs, and the HRV features were calculated for every 2-minute window in each of the signals. Although only the data from 2 subjects survived the criteria followed by the Swedish National Road and Transport Research Institute dataset, the model trained with chest ECG remained fairly performing when applied to wrist PPG, with the scores ranging from 0.34 to 0.61 MCC.

Summing up the obtained results, this work found that even though the proposed system based on peripheral cardiac signal is feasible, future efforts should be devoted to tackling the limited size of the analyzed population. Furthermore,

given that it is hypothesized that a limited set of individual models can be representative of the possible ranges of HRV for a general population, gathering and combining this set in a voting scheme or other ensemble classification framework might be a promising approach to finally developing a generalized, subject-independent system to detect driver drowsiness.

The main contribution of this work for the AUTOMOTIVE project can be summarized as follows:

- **A reference scheme for researchers in driver handover strategies:** the AUTOMOTIVE project moves forward with a baseline study on the subject-dependent binary classification of driver drowsiness based on the peripheral cardiac signal, which is acquired non-intrusively. The present study concurs with the subject-dependent study presented on 2) in terms of the reliability of the sleepiness ground truth. Upon further analysis of the data where the models were constantly performing badly, an unbalance in classes was observed, as not all participants have managed to provide enough ratings associated with being sleepy for the model to properly learn the separation boundary between the two classes, indicating that a balanced class distribution is important for the topic at hand.
- **Application Impact:** This work showed that on a restricted dataset driver drowsiness detection systems can be capable of using any format of cardiac rhythm sensor to assess the driver's state. More research in a more variety of sensors can contribute to the development of more flexible driver monitoring systems and advanced driver-assistance systems (ADAS).

## IV. CONCLUSION AND FUTURE WORK

This paper presented the breakthroughs and achievements towards driver drowsiness monitoring throughout the AUTOMOTIVE project. Despite the diversity of explored research topics, AUTOMOTIVE has kept its single central target: to usher in the new generation of driver drowsiness systems.

AUTOMOTIVE's efforts towards immersive driving simulators with realistic procedural road generation and state-of-the-art labeling strategies brought us the ability to collect more and better data. The extensive research on ECG-based biometric recognition opened the door to user-tuned continuously-learning drowsiness algorithms. The developed emotion and drowsiness methodologies have addressed the main problems currently besieging the research community and delaying the real deployment of reliable commercial applications. And, the innovative studies on interpretability lead us to more transparent and trustworthy monitoring systems.

Despite these contributions, many hurdles are still unconquered. Further effort should be devoted to the full integration of these algorithms into a final robust drowsiness monitoring system. Other relevant challenges that remain are the development of even more realistic simulator prototypes or even larger initiatives for the acquisition of naturalistic data. Furthermore, new approaches are needed for model training

less focused on subjective labels, strategies to learn from continuous unlabeled data sources, and models that maintain high-performance levels in hyper-realistic long-term usage scenarios.

## ACKNOWLEDGMENTS

The presented research work was financed by the European Regional Development Fund (ERDF) through the Operational Programme for Competitiveness and Internationalization (COMPETE 2020 Programme), and by national funds through the Portuguese funding agency, Fundação para a Ciência e a Tecnologia (FCT), within the “POCI-01-0145-FEDER-030707” project and the “SFRH/BD/137720/2018” Ph.D. grant. The authors also wish to acknowledge all other collaborators in the AUTOMOTIVE project, especially Licínio Oliveira and Cláudia Sofia Silveira, for their valuable contributions.

## REFERENCES

- [1] Y. Choi, S. I. Han, S.-H. Kong, and H. Ko, “Driver status monitoring systems for smart vehicles using physiological sensors: A safety enhancement system from automobile manufacturers,” *IEEE Signal Processing Magazine*, vol. 33, no. 6, pp. 22–34, 2016.
- [2] S.-N. Lu, H.-W. Tseng, Y.-H. Lee, Y.-G. Jan, W.-C. Lee et al., “Intelligent safety warning and alert system for car driving,” *Journal of Applied Science and Engineering*, vol. 13, no. 4, pp. 395–404, 2010.
- [3] D. R. Mayhew, H. M. Simpson, K. M. Wood, L. Lonero, K. M. Clinton, and A. G. Johnson, “On-road and simulated driving: Concurrent and discriminant validation,” *Journal of safety research*, vol. 42, no. 4, pp. 267–275, 2011.
- [4] J.-M. Auberlet, F. Rosey, F. Anceaux, S. Aubin, P. Briand, M.-P. Pacaux, and P. Plainchault, “The impact of perceptual treatments on driver’s behavior: From driving simulator studies to field tests—first results,” *Accident Analysis & Prevention*, vol. 45, pp. 91–98, 2012.
- [5] J. H. Hansen, K. Takeda, S. M. Naik, M. M. Trivedi, G. U. Schmidt, Y. J. Chen, and W. Trappe, “Signal processing for smart vehicle technologies: Part 2,” *IEEE Signal Processing Magazine*, vol. 34, no. 2, pp. 18–21, 2017.
- [6] J. C. Monteiro and J. S. Cardoso, “A cognitively-motivated framework for partial face recognition in unconstrained scenarios,” *Sensors*, vol. 15, no. 1, pp. 1903–1924, 2015.
- [7] G. Kaur, G. Singh, and V. Kumar, “A review on biometric recognition,” *International Journal of Bio-Science and Bio-Technology*, vol. 6, no. 4, pp. 69–76, 2014.
- [8] A. Lourenço, A. P. Alves, C. Carreiras, R. P. Duarte, and A. Fred, “CardioWheel: ECG biometrics on the steering wheel,” in *Joint European conference on machine learning and knowledge discovery in databases*. Springer, 2015, pp. 267–270.
- [9] F. Agrafioti, K. Martin, and S. Oung, “Preauthorized wearable biometric device, system and method for use thereof,” 2016, nymi Inc., US9472033B2, USA. [Online]. Available: <https://patents.google.com/patent/US9472033B2>
- [10] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, “Deepface: Closing the gap to human-level performance in face verification,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1701–1708.
- [11] F. Schroff, D. Kalenichenko, and J. Philbin, “Facenet: A unified embedding for face recognition and clustering,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.
- [12] J. R. Pinto, J. S. Cardoso, and A. Lourenço, “Evolution, current challenges, and future possibilities in ECG biometrics,” *IEEE Access*, vol. 6, pp. 34 746–34 776, 2018.
- [13] M. Doudou, A. Bouabdallah, and V. Berge-Cherfaoui, “Driver Drowsiness Measurement Technologies: Current Research, Market Solutions, and Challenges,” *International Journal of Intelligent Transportation Systems Research*, vol. 18, no. 2, pp. 297–319, 2020.
- [14] S. Shah, D. Dey, C. Lovett, and A. Kapoor, “Airsim: High-fidelity visual and physical simulation for autonomous vehicles,” in *Field and Service Robotics*, 2017. [Online]. Available: <https://arxiv.org/abs/1705.05065>
- [15] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, “Carla: An open urban driving simulator,” in *Conference on robot learning*. PMLR, 2017, pp. 1–16.
- [16] Z. Greenstein, “NVIDIA’s AV Fleet Drives in the Data Center | NVIDIA Blog,” 2020. [Online]. Available: <https://blogs.nvidia.com/blog/2020/05/19/nvidia-fleet-drives-in-the-data-center/>
- [17] M. Doudou, A. Bouabdallah, V. Cherfaoui, and V. Charfaoui, “A Light on Physiological Sensors for Efficient Driver Drowsiness Detection System,” *Sensors & Transducers Journal, International Frequency Sensor Association (IFSA)*, vol. 224, no. 8, pp. 39–50, 2019. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-02162758>
- [18] Q. Massoz, T. Langohr, C. Francois, and J. G. Verly, “The ULg multimodality drowsiness database (called DROZY) and examples of use,” in *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*. IEEE, may 2016.
- [19] R. Ghoddoosian, M. Galib, and V. Athitsos, “A Realistic Dataset and Baseline Temporal Model for Early Drowsiness Detection,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2019-June, pp. 178–187, apr 2019. [Online]. Available: <http://arxiv.org/abs/1904.07312>
- [20] C. Fors, C. Ahlström, P. Sörner, J. Kovaceva, E. Hasselberg, M. Krantz, J.-F. Grönvall, K. Kircher, and A. Anund, *Camera-based sleepiness detection: final report of the project SleepEYE*. Statens väg-och transportforskningsinstitut, 2011.
- [21] C. H. Weng, Y. H. Lai, and S. H. Lai, “Driver drowsiness detection via a hierarchical temporal deep belief network,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10118 LNCS. Springer Verlag, nov 2017, pp. 117–133. [Online]. Available: [https://link.springer.com/chapter/10.1007/978-3-319-54526-4\\_9](https://link.springer.com/chapter/10.1007/978-3-319-54526-4_9)
- [22] S. Wahabi, S. Pouryayevali, S. Hari, and D. Hatzinakos, “On Evaluating ECG Biometric Systems: Session-Dependence and Body Posture,” *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 11, pp. 2002–2013, Nov. 2014.
- [23] R. Boussejot, D. Kreiseler, and A. Schnabel, “Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet,” *Biomedizinische Technik*, vol. 40, no. s1, pp. 317–318, 1995.
- [24] H. P. Silva, A. Lourenço, A. Fred, N. Raposo, and M. Aires-de-Sousa, “Check Your Biosignals Here: A new dataset for off-the-person ECG biometrics,” *Computer Methods and Programs in Biomedicine*, vol. 113, no. 2, pp. 503–514, 2014.
- [25] J.-P. Couderc, X. Xiaojuan, W. Zareba, and A. J. Moss, “Assessment of the stability of the individual-based correction of qt interval for heart rate,” *Annals of Noninvasive Electrocardiology*, vol. 10, no. 1, pp. 25–34, 2005.
- [26] L. Wolf, T. Hassner, and I. Maoz, “Face recognition in unconstrained videos with matched background similarity,” in *CVPR 2011*, June 2011, pp. 529–534.
- [27] J. A. Healey, “Wearable and Automotive Systems for Affect Recognition from Physiology,” Ph.D. dissertation, Massachusetts Institute of Technology, 2000. [Online]. Available: <https://dspace.mit.edu/handle/1721.1/9067>
- [28] P. Costa, A. Lourenço, C. Carreiras, D. Velez, P. Jorge, and A. Abrantes, “Multimodal biometrics in vehicles: a simulator study,” *IWBF2020*, 2020.
- [29] C. Jacobé de Naurois, C. Bourdin, A. Stratulat, E. Diaz, and J. L. Vercher, “Detection and prediction of driver drowsiness using artificial neural network models,” *Accident Analysis and Prevention*, vol. 126, pp. 95–104, 2019. [Online]. Available: <https://doi.org/10.1016/j.aap.2017.11.038>
- [30] J. F. Golding, “Predicting individual differences in motion sickness susceptibility by questionnaire,” *Personality and Individual Differences*, vol. 41, no. 2, pp. 237–248, 2006.
- [31] M. W. Johns, “A new method for measuring daytime sleepiness: the Epworth sleepiness scale,” *Sleep*, vol. 14, no. 6, pp. 540–545, dec 1991.
- [32] J. Horne and L. Reyner, “Vehicle accidents related to sleep: a review,” *Occupational and Environmental Medicine*, vol. 56, no. 5, pp. 289–294, 1999. [Online]. Available: <https://oem.bmj.com/content/56/5/289>
- [33] R. Delgado-Gonzalo, J. Parak, A. Tarniceriu, P. Renevey, M. Bertschi, and I. Korhonen, “Evaluation of accuracy and reliability of pulseox optical heart rate monitoring device,” in *2015 37th Annual International*

- Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 430–433.
- [34] J. Parak, A. Tarniceriu, P. Renevey, M. Bertschi, R. Delgado-Gonzalo, and I. Korhonen, "Evaluation of the beat-to-beat detection accuracy of pulseon wearable optical heart rate monitor," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 8099–8102.
- [35] L. A. Rodrigues, "Driver Drowsiness Detection with Peripheral Cardiac Signals," Master Thesis, Instituto Superior Técnico, 2021.
- [36] J. R. Pinto and J. S. Cardoso, "ECG Biometrics," in *Encyclopedia of Cryptography, Security and Privacy*, S. Jajodia, P. Samarati, and M. Yung, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2021, pp. 1–4.
- [37] J. R. Pinto, J. S. Cardoso, and A. Lourenço, "Deep Neural Networks For Biometric Identification Based On Non-Intrusive ECG Acquisitions," in *The Biometric Computing: Recognition and Registration*, K. V. Arya and R. S. Bhadoria, Eds. Boca Raton FL, United States: CRC Press, 2019, ch. 11, pp. 217–234.
- [38] J. R. Pinto and J. S. Cardoso, "An End-to-End Convolutional Neural Network for ECG-Based Biometric Authentication," in *IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, Tampa, FL, United States, Sep. 2019.
- [39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.
- [40] G. Chechik, V. Sharma, U. Shalit, and S. Bengio, "Large scale online learning of image similarity through ranking," *Journal of Machine Learning Research*, vol. 11, pp. 1109–1135, 2010.
- [41] F. Agrafioti, F. M. Bui, and D. Hatzinakos, "Secure Telemedicine: Biometrics for Remote and Continuous Patient Verification," *Journal of Computer Networks and Communications*, vol. 2012, p. 11, 2012.
- [42] A. Eduardo, H. Aidos, and A. L. N. Fred, "ECG-based Biometrics using a Deep Autoencoder for Feature Learning: An Empirical Study on Transferability," in *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2017)*, vol. 1, Porto, Portugal, Feb. 2017, pp. 463–470.
- [43] J. R. Pinto, J. S. Cardoso, A. Lourenço, and C. Carreiras, "Towards a Continuous Biometric System Based on ECG Signals Acquired on the Steering Wheel," *Sensors*, vol. 17, no. 10, p. 2228, 2017.
- [44] R. Labati, R. Sassi, and F. Scotti, "ECG biometric recognition: Permanence analysis of QRS signals for 24 hours continuous authentication," in *2013 IEEE International Workshop on Information Forensics and Security (WIFS)*. Guangzhou, China: IEEE, Nov. 2013, pp. 31–36.
- [45] R. J. A. Schijvenaars, "Intra-individual variability of the electrocardiogram: assessment and exploitation in computerized ECG analysis," Ph.D. dissertation, Erasmus University Rotterdam, 2000.
- [46] G. Lopes, J. R. Pinto, and J. S. Cardoso, "Don't You Forget About Me: A Study on Long-Term Performance in ECG Biometrics," in *Proceedings of Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA)*, Madrid, Spain, 2019.
- [47] K. Plataniotis, D. Hatzinakos, and J. Lee, "ECG Biometric Recognition Without Fiducial Detection," in *2006 Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference*. IEEE, Sep. 2006, pp. 1–6.
- [48] M. M. Tawfik, H. Selim, and T. Kamal, "Human identification using time normalized QT signal and the QRS complex of the ECG," in *2010 7th International Symposium on Communication Systems, Networks Digital Signal Processing (CSNDSP 2010)*, July 2010, pp. 755–759.
- [49] N. Belgacem, A. Nait-Ali, R. Fournier, and F. Bereksi-Reguig, "ECG Based Human Authentication using Wavelets and Random Forests," *International Journal on Cryptography and Information Security*, vol. 2, no. 2, pp. 1–11, Jun. 2012.
- [50] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artificial intelligence*, vol. 267, pp. 1–38, 2019.
- [51] B. Kim, R. Khanna, and O. O. Koyejo, "Examples are not enough, learn to criticize! criticism for interpretability," *Advances in neural information processing systems*, vol. 29, 2016.
- [52] C. Seibold, W. Samek, A. Hilsman, and P. Eisert, "Accurate and robust neural networks for security related applications exemplified by face morphing attacks," *arXiv preprint arXiv:1806.04265*, 2018.
- [53] T. Zee, G. Gali, and I. Nwogu, "Enhancing human face recognition with an interpretable neural network," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 0–0.
- [54] J. R. Pinto and J. S. Cardoso, "Explaining ECG Biometrics: Is It All In The QRS?" in *2020 International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE, 2020, pp. 1–5.
- [55] A. F. Sequeira, W. Silva, J. R. Pinto, T. Gonçalves, and J. S. Cardoso, "Interpretable biometrics: Should we rethink how presentation attack detection is evaluated?" in *2020 8th International Workshop on Biometrics and Forensics (IWBF)*. IEEE, 2020, pp. 1–6.
- [56] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine learning interpretability: A survey on methods and metrics," *Electronics*, 2019.
- [57] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," in *2014 European Conference on Computer Vision (ECCV)*, 2014, pp. 818–833.
- [58] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps," in *Workshop at International Conference on Learning Representations*, 2014.
- [59] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems*, 2017, pp. 4765–4774.
- [60] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning Important Features through Propagating Activation Differences," in *34th International Conference on Machine Learning*, 2017, pp. 3145–3153.
- [61] N. Kohlikeyan, V. Miglani, M. Martin, E. Wang, J. Reynolds, A. Melnikov, N. Lunova, and O. Reblitz-Richardson, "Pytorch captum," <https://github.com/pytorch/captum>, 2019.
- [62] M. Gomez-Barrero, C. Rathgeb, J. Galbally, C. Busch, and J. Fierrez, "Unlinkable and irreversible biometric template protection based on bloom filters," *Information Sciences*, vol. 370–371, pp. 18–32, 2016.
- [63] P. Drozdowski, N. Buchmann, C. Rathgeb, M. Margraf, and C. Busch, "On the Application of Homomorphic Encryption to Face Identification," in *2019 International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2019.
- [64] J. R. Pinto, J. S. Cardoso, and M. V. Correia, "Secure Triplet Loss for End-to-End Deep Biometrics," in *2020 8th International Workshop on Biometrics and Forensics (IWBF)*, Porto, Portugal, Apr 2020.
- [65] J. R. Pinto, M. V. Correia, and J. S. Cardoso, "Secure Triplet Loss: Achieving Cancelability and Non-Linkability in End-to-End Deep Biometrics," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 3, 2021.
- [66] J. R. Pinto and J. S. Cardoso, "Self-learning with stochastic triplet loss," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, Glasgow, United Kingdom, 2020.
- [67] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)*, 2017.
- [68] C. A. Corneanu, M. O. Simón, J. F. Cohn, and S. E. Guerrero, "Survey on rgb, 3d, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 8, pp. 1548–1568, 2016.
- [69] Z. Yu and C. Zhang, "Image based static facial expression recognition with multiple deep network learning," in *Proceedings of the 2015 ACM on international conference on multimodal interaction*, 2015, pp. 435–442.
- [70] B.-K. Kim, J. Roh, S.-Y. Dong, and S.-Y. Lee, "Hierarchical committee of deep convolutional neural networks for robust facial expression recognition," *Journal on Multimodal User Interfaces*, vol. 10, no. 2, pp. 173–189, 2016.
- [71] M. Liu, S. Li, S. Shan, and X. Chen, "Au-inspired deep networks for facial expression feature learning," *Neurocomputing*, vol. 159, pp. 126–136, 2015.
- [72] A. Dawel, L. Wright, J. Irons, R. Dumbleton, R. Palermo, R. O'Kearney, and E. McKone, "Perceived emotion genuineness: normative ratings for popular facial expression stimuli and the development of perceived-as-genuine and perceived-as-fake sets," *Behavior Research Methods*, vol. 49, no. 4, pp. 1539–1562, 2017.
- [73] I. A. T. A. Magalhães, "Feel My Heart: Emotion Recognition Using the Electrocardiogram," Master's thesis, Universidade do Porto, Porto, Portugal, 2021.
- [74] Y. L. Hsu, J. S. Wang, W. C. Chiang, and C. H. Hung, "Automatic ECG-Based Emotion Recognition in Music Listening," *IEEE Transactions on Affective Computing*, vol. 11, no. 1, pp. 85–99, 2017.



- [75] A. Goshvarpour, A. Abbasi, and A. Goshvarpour, "An accurate emotion recognition system using ECG and GSR signals and matching pursuit method," *Biomedical Journal*, vol. 40, no. 6, pp. 355–368, 2017.
- [76] C. Xiefeng, Y. Wang, S. Dai, P. Zhao, and Q. Liu, "Heart sound signals can be used for emotion recognition," *Scientific Reports*, pp. 1–11, 2019.
- [77] P. Sarkar and A. Etemad, "Self-Supervised Learning for ECG-Based Emotion Recognition," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 3217–3221.
- [78] S. Katsigiannis and N. Ramzan, "DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 98–107, 2015.
- [79] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, "Amigos: A dataset for affect, personality and mood research on individuals and groups," *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 479–493, 2021.
- [80] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42–55, 2012.
- [81] P. M. Ferreira, F. Marques, J. S. Cardoso, and A. Rebelo, "Physiological inspired deep neural networks for emotion recognition," *IEEE Access*, vol. 6, pp. 53 930–53 943, 2018.
- [82] S. Zepf, J. Hernandez, A. Schmitt, W. Minker, and R. W. Picard, "Driver emotion recognition for intelligent vehicles: A survey," *ACM Comput. Surv.*, vol. 53, no. 3, Jun. 2020. [Online]. Available: <https://doi.org/10.1145/3388790>
- [83] American Sleep Association / Reviewers and Writers, "What is Sleep and Why is It Important?" 2016. [Online]. Available: <https://www.sleepassociation.org/about-sleep/what-is-sleep/>
- [84] S. A. Keenan, M. Hirshkowitz, and H. Casseres, "Monitoring and staging human sleep," in *Principles and practice of sleep medicine*, 5th ed., M. Kryger, T. Roth, and W. Dement, Eds. St. Louis: Elsevier Saunders, 2011, ch. Normal Hum, pp. 16–26.
- [85] M. W. Johns, "A sleep physiologist's view of the drowsy driver," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 3, no. 4, pp. 241–249, 2000.
- [86] N. Lamond and D. Dawson, "Quantifying the performance impairment associated with fatigue," *Journal of Sleep Research*, vol. 8, no. 4, pp. 255–262, 1999.
- [87] Y. Liang, W. J. Horrey, M. E. Howard, M. L. Lee, C. Anderson, M. S. Shreeve, C. S. O'Brien, and C. A. Czeisler, "Prediction of drowsiness events in night shift workers during morning driving," *Accident Analysis and Prevention*, vol. 126, no. November 2017, pp. 105–114, 2019. [Online]. Available: <https://doi.org/10.1016/j.aap.2017.11.004>
- [88] H. J. Moller, L. Kayumov, E. L. Bulmash, J. Nhan, and C. M. Shapiro, "Simulator performance, microsleep episodes, and subjective sleepiness: normative data using convergent methodologies to assess driver drowsiness," *Journal of Psychosomatic Research*, vol. 61, no. 3, pp. 335–342, 2006.
- [89] L. N. Boyle, J. Tippin, A. Paul, and M. Rizzo, "Driver performance in the moments surrounding a microsleep," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 11, no. 2, pp. 126–136, 2008.
- [90] C. C. Liu, S. G. Hosking, and M. G. Lenné, "Predicting driver drowsiness using vehicle measures: Recent insights and future challenges," *Journal of Safety Research*, vol. 40, no. 4, pp. 239–245, 2009. [Online]. Available: <http://dx.doi.org/10.1016/j.jsr.2009.04.005>
- [91] E. Hoddes, V. Zarcone, H. Smythe, R. Phillips, and W. C. Dement, "Quantification of sleepiness: a new approach," *Psychophysiology*, vol. 10, no. 4, pp. 431–436, jul 1973.
- [92] T. H. Monk, "A Visual Analogue Scale technique to measure global vigor and affect," *Psychiatry research*, vol. 27, no. 1, pp. 89–99, jan 1989.
- [93] K. Kaida, M. Takahashi, T. Åkerstedt, A. Nakata, Y. Otsuka, T. Haratani, and K. Fukasawa, "Validation of the Karolinska sleepiness scale against performance and EEG variables," *Clinical Neurophysiology*, vol. 117, no. 7, pp. 1574–1581, 2006. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1388245706001428>
- [94] J. Vicente, P. Laguna, A. Bartra, and R. Bailón, "Drowsiness detection using heart rate variability," *Medical & biological engineering & computing*, vol. 54, no. 6, pp. 927–937, 2016.
- [95] E. Michail, A. Kokonozi, I. Chouvarda, and N. Maglaveras, "EEG and HRV markers of sleepiness and loss of control during car driving," *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS'08 - "Personalized Healthcare through Technology"*, pp. 2566–2569, 2008.
- [96] G. D. Furman, A. Baharav, C. Cahan, and S. Akselrod, "Early detection of falling asleep at the wheel: A heart rate variability approach," *Computers in Cardiology*, vol. 35, no. July, pp. 1109–1112, 2008.
- [97] A. Baharav, S. Kotagal, V. Gibbons, B. K. Rubin, G. Pratt, J. Karin, and S. Akselrod, "Fluctuations in autonomic nervous activity during sleep displayed by power spectrum analysis of heart rate variability," *Neurology*, vol. 45, no. 6, pp. 1183–1187, jun 1995.
- [98] L. K. McCorry, "Physiology of the Autonomic Nervous System," *American Journal of Pharmaceutical Education*, vol. 71, no. 4, p. 78, sep 2007. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1959222/>
- [99] Task Force of the European Society of Cardiology and The North American Society of Pacing Electrophysiology, "Heart Rate Variability," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.
- [100] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on intelligent transportation systems*, vol. 6, no. 2, pp. 156–166, 2005.
- [101] M. Miyaji, "Method of drowsy state detection for driver monitoring function," *International Journal of Information and Electronics Engineering*, vol. 4, no. 4, p. 264, 2014.
- [102] G. Rigas, Y. Goletsis, P. Bougia, and D. I. Fotiadis, "Towards driver's state recognition on real driving conditions," *International Journal of Vehicular Technology*, vol. 2011, 2011.
- [103] B.-G. Lee and W.-Y. Chung, "A smartphone-based driver safety monitoring system using data fusion," *Sensors*, vol. 12, no. 12, pp. 17 536–17 552, 2012.
- [104] J. Vicente, P. Laguna, A. Bartra, and R. Bailón, "Detection of driver's drowsiness by means of HRV analysis," in *2011 Computing in Cardiology*. IEEE, 2011, pp. 89–92.
- [105] J.-S. Wang, P.-C. Chung, W.-H. Wang, and C.-W. Lin, "Driving conditions recognition using heart rate variability indexes," in *2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. IEEE, 2010, pp. 389–392.
- [106] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, "Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm," *IEEE transactions on biomedical engineering*, vol. 58, no. 1, pp. 121–131, 2010.
- [107] S. Begum, M. U. Ahmed, P. Funk, and R. Filla, "Mental state monitoring system for the professional drivers based on heart rate variability analysis and case-based reasoning," in *2012 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2012, pp. 35–42.
- [108] S. Hu and G. Zheng, "Driver drowsiness detection with eyelid related parameters by support vector machine," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7651–7658, 2009.
- [109] M. Akin, M. B. Kurt, N. Sezgin, and M. Bayram, "Estimating vigilance level by using eeg and emg signals," *Neural Computing and Applications*, vol. 17, no. 3, pp. 227–236, 2008.
- [110] M. B. Kurt, N. Sezgin, M. Akin, G. Kirbas, and M. Bayram, "The ann-based computing of drowsy level," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2534–2542, 2009.
- [111] K. Karrer, T. Vöhringer-Kuhnt, T. Baumgarten, and S. Briest, "The role of individual differences in driver fatigue prediction," in *Third International Conference on Traffic and Transport Psychology*, Nottingham, UK. Citeseer, 2004, pp. 5–9.
- [112] L. Oliveira, J. S. Cardoso, A. Lourenço, and C. Ahlström, "Driver drowsiness detection: a comparison between intrusive and non-intrusive signal acquisition methods," in *2018 7th European Workshop on Visual Information Processing (EUVIP)*. IEEE, 2018, pp. 1–6.
- [113] C. S. Silveira, J. S. Cardoso, A. L. Lourenço, and C. Ahlström, "Importance of subject-dependent classification and imbalanced distributions in driver sleepiness detection in realistic conditions," *IET Intelligent Transport Systems*, vol. 13, no. 2, pp. 347–355, 2019.



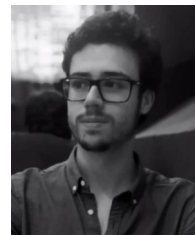
**TELMA ESTEVES** is enrolled in the master's program in biomedical engineering at the NOVA School of Science and Technology. During her academic education, she took courses on electrophysiology, signal analysis and processing, medical imaging, medical information systems, and electronics, among others. She is currently developing her M.Sc. thesis in collaboration with CardioID Technologies LDA and the INESC TEC's Visual Computing and Machine Intelligence (VCMI) research group. Her ongoing thesis work is related to the issue of drowsy driving, a fact that made her connect with the AUTOMOTIVE project. She aims to mitigate by analyzing current detection methods and proposing new solutions to recognize sleepy states from the driver's facial expressions, overcoming limitations such as inter-subject variability. Her main research interests include computer vision, machine learning, and image processing.



**PEDRO AMARO COSTA** received the B.Sc. in Communication Sciences from Universidade Nova de Lisboa and the M.Sc. degree in Multimedia Communication Systems at Universidade Lusófona de Humanidades e Tecnologias. He then went on to obtain a second B.Sc. degree at Instituto Superior de Engenharia de Lisboa in 2020, where he is currently undergoing the subsequent M.Sc. studies on Multimedia Software Engineering, combining software engineering, computer vision, and artificial intelligence. In 2019, he became a Researcher for AUTOMOTIVE - AUTOMATIC multimodal drowsiness detection for smart VEHICLES, where he had the opportunity to develop a virtual simulation prototype platform for driver monitoring and driving assistance, in close collaboration with partners CardioID Technologies. The system includes multimodal sensory acquisition for driver state classification and an automated driving module (L1) as a safety-critical measure for lane-path recovery, part of a collaborative vehicle control model between human/artificial agents. He currently works as Virtual Testing and Simulation Engineer at Continental, developing solutions based on state-of-the-art simulation environments for efficient development of Advanced Driving Assistance Systems and Autonomous Driving.



**JOÃO RIBEIRO PINTO (S'15)** received the M.Sc. degree in bioengineering (field of biomedical engineering) from the Faculdade de Engenharia da Universidade do Porto (FEUP), in 2017, with a thesis on the use of the electrocardiogram (ECG) for biometric recognition of vehicle drivers. He is currently pursuing his Ph.D. degree in electrical and computer engineering with FEUP. He is a Research Assistant with the Visual Computing and Machine Intelligence Research Group, INESC TEC, Porto, Portugal. His research has focused on contributing to make the ECG a viable and stronger biometric characteristic in realistic conditions. His Ph.D. studies are focused on using ECG and face, both acquired almost unnoticeably from vehicle drivers, to recognize them and continuously monitor their wellbeing. His research interests include biometrics, biosignals, pattern recognition, computer vision, and machine learning in general.



**LOURENÇO ABRUNHOSA RODRIGUES** is a machine learning engineer at CardioID Technologies Lda. He completed his MSc Degree in Biomedical Engineering from Instituto Superior Técnico (IST), Universidade de Lisboa, Portugal, in 2021. During academic training he devoted himself to signal processing, machine learning, and programming. Since joining CardioID, his work has been focused on the development of algorithms for heart rate variability, namely from wearables and off-the-person devices, as well as drowsiness detection in driving settings.



**PEDRO M. FERREIRA** received the degree in Biomedical Engineering from the Politécnico do Porto in 2009, the M.Sc. degree in Biomedical Engineering from the Faculdade de Engenharia da Universidade do Porto (FEUP) in 2012, and the Ph.D. degree in Deep Learning and Computer Vision in 2020, also from FEUP, where he was a Researcher at INESC TEC within the Visual Computing and Machine Intelligence Group. He is currently a Senior Data Scientist at FARFETCH. His main research interests include Computer Vision, Machine Learning, and Artificial Intelligence.



**INÊS ANTUNES** received the M.Sc degree in Bioengineering from the University of Porto in 2021. Thus, her academic education focused on a variety of themes and areas that encountered Bioengineering at some level. During the Biomedical Engineering specialty, subjects like signal and image processing, machine and deep learning, bionics, sensors, software engineering, etc, were more deeply studied and developed. Since 2018, Inês is part of the National Biomedical Engineering Students Association (ANEEB), being a member of the board and a department director since 2020. In the same year, she did a research internship at the Biomechanical Department of the University of Twente, working with raw EMG data and developing controllers for exoskeletons to help paraplegic patients. Furthermore, she has developed her master thesis at INESC-TEC, addressing the field of emotion recognition using the electrocardiogram and deep learning architectures.



**GABRIEL LOPES** received the M.Sc. degree in bioengineering from the University of Porto in 2019. His academic education was focused on a wide array of subjects, such as signal and image processing, machine learning, computer-aided diagnosis, sensors, electronics, and software engineering. He has addressed the topic of long-term performance in ECG Biometrics, including template update techniques with realist data. He is currently a consultant at Deloitte. He belongs to the Center of Excellence (CoE) of Low code/No Code.



**PEDRO M. JORGE** received the Licenciatura (five years), the M.Sc., and Ph.D. degrees in electrical and computer engineering from the Technical University of Lisbon in 1992, 1998, and 2007, respectively. Currently, he is a Coordinating Professor at the Instituto Superior de Engenharia de Lisboa (ISEL) from Politécnico de Lisboa (IPL), teaching courses in computer vision, machine learning, and mixed reality. He has been supervising several final projects of B.Sc. degrees and M.Sc. dissertations at ISEL/IPL and other institutions. He is the author and co-author of one European patent, one book chapter, journal papers, and several conference communications, and other scientific meetings, both in Portugal and abroad. He has been participating in several R&D projects, both with industry and funded by the National Science Foundation. He is currently the coordinator of the Computer Science and Multimedia Engineering B.Sc. Degree and of the Multimedia and Machine Learning R&D group from ISEL.



**PEDRO GAMITO** earned his Ph.D. from the University of Salford (UK) in Built and Human Environment in 2002, and the title of Aggregate in Rehabilitation by the University of Lisbon in 2012. Pedro is a Full Professor and the Director of the Computational Psychology Laboratory at the School of Psychology and Life Sciences and the director of the Research Centre - HEI-Lab – Digital Human-Environment Interaction Lab, at Lusófona University. Since 2004 he has been

developing, validating, and applying numerous virtual reality applications designed for the treatment of anxiety disorders and for the cognitive stimulation of patients with stroke, TBI, and Substance Use Disorders. His apps are being used daily in the two largest Portuguese psychiatric and rehab hospitals. He has participated in several projects in the field of eHealth and contributes to the most relevant professional associations within his line of work. In 2018, he was awarded the Lifetime Achievement Award of the International Association of CyberPsychology, Training & Rehabilitation.



**ANDRÉ LOURENÇO** received the M.Sc. and Ph.D. degrees in electrical and computer engineering from the Instituto Superior Técnico (IST), Universidade de Lisboa, Portugal. He is the Co-Founder and the CEO of the CardioID Technologies LDA. Since 2005, he has been Professor at the Electronics, Telecommunications and Computers Department, ISEL, where he is lecturing topics in information and signal processing and programming and supervising over 40 final year projects

and 13 master theses for post-graduate students. Apart from technical qualities, he has key business and entrepreneurial experience participating in several startups (Albatroz Engineering, Lusospace, and Minalytics), and R&D projects (including H2020). His scientific contributions include 12 book chapters, 11 peer-reviewed papers, 77 international conference papers, and a family of patents “Device and Method for Continuous Biometric Recognition Based on Electrocardiographic Signals”.



**ARNALDO J. ABRANTES** received the Licenciatura in Electronic Engineering and Telecommunications from University of Aveiro, in 1984, and the M.Sc. and Ph.D. diplomas, both in Electrotechnical and Computer Engineering, from the Technical University of Lisbon, in 1992 and 1998, respectively. He is currently a Coordinating Professor at Instituto Superior de Engenharia de Lisboa (ISEL), where he has been teaching (since 1985) courses in the areas of signal and image

processing, machine learning, system dynamics, agent-based modeling, and intelligent virtual environments. He has co-authored about sixty articles in the area of signal and image processing, participated in R&D projects, and co-supervised two doctoral theses, in the areas of video surveillance and multimedia applications.



**ANA F. SEQUEIRA** collaborates as Assistant Researcher at INESC TEC in the field of computer vision and machine learning. Ana has a Ph.D. in Electrical and Computers Engineering and a Licenciatura Degree (5-year) and a Master in Mathematics. Ana’s research comprises liveness detection techniques (for iris, face, and fingerprint); biometrics for border control; as well as facial analysis topics, such as emotion recognition, image compliance with standardization requirements, among others; and more recently interpretability/explainability of ML for biometrics. In the past, Ana worked at the University of Reading, UK, collaborating in EU projects related to the application of biometric recognition in Border Control (FASTPASS and PROTECT projects). Moreover, Ana had a short-term collaboration with the company Iris Guard UK researching the vulnerabilities and developing a proof-of-concept of a robust iris anti-spoofing measure for their EyePay® Technology. Ana led the construction of several biometric databases, managed biometric competitions, and has co-authored several research publications recognized by her peers with over 300 citations.

processing, machine learning, system dynamics, agent-based modeling, and intelligent virtual environments. He has co-authored about sixty articles in the area of signal and image processing, participated in R&D projects, and co-supervised two doctoral theses, in the areas of video surveillance and multimedia applications.





**JAIME S. CARDOSO** (SM'11) received the Licenciatura (five-year degree) in electrical and computer engineering, the M.Sc. degree in mathematical engineering, and the Ph.D. degree in computer vision from the University of Porto in 1999, 2005, and 2006, respectively. He is a Full Professor with the Faculty of Engineering, University of Porto (FEUP), where he has been teaching machine learning and computer vision in Doctoral Programs and multiple courses for graduate studies.

He is currently the Coordinator of the Centre for Telecommunications and Multimedia, INESC TEC, and the Co-Founder and the Co-Leader of the Breast Research Group and the Visual Computing and Machine Intelligence Group. He is also the Co-Founder of ClusterMedia Labs. He has co-authored more than 250 papers, 80+ of which were in international journals. His research can be summed up in three major topics: computer vision, machine learning, and decision support systems. He has been a recipient of numerous awards, including the Honorable Mention in the Exame Informática Award 2011, in the software category, for the project "Semantic PACS" and First Place in the ICDAR 2013 Music Scores Competition: Staff Removal (task: staff removal with local noise), August 2013. His research results have been recognized both by his peers, with 5900+ citations to his publications, and the advertisement in the mainstream media several times.



**ANA REBELO** was born in Porto, Portugal, in 1985. She graduated in Mathematics Applied to Technology at the School of Sciences of the University of Porto, Portugal, in 2007. She received the M.Sc. degree in Mathematical Engineering from the School of Sciences of the University of Porto, Portugal, in 2008. That year, she started her Ph.D. studies at the School of Engineering, University of Porto, Portugal. She has been working since 2007 as a researcher at INESC TEC, an R&D

institute affiliated with the University of Porto, in the Visual Computing and Machine Intelligence Group (VCMi). Ana Rebelo was a Project Member of one FCT (Foundation of Science and Technology - Portugal) research project in the area of optical music recognition (OMR). She is a Researcher at INESC TEC. Her main research interests include computer vision, deep learning, biometrics, digital forensics, and document analysis.

• • •