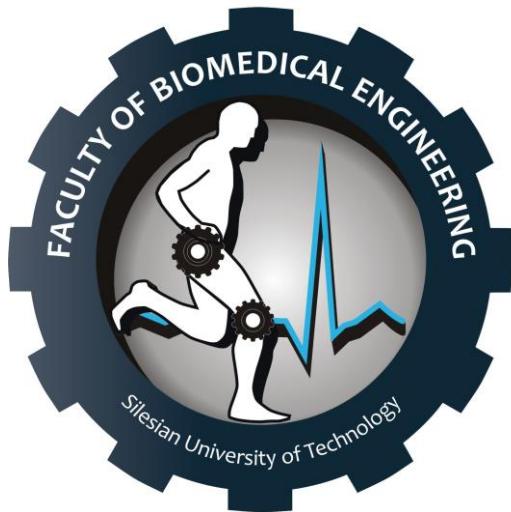


Faculty of Biomedical Engineering
SILESIAN UNIVERSITY OF TECHNOLOGY



MASTER'S THESIS

Development of a classification algorithm for basic and distracting driving activities, based on multimodal signals and (deep) machine learning methods.

Justyna Konior
Major: Biomedical Engineering
Specialization: IT in medicine

SUPERVISOR
Dr inż. Rafał Doniec

Table of contents

| | |
|---|----|
| 1. <i>Introduction</i> | 5 |
| 1.1 The Problem | 6 |
| 1.2 Aim of the work | 7 |
| 1.3 Outline of the Thesis | 7 |
| 1.4 Overview of the area in related literature | 7 |
| 1.4.1 Evaluating driver cognitive distraction by eye tracking: From simulator to driving [1] | 7 |
| 1.4.2 Activity Classification in Independent Living Environment with JINS MEME Eyewear [8] | 8 |
| 1.4.3 Rank Pooling Approach for Wearable Sensor-Based ADLs Recognition [30] | 10 |
| 1.4.4 Towards Reading Trackers in the Wild: Detecting Reading Activities by EOG Glasses and Deep Neural Networks [16] | 12 |
| 1.4.5 EOG-Based Eye Movement Classification and Application on HCI Baseball Game [22] | 12 |
| 1.4.6 A novel eye movement detection algorithm for EOG driven human computer interface [25] | 14 |
| 1.4.7 Hybrid mean fuzzy approach for attention detection (2021) [28] | 14 |
| 1.5 Challenges in Human Activity Recognition | 15 |
| 2. <i>Materials and Methods</i> | 17 |
| 2.1 Hardware and software | 17 |
| 2.1.1 JINS MEME Academic Pack Smart Glasses | 17 |
| 2.1.1.1 EOG signal | 18 |
| 2.1.1.2 JINS MEME Academic Pack | 20 |
| 2.1.1.2.1 Connecting the glasses | 20 |
| 2.1.1.2.2 Data Acquisition Control | 21 |
| 2.1.2 Driving simulator | 22 |
| 2.2 Data collection | 23 |
| 2.2.1 Study participants | 24 |
| 2.2.2 Scenarios | 24 |
| 2.2.2.1 Basic driving activities | 24 |
| 2.2.2.2 Distracting driving activities | 27 |
| 2.2.3 Obtained data | 27 |
| 2.3 The algorithm | 28 |

| | | |
|---------|---|----|
| 2.3.1 | Data format, loading and labeling | 29 |
| 2.3.1.1 | Data format | 29 |
| 2.3.1.2 | Data loading and labeling | 30 |
| 2.3.2 | Preprocessing | 30 |
| 2.3.2.1 | Filtering the accelerometer signal | 31 |
| 2.3.2.2 | Baseline drifts removal and filtering the EOG signals . | 32 |
| 2.3.2.3 | Resampling | 33 |
| 2.3.3 | Neural Network architecture | 34 |
| 3. | <i>Results and Discussion</i> | 39 |
| 3.1 | Evaluation methods | 40 |
| 3.2 | Accuracy and Loss while training | 41 |
| 4. | <i>Summary</i> | 47 |
| | <i>Bibliography</i> | 51 |

List of Figures

| | | |
|------|--|----|
| 2.1 | JINS MEME Smart Glasses [7] | 18 |
| 2.2 | JINS MEME Smart Glasses scheme [23] | 18 |
| 2.3 | Scheme of the eye. [21] | 19 |
| 2.4 | Jins MEME Academic Pack software [24] | 20 |
| 2.5 | Driving simulator | 22 |
| 2.6 | Roundabout scenarios. | 25 |
| 2.7 | Crossroad scenarios. | 25 |
| 2.8 | Parking scenarios. | 26 |
| 2.9 | Scheme of the methodology. | 28 |
| 2.10 | 3-axis accelerometer signal filtering of going straight at a roundabout. | 31 |
| 2.11 | Raw EOG signal of going straight at a roundabout. | 32 |
| 2.12 | Preprocessed accelerometer signal of going straight at a roundabout. | 33 |
| 2.13 | Preprocessed EOG signal of going straight at a roundabout. | 34 |
| 2.14 | Network architecture. | 35 |
| 2.15 | Convolution operation. [26] | 36 |
| 3.1 | Distribution of activity samples in the training and testing sets. | 39 |
| 3.2 | Accuracy and loss graph for validation and training . | 41 |
| 3.3 | Scatter plot of the training and testing sets according to predicted labels. | 42 |
| 3.4 | Confusion matrix. | 44 |

List of Tables

| | | |
|-----|---|----|
| 1.1 | Scenarios for the activities of Daily Living. | 9 |
| 2.1 | Sensor parameters of 2 possible measurement modes. | 21 |
| 2.2 | Possible values of sensor parameters. | 21 |
| 2.3 | Secondary driving activities scenarios. | 27 |
| 2.4 | The format of acquired data. | 29 |
| 2.5 | Driving activities labels. | 30 |
| 3.1 | Evaluation scores of driving activities classification. | 43 |
| 3.2 | Evaluation scores of the type-based classification. | 45 |

Acronyms

ADLs Activities of daily living. [9, 10](#)

BADLs Basic Activities of daily living. [9](#)

CNN Convolutional Neural Network. [34](#)

EOG Electrooculography. [8, 9](#)

GPS Global Positioning System. [5](#)

IADLs Instrumental Activities of Daily Living. [9](#)

IMU Inertial Measurement Unit. [47](#)

LSTM Long short-term memory. [11, 12](#)

MOSFET metal-oxide-semiconductor field-effect transistor. [23](#)

NHTSA National Highway Transportation and Safety Administration's. [5](#)

OKR Optokinetic Reflex. [8](#)

PCA Principal Component Analysis. [41](#)

RDS(on) Drain-source on resistance. [23](#)

ReLU Rectified Linear Unit. [36](#)

TVMV Time Varying Mean Vector. [11](#)

VOR Vestibular-Ocular Reflex. [8](#)

WORD Wojewódzki Ośrodek Ruchu Drogowego. [24](#)

Abstract

The aim of this paper is the creation of a deep learning-based system for identifying driver activities. This study is entirely performed on the basis of the signal acquired from JINS MEME Smart Glasses equipped with a set of EOG sensors, and a set of accelerometer and velocity sensors. A 1D CNN model was developed, which then uses the preprocessed EOG and accelerometer readings to categorize the subsequent driving behaviors: primary activities - crossroad turning left, right and going straight, angle parking left and right, parallel parking left and right, perpendicular parking left and right, roundabout turning left, right and going straight and secondary activities - bending, drinking, eating and turning back. The results obtained were divided into three categories. For the binary classification between secondary and primary activities, an accuracy rate of 99.5 % was achieved. An analysis of classifications for a specific set of activities, comprising crossroad, parking, roundabout, and secondary activities, produced accuracy values of 97.9%, 96.8%, 97.4% and 99.5%, respectively. Finally, recognizing each behavior independently demonstrated satisfactory accuracy, precision, and f1 score at the level of 80%. The secondary activity of drinking was predicted with the greatest degree of accuracy, while primary parallel parking on the left side and perpendicular parking on the right side with the smallest.

1. Introduction

To drive safely, one must be sufficiently aware of their surroundings, constantly pay attention to the road and the traffic, and be sufficiently awake to react to unforeseen circumstances [18]. Driver fatigue and inattentiveness are still key contributors to fatalities and serious injuries on the roads, and they continue to be issues for international road safety measures, since it impacts not only the driver but also everyone else on the road.

According to the [National Highway Transportation and Safety Administration's \(NHTSA\)](#) most recent figures, distracted driving claimed the lives of over 3,000 individuals in 2020 in America, accounting for 8.1% of all car crash fatalities. In addition, around 324,000 people were injured due to motorists' inattentiveness [29]. There are no exact data concerning Europe, however the Directorate General for Transport suggests that it potentially accounts for more than 25% of all fatal car accidents [11].

The tasks that are directly related to maneuvering the vehicle are called primary driving activities. The term "distraction" refers to a specific sort of inattention that happens when drivers divert their attention away from the primary tasks of navigating the vehicle to focus on another activity. These diversions can come from more commonplace activities like talking to other passengers and eating, as well as more technological ones like using cell phones and [GPS](#) systems. These diverting activities can have various effects on drivers, and according to [NHTSA](#) they can be divided into the following categories:

- Visual distraction: Activities that force the motorist to look away from the road
- Manual distractions: Activities that require the driver to remove their hand from the wheel
- Cognitive distraction: Activities that demand the driver to shift their mental focus away from the driving task

Even though there are three distinct categories, many activities can be classified into more than one of them. For instance, using the [GPS](#) device requires engaging in all of the listed distracting activities, otherwise described as secondary driving activities. The cognitive distraction, which takes place in the driver's brain, is the most challenging to identify. This phenomenon is also known as 'looking but not seeing' [32].

The attentional demands of the distracting work and the prevalence of multitasking among drivers make up the two fundamental elements of the distracted driving safety issue. The task demand is the total amount of visual, physical, and cognitive resources required to execute the activity. The second issue is the frequency with which drivers undertake the task. Even a brief task that is carried out regularly might create a safety issue.

According to the study [19] the findings suggest that activities requiring the driver to look away from the road and/or conduct manual tasks greatly enhance the likelihood of a collision. The risk of a driving accident is increased by using a handheld phone by a ratio of 2.05, particularly when dialing (x12) and messaging (x6). The lengthened period of time spent looking away from the road is also a significant influence. According to certain research, looking away from the road for longer than two seconds considerably increases the probability of safety-critical occurrences [4]. Following the research, the US Department of Transportation advises against doing this and glancing away repeatedly over a period of 12 seconds while operating a motor vehicle.

1.1 The Problem

There are a number of steps that can be taken to raise the standard of the transportation system, from governmental regulations to driver performance. There are a number of campaigns that raise awareness about distracted driving and its consequences. They mainly concentrate on using electronic devices, but even for briefer tasks, driving during that period diminishes attention to the road, which can be risky and result in frequent rear-end collisions. The cost of such campaigns is significant, whilst "evaluations of distracted driving awareness campaigns have had mixed results." [2] According to some research, there saw a significant decrease in distracted driving after the intervention, but this didn't last during a follow-up assessment six months later. In the spirit of Vision Zero, roads, streets, and vehicles must be considerably more tailored to human capacity and tolerance. It envisions a future, in which there are no victims of traffic accidents [33]. Although Vision Zero covers a much larger range of issues, one method to enhance traffic safety and driving performance is by developing sophisticated monitoring and assistance systems for motorists that analyze the driver and their surrounding environment. Recently, a lot of research has been done on this topic. Some studies suggest that utilizing a reliable driver behavior monitoring system can reduce accidents by 10% to 20% [34]. The purpose of the driver monitoring systems is to assess the driver's level of attention in order to take the appropriate remedial action to maintain driving safety. Some automakers have already included these systems into their automobiles, but more dependable and quick-response monitoring systems are still required. It is essential to make a distinction between the many types of distraction and weariness in order to develop an in-vehicle technology that is appropriate for the detrimental effects brought on by each source of drowsiness or distraction. Thus, one of the most crucial challenges is the identification of driving behavior.

1.2 Aim of the work

The aim of this paper, is a verification of the thesis that EOG and accelerometer signals are sufficient to identify primary and secondary driving activities. The goal is to create a deep learning-based system for identifying driving behaviors. Using the physiological data collected, the system should be able to separate primary and secondary driving actions and correctly identify the type of task performed.

1.3 Outline of the Thesis

The remaining sections of this paper are organized as follows. Section 1.4 serves as an overview of available literature concerning similar topics. Section 2.3 presents the hardware along with its functionality, describes the data collection technique and the methodology of activity detection algorithm. Section 3 is related to experimental results and discussion. Section 4 concludes the research study.

1.4 Overview of the area in related literature

A variety of sensing modalities have been employed to identify driving behavior. The literature defines three categories: vehicle control data, visual data and physiological data [10]. The first type identifies an activity based on information about the vehicle's dynamics, such as the steering wheel's or pedal's position, or the throttle hold rate [6]. Some studies show that distracted drivers regularly increase the separation distance or travel at a higher speed than usual [17]. The visual data classification makes use of photos and videos to infer the behavior of the driver by capturing the driver's eye movements, head movements, or body motions [6]. It was mostly utilized while detecting secondary driving activities [34]. The last type utilizes the physiological signal, such as the heart rate or brain activity, to assess the driving activity. These signals are mainly utilized to identify driver weariness as the primary source of attention and are frequently used for real-time driver condition monitoring [35].

A survey of related literature that served as a writing guide for this thesis is to be presented in this section.

1.4.1 Evaluating driver cognitive distraction by eye tracking: From simulator to driving [1]

The article, similarly, to this paper, concentrates on analysing eye movements as a way of recognizing the driver's behavior. The authors of the article focused on the analysis of the cognitive distraction of drivers, which is described as the mental workload. Since it is not an activity that involves additional movements, rather a thought process that does not concern the main activity, which is driving, the task is not trivial and requires a good method for data acquisition and processing. The authors proposed a method for simulating involuntary eye movement by combining the

vestibuloocular reflex model with the optokinetic response, which is then compared with the observed eye movements. The difference between those values is assumed to be a measure of the level of cognitive distraction. For this purpose, the authors decided to use head and eye tracking, specifically Smart Eye for eye movements and Fastrak electromagnetic tracker for head movements (in driving simulator room) and EyeSeeCam equipment, which is a head-mounted device for collecting eye and head movement (in natural condition). Nine participants (5 men and 4 women) were asked to drive for 3 minutes first without any distractions followed by the addition of auditory simulations, in both real environment and on a simulator. In the second task, every 3 seconds the drivers heard a number and had to decide whether it matches the previous one. They answered by pressing an appropriate button. In such a task the participants not only had to focus on the provided number but also remember the previous one.

In the preprocessing part, the authors used filtering for saccades, removed blink points and used the Kalman Filter for noise reduction. For parameter identification samples of 10 second in length were taken from the experiment without additional mental workload and such sets were used to simulate eye movements in **VOR + OKR** models.

The performance was measured by comparing the mean-squared error between a 30s time windows of simulated and real signal. Next, to determine the effect of cognitive distraction on vertical eye movement, which is the direction in which the involuntary eye movement while driving usually happens due to the vibration from the vehicle, the authors performed a t-test analysis.

The results showed that the eye movement was abnormal under cognitive distraction yielding the results of $t=-5.25$, $p=0.00038$ on the simulator and $t=-5.48$, $p < 0.005$ in real life environment.

1.4.2 Activity Classification in Independent Living Environment with JINS MEME Eyewear [8]

The article described in this section utilizes the same equipment, which is to be used in this thesis. Therefore, it is gainful to study its performance in activity detection.

The goal of the article according to the authors is:

- examination of the sensors embedded within JINS MEME glasses as a source of data to be used in the classification of activities of daily living,
- development of a method for calculating information about eye and head movements, in addition to basic characteristics of **EOG** and motion signals, to be used as attributes in the classification process,
- demonstration of techniques for mitigating the problem of imbalanced classes.

The aim of the work was a classification of various activities described as [Activities of daily living \(ADLs\)](#). They were divided into two categories, namely [Basic Activities of daily living \(BADLs\)](#) and [Instrumental Activities of Daily Living \(IADLs\)](#). The first ones are associated with activities that are necessary in everyday live, whereas the latter describes more complex tasks important for independent living.

| Activities of Daily Living | |
|----------------------------|----------------|
| Basic | Instrumental |
| Bathing | Cooking |
| Dressing | Exercise |
| Feeding | Housework |
| Grooming | Typing/Writing |
| Toilet Use | Watching TV |
| Walking | |

Tab. 1.1: Scenarios for the activities of Daily Living.

In the study, twenty six participants were divided into 7 pairs and the rest stayed as individuals. They were asked to spend two hours at The Smart Condo, which is a smart home environment at the University of Alberta and perform specified activities. Additionally, a video recording of all the procedures was acquired to determine the ground truth of participant activities. The JINS MEME smartglasses were worn by 12 participants.

The Data acquired from the eyewear was cut into 5.6 s overlapping windows. For the preprocessing part, low pass filter as well as bandpass filter were utilized. The first one allowed for noise reduction, whilst the latter helped to eliminate static signal components.

Additionally, higher order attributes were extracted from the [EOG](#) signal - blinks, horizontal and vertical saccades. From accelerometer and gyroscope the head movements about Euler angles were determined. In order to remove signal noise without disturbing the overall form of physical movements, a moving average over a duration of 0.05 s was applied. On top of that, a new attribute that corresponds to the nature of used electrodes was introduced. Since dry electrodes do not adhere to the skin surface, in the [EOG](#) signal abnormally large peaks appear in response to facial movements or eyewear adjustments. To minimize the influence of those disturbances, the time of occurrence of these peaks is calculated and presented as a percent of the whole sample's length. In the paper, four different machine-learning classifiers were used. However, before feeding the data into the classifiers, class balancing problem had to be addressed. Due to different nature of activities, their duration varied greatly. To avoid the problem of some classes being more favourable than others, the classes that

yielded in less than 50 analysis windows were removed from the dataset. For the other classes a technique called SMOTE (Synthetic Minority Over sampling Technique) was used. Finally, the classification could be performed. The four classifiers were:

- K Nearest Neighbor
- PART
- Random Forest
- Sequential Minimal Optimization (SMO)

The best results were obtained using the Random Forest giving the classification accuracy for the activities, using cross validation, between 87.15% and 100% giving the overall activity equal to 93.03%.

1.4.3 Rank Pooling Approach for Wearable Sensor-Based ADLs Recognition [30]

The topic of this study is the wearable-based recognition of [ADLs](#), which are made up of a number of time-dependent, repetitive, and brief movements. For the purpose of recognizing ADLs, the authors presented a two-level hierarchical model. The data collection process is split into two parts. First, atomic activities are performed, focusing on 61 distinct activities involving 8 subjects. By the term atomic activities it is understood that an action requires only one short-term movement. Exemplary actions used in this study are:

- Take off jacket
- Gargle
- Close tap water
- Rub hands

Additionally, six participants were subjected to data collection for seven composite activities:

- Brushing Teeth
- Cleaning Room
- Handling Medication
- Preparing Food
- Styling Hair
- Using Phone

- Washing Hands

These activities were composed of a series of atomic actions.

The data acquisition was performed using five separate sensory modalities

- the LG G5 smartphone inserted into the subject's left front pocket on jeans,
- Huawei watch,
- JINS MEME glasses.

The smartphone was responsible for recording bodily motion using a 3-axis accelerometer, a gravity sensor, a gyroscope, a linear accelerometer (all captured at 200 Hz), and a magnetometer (100 Hz). The watch was worn on the left arm and acquired accelerometer and a gyroscope signals sampled at 100 Hz. Finally, the glasses provided 3-dimensional accelerometer data sampled at 20 Hz.

Instead of analysing the raw signal, a two level recognition is proposed. The second level recognizes ADLs by performing rank pooling on the recognition scores of atomic activities that were generated by the previous level utilizing feature extraction based on the codebook technique.

Features that symbolize the properties of the associated sequence for each sensor are derived. the attributes for each sensor are combined to create a single high-dimensional attribute. This method transforms a variety of atomic activity instances from training data into high-dimensional characteristics. Then, a recognition model is constructed using these high-dimensional features throughout the model training/testing phase. Using the computed atomic scores of the atomic activities altered by applying a vector-valued function obtained using [Time Varying Mean Vector](#), the ranking function learns how the vector changes over time. The authors implemented a point-wise rank pooling technique by resolving the constrained minimization problem and looking at the direct mapping from the input vector to its time variable based on the linear parameters. Additionally, to tackle the non-linearity issue of the input data, Hellinger and power transformations on [TVMV](#) were employed .

As a result, rank pooling outperforms other pooling methods resulting in an improvement of 5-13% over the other widely-used approaches including max, average, Hidden Markov Models, and [LSTM](#) pooling for the recognition of composite activities obtaining average accuracy, f1 score and weighted f1 score equal to 61.48, 60.91, 60.95, respectively for the 124-dimensional feature vector created on the concatenated Hellinger and power transformation of forward rank pooling feature vector. The outcomes of combining rank pooling with average and maximum pooling increased those scores to 63.64, 63.65, 63.48.

1.4.4 Towards Reading Trackers in the Wild: Detecting Reading Activities by EOG Glasses and Deep Neural Networks [16]

The purpose of this paper was an accurate detection of reading activity in conditions that resemble real life situations, in which people read. The authors state that most other research are performed in a fully controlled environment, while people tend to read in a variety of situations from relaxed at home, through in a busy public transportation to while waiting in a line.

The data consisted of EOG signal and accelerometer data from JINS MEME Glasses. & participants delivered over 980 hours of real life reading signal. Precisely 22 hours of controlled reading, 427 hours of natural reading, 156 hours of social interactions and 375 hours of other activities. The participants were asked to add annotations to every minute of the data labelling it as one of the three: "reading", "talking", and "other activities". Such prepared signals were then subjected to three different classification algorithms. Namely, manual feature extraction with Supported Vector Machine and classification of raw signal using Convolutional Neural Network and [Long short-term memory \(LSTM\)](#). The second two approaches, additionally to recorded movement, computed the blink speed and inputted it into the network.

400 frames of 6 sensors' values were mapped and input to the networks. The CNN architecture consisted of two sets of convolution and Pooling layers, followed by two Fully connected layers. The first pair utilizes a filter of size 12x1 and max pooling with stride equal to 3. The second one filtering 11x1 and stride 2. In all layers a drop out with rate 0.5 is employed. Afterwards, the fully connected layer reduces the size to 100 units and finally, utilizing rectified linear units, to one of two states: reading or not reading. The parameters of [LSTM](#) algorithm were: number of nodes in the hidden layer - 32, loss function - Adam, learning rate 0.002, dropout - 0.1. The output followed sigmoid function and classified an activity as reading or not reading.

The results demonstrate that in natural reading environment deep learning techniques perform much better than classic approaches, which yield better results in controlled environments. Nevertheless, the classification accuracy decreased in all of the proposed methods. The user dependent natural reading accuracy was best for [LSTM](#) - 73.8%, whereas the user independent for CNN with 69.6%.

1.4.5 EOG-Based Eye Movement Classification and Application on HCI Baseball Game [22]

The paper focuses on the investigation of eye movement tracking using electrooculography signal and its use in human-machine systems. The authors proposed eight directional eye movement classification algorithm based on analysis of saccades, therefore they focus greatly on denoising and processing the EOG signal.

The eight eye movement to be determined in this study are:

- look-up
- look-left
- look-down
- look-right
- look-up-and-left
- look-up-and-right
- look-down-and-left,
- look-down-and-right

The last four are more complex and are composed of two previous activities combined.

The raw signal was acquired with the use of EOG Mindo device from National Chiao Tung University Brain Research Center. There is a set of four electrodes placed around eyes and a reference electrode on the forehead, which result in two horizontal and two vertical signal readings. The data acquired from the sensors is preamplified and denoised thanks to the use of an instrumentation amplifier controlled by a microcontroller program. Using the moving average, the power line noise is reduced and finally, the digitized signal with a sampling rate of 256 Hz is forwarded to the computer via Bluetooth. Once the buffer is full, the obtained signal is passed to the software and prepared for classification.

According to the authors, one of the most crucial steps in the classification is detection and elimination of blinks, since they are the main source of misclassification. Due to the nature of blinks, which have higher amplitude than saccades, they can be easily distinguished by their amplitude. Therefore, by using a peak detection algorithm, they can be filtered out from the signal.

The same algorithm allows for recognizing the eye movement. By analysing the peak upper and lower threshold value, the classification to one of the eight classes can be performed. When the eyes moves straight in one direction, there only one channel becomes activated, however for the complex movements both channels become activated at the same time.

The effectiveness of the proposed method was examined in three set ups - first, a participants had to follow cues appearing on the screen, second, they were allowed to move their sight independently and third, they repeated the previous experiment on a smaller screen.

The first experiment yielded almost perfect results, providing the correct rate between 94 and 100%. The experiment without cues also produced satisfactory yet less accurate results with the lowest score of 90% for three of the movements. The last test, performed on a small scale device, produced results that were mostly above 90%.

The tests performed on the target activity revealed that the application works with 90% mean accuracy, after performing 5 HCI Baseball Games.

1.4.6 A novel eye movement detection algorithm for EOG driven human computer interface [25]

A similar study to 1.4.5 concerning eye movement detection and EOG-based human computer interface was conducted, however the authors proposed extracting linear predictive coding cepstrum (LPCC) coefficients as feature vectors and used them for pattern matching.

The preprocessor, the feature extracting unit, and the training/recognizing module make up the three basic pieces of the proposed eye movement detection algorithm. The first block consists of such steps as endpoint detecting, band-pass filtering, frame blocking, and windowing. The second block contains extracting linear predictive coding (LPC) and LPCC coefficients and finally the last segment aims at recognizing the movement based on dynamic time warping (DTW).

In the preprocessing step, the authors used the entropy-based algorithm to be able to distinct the eyeball and eyelid movement from the remaining signal. Following, a band-pass filter was applied to remove the base-line shift, power frequency interference as well as any other undesirable signal. In the next step, a frame blocking method is used to guarantee the stationarity of noisy EOG fragments and extract additional feature parameters. Finally, each individual frame is subjected to windowing as to reduce the signal discontinuities at the start and end of each frame.

The feature extracting unit includes calculating LPC and LPCC coefficients of the previously prepared data.

Finally, the template matching method is used for the eye movement classification. To reduce the problem of different amplitudes and durations of the movements, a method called dynamic time warping was introduced, which utilizes optimal alignments between sample points in the reference and testing time series.

The algorithm was tested on 10 users, six males and four females, with their recognition rate averaged. The actions to be recognized are: looking up, down, left, right, blinking twice, three times and four times. The results show that the algorithm works best when the signal-to-ratio is higher. In a low SNR environment it produced the worst results equal to 62.7% for 0 dB and 82.6% for 10 dB. However, it performed much better in comparison to the energy-based algorithms, that were not able to recognize the pattern. In a medium SNR environment the recognition rate is equal to 90.4% is achieved, whereas in big SNR environment it reached 94.5 %. As a result, the suggested spectral entropy-based eye movement detection technique is more reliable in a variety of noisy settings.

1.4.7 Hybrid mean fuzzy approach for attention detection (2021) [28]

This study focuses on detecting the driver's level of attention as one of the crucial parts of preserving the road safety. The data was acquired in the form of electroencephalography (EEG) signal and aimed to determine the attention-related characteristic. In order to specifically predict the driver's degree of attention during driving, this

study offers a hybrid mean fuzzy (HMF) system that combines the simple averaging (Mean) approach with fuzzy system. The former is used for data analysis and pre-classification process, whereas the latter in the step of data classification. Additionally, the driver warning system hardware implementation was suggested. A total of 150 adults between the ages of 18 and 30 were subject to the study. They were divided into three equal groups and performed assigned tasks. The first two units, treated as the training and testing data, were asked to drive the driving simulator in normal driving conditions without any stimulation and while listening to the radio. The last one served as control data. Half of the group was asked to watch the same video twice, which lowered their interest the second time, whereas the rest watched two different videos. Since each person has individual traits that might significantly demonstrate attention impairment, the goal of this study was to minimize the variability gap between the EEG signal and the fuzzy method.

The BIOPAC device, whose impedance is guaranteed to be under $5\text{ k}\Omega$, was used to capture the EEG data. Its bandpass filter ranges from 4 to 30 Hz, and its sampling frequency is 500 Hz. To minimize artifacts in the recorded EEG data, the BIOPAC system is configured to a 50 Hz notch filter. In the pre-processing step artifacts were removed from the EEG signal using visual inspection and MATLAB coding. Data analysis was performed with the use of the mean approach, which implied calculating the amplitude of complex waveform of a visually chosen time window, containing the desired event-related potential (ERP) component, for the extracted averages of 500 responses. The average of all successful trials was calculated in groups. The obtained signal is then subjected to a fuzzy system, specifically a fuzzy rule-based method, where it, together with the accident score, outputs the level of attentiveness of the drivers.

To measure the effectiveness of the proposed method, the authors introduced the Attention Degradation Scale (ADS), which allowed for assigning numerical values to one of three states 1 being inattentive and 3 attentive. The results of the experiment showed that without the radio stimulation, the driver's attention degraded, which led to raising the possibility of an accident, while with the stimulation helped with focusing and therefore improved the driver's performance.

1.5 Challenges in Human Activity Recognition

Many machine learning techniques have been used in the recognition of human behavior, though they still face a lot of technological difficulties. Some of the issues are common to other pattern recognition domains, while others are specific to sensor-based activity recognition and necessitate dedicated solutions for real-world applications. The problems that need to be faced are:

- The number of labeled data samples for training and evaluation To develop a reliable system able to recognize activity patterns, a large dataset is required. First of all, a certain amount of data is needed for the neural network to distinguish important features and second of all, it is essential to provide a wide variety of signals. However, gathering and annotating data on sensory activity

is costly and time-consuming. The signals gathered from a single subject may change over time, just as different persons have varied approaches to performing the same activity. Additionally, the nature of activity plays an important role in the accuracy of its acquisition. The collection of data from unexpected events is highly challenging. In such cases a class imbalance problem may arise.

- A difficulty in separating activities As stated above, the same task can be carried out in different manners, what makes them person-dependent. When comparing similar activities for more than one person, differences that are first discernible may become indistinguishable or highly misclassified. The problem escalates when novel behaviors appear. Another obstacle might arise while dealing with composite activities, which can be broken down into a series of subsequent tasks that must be accurately segmented in order to move forward with the classification. Furthermore, separating concurrent activities, which imply that more than one activity is being completed at a time, demonstrates a significant issue.
- Feature extraction A very important step in a classification algorithm is the feature extraction, on the basis of which the activities are separated. However, it might be challenging to develop distinctive elements to depict different activities due to inter-activity similarities, what makes feature extraction for sensor-based activity detection very challenging.
- The human activity identification system's response This point addresses a different issue that needs to be taken into consideration when dealing with human data. Since the main goal of any recognition system is its use in daily life, it needs to be both quick and safe. Therefore, the computational cost should be as low as possible while maintaining the highest possible accuracy. Due to the nature of sensory input, which frequently contains other undesired signals and depends on proper utilization, this state is challenging to establish. Therefore, trustworthy recognition solutions must be able to analyze sensory input and understand which elements of information aid in recognition and which hinder it.

2. Materials and Methods

This chapter introduces the data collection, preprocessing and the architecture of developed neural network.

2.1 Hardware and software

The data collection station utilized in the development of this study is composed of the following components:

- a driving simulator
- Jins MEME Academic Pack glasses
- a computer used for data collection

This section provides a thorough description of the hardware and software.

2.1.1 JINS MEME Academic Pack Smart Glasses

This study is entirely performed on the basis of the signal acquired from JINS MEME Smart Glasses. They were designed by the Japanese brand called Jins, which specializes in making eyewear for everyday use. According to the developers, obtaining signal from the location of human's head gives the most detailed and accurate data about posture and allows for monitoring movement, whereas the eyes reveal information about health and state of mind. The model utilized in this study, shown in figure 2.1, looks like a regular pair of glasses, however is equipped with a set of EOG sensors mounted to the frame, one electrode on the bridge of the nose, which serves as reference signal and one on each of the nose pads, and a set of accelerometer and velocity sensors hidden in one of the arms. It combines the functionality of a fitness tracker with the benefits of other wearable smart devices assessing the spine.

The glasses are light, approximately 36 grams without the lenses, unisex and come in the Wellington design. They are made from a durable, waterproof material and the battery life is said to last around 16 hours, which is worth an average day of use. They can be charged with a Micro USB cable and the process takes approximately 2 hours.

All important components of the glasses are presented in the figure 2.2. To switch them on, one should press and hold the device's power button for 2 seconds, then release it. After that time, the LED located right above should start flashing blue, indicating that the glasses enter advertising mode. They wait for connection with the

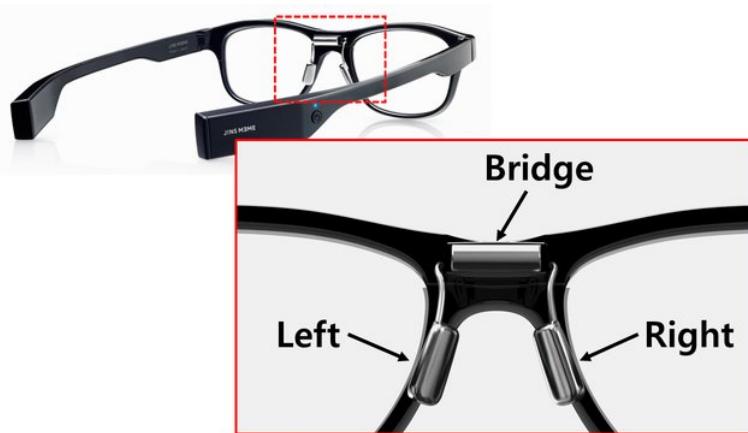


Fig. 2.1: JINS MEME Smart Glasses [7]

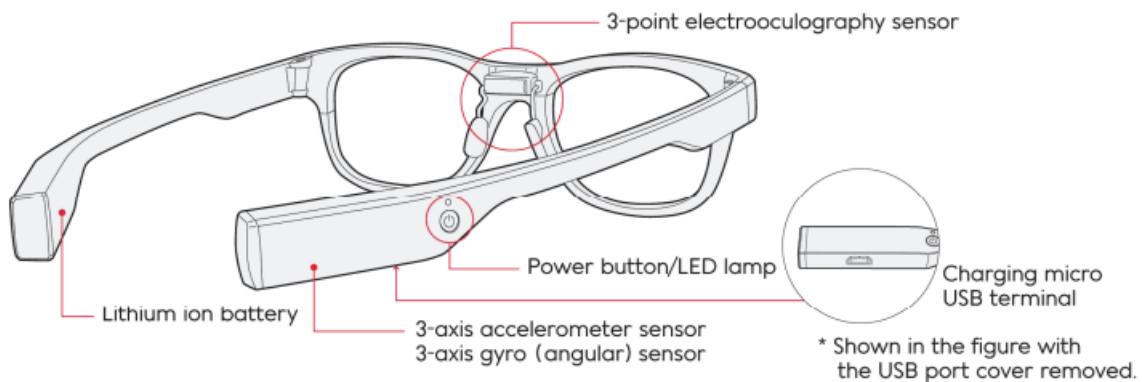


Fig. 2.2: JINS MEME Smart Glasses scheme [23]

PC for 10 seconds and if after that time the device is not connected, the LED switches to red and the device is turned off.

2.1.1.1 EOG signal

Electrooculography is a technique that utilizes the electrical features generated by the eye. The objective is to record eye movements by placing electrodes on the skin adjacent to the eyes. Specifically, the obtained signal represents the voltage difference between the cornea and the retina as shown in the figure 2.3. The detected potential changes as the eyes perform various types of movements due to the change in the electric field vector. The recorded time-variant signal is called electrooculogram or EOG.

Attached to the globe of an eye, there are three antagonistic pairs of muscles that control horizontal, vertical and torsional movement, as well as the position of an eye.

Namely, the muscles that are in control of:

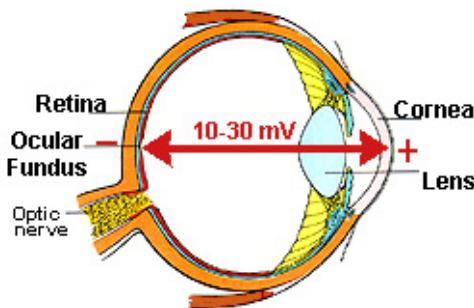


Fig. 2.3: Scheme of the eye. [21]

- horizontal movement (left versus right)
 - **the medial rectus muscle** responsible for adduction
 - **the lateral rectus muscle** responsible for abduction
- vertical movement (up versus down)
 - **the superior rectus muscle** elevates the eye with a slight intorsion
 - **the inferior rectus muscle** depresses the eye with a slight extorsion
- torsional movement (clockwise versus counter clockwise)
 - **the superior oblique muscle** responsible for intorsion with a slight depression
 - **the inferior oblique muscle** responsible for extorsion with a slight elevation (abducts)

Electrically active motor neurons innervate these muscles and generate movements that fit into one of the two categories:

- **Reflex eye movements**
function: stabilization of the eye position during head movements
- **Voluntary eye movements**
function: redirection of the line of sight

Different types of eye movements, each with a specific purpose and distinguishing characteristics, are controlled by the integration of information from several groups of neurons in the oculomotor system. The four basic types are: saccades, smooth pursuit movements, vestibular ocular reflex (VOR), vergence movements.

Saccades are rapid, ballistic movements that occur when the focal point moves from one area of the visual field to another. For instance, they can be as little as the motions performed when reading or as vast as those made while looking around a room. Although they can be induced purposefully, saccades happen reflexively whenever the eyes are open, even when they are fixed on a target.

Smooth pursuit movements also represent a type of eye movement, in which the eyes remain fixated on a moving object, however, much slower than saccades and fully voluntary.

Vestibulo-ocular movements balance out head movements by stabilizing the eyes' position in relation to the outside world. For instance, when turning head to the right, the eyes will adjust to the new position by turning to the left.

Vergence movements are, unlike in the other types, simultaneous movements of the pupils in the opposite direction while focusing on a target. When the object is closer, they move towards each other, thus converge, whereas if it is further away, they diverge.

2.1.1.2 JINS MEME Academic Pack

JINS MEME Academic Pack glasses are primarily designed for commercial use and come with designated applications that analyze the acquired signals and display the user's state of body and mind in real time. On the other hand, since the hardware is also utilized in academic research, the designers proposed the JINS MEME ACADEMIC PACK, which allows for custom acquisition of the raw sensor data. The results from all specified channels are then saved together with the exact time of occurrence in CSV format.

The Graphical User Interface is shown in the figure 2.4.

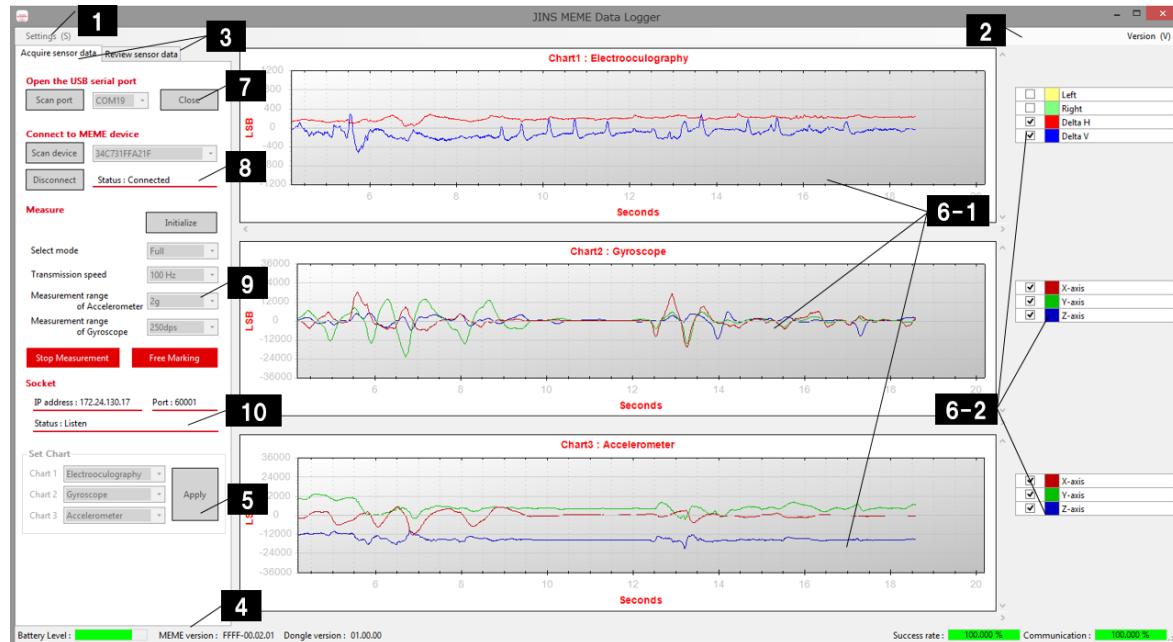


Fig. 2.4: Jins MEME Academic Pack software [24]

2.1.1.2.1 Connecting the glasses To begin data acquisition, the connection with Jins MEME glasses needs to be established. The process consists of the following steps:

- the dongle connection, label 7

First, the dongle that comes with the glasses needs to be inserted in an USB port on the PC. Afterwards, the destination port shall be detected by the software by clicking the 'Scan port' button. Following the correct selection of the port, using the 'Open' button, the connection is to be established.

- JINS MEME ES_R connection, label 8

In this step, the glasses are to be connected with the software. Similarly to the previous step, there are two buttons used for that purpose. When the glasses are in close distance to the PC and are turned on, the 'Scan device' button allows for their detection. Once the device is found, the connection is established using the 'Connect' button.

2.1.1.2.2 Data Acquisition Control This segment, label 9, is responsible for setting measurement parameters. There are three modes that can be chosen, that define the measuring range and the communication speed.

| Measurement Mode | | |
|---------------------------|-----------|---------------|
| Sensors | Full Mode | Standard Mode |
| Electrooculography sensor | 100 Hz | 200 Hz |
| Accelerometer sensor | 100 Hz | 100 Hz |
| Gyroscope sensor | 100 Hz | - |

Tab. 2.1: Sensor parameters of 2 possible measurement modes.

The 3rd possible measurement mode, not presented in table 2.1, is the Quaternion Mode, which outputs quaternions at a frequency of 100Hz.

Apart from the measurement mode, the following parameters as presented in table 2.2 are to be specified.

| Transmission Speed | Accelerometer measurement range | Gyroscope measurement range |
|--------------------|--|---|
| 50 or 100 [Hz] | $\pm 2, \pm 4,$ ± 8 or ± 16 [g] | $\pm 250, \pm 500,$ ± 1000 or ± 2000 |

Tab. 2.2: Possible values of sensor parameters.

Once all of the necessary parameters are correctly set, the environment and subject prepared, the data collection can be initiated by clicking the 'Start Measurement'

button. The graphs of the received signals are presented in the middle section of the UI, labeled as 6-1 in the figure 2.4. To stop the sequence, the same button as previously, which now says 'Stop Measurement' needs to be clicked. The acquired data is then saved in the location "\Documents\JINS\MEMEacademic\SensorData" in a CSV format.

2.1.2 Driving simulator

The experiments were performed in a simulated environment as presented in the figure 2.5.



Fig. 2.5: Driving simulator

The simulator is composed of the following components:

- a central unit equipped with:
 - an i7 processor,
 - XFX RADEON HD 5770 1 GB graphic card with NVIDIA processor and 3D VISION system,
 - 4 GB memory,

- Gigabyte's Ultra Durable 3 motherboard
- MOSFET transistors with low RDS(on), which in comparison to ordinary transistors have lower resistance and power consumption, leading to lower charging and discharging times, as well as reduced heat generation,
- Ferrite-core coils, characterized by lower energy losses compared to coils equipped with metallic cores, especially in the high-frequency range,
- Japanese Solid capacitors, are characterized by much longer failure-free operation, averaging 50,000 hours.
- standard computer peripherals,
- LED illumination,
- a spacial construction made of steel,
- a two way adjustable seat,
- a Logitech set: steering wheel, pedals and gearbox,
- 3 LED 27 monitors suitable for long operation,
- a sound system,
- dedicated software "Nauka jazdy".

The device allows to master the fundamental skills of driving such as:

- driving in a training area,
- starting, shifting gears,
- taking turns,
- parking,
- driving in traffic,
- following the traffic rules.

2.2 Data collection

The study was composed of two separate experiments conducted independently. Both were carried out using the JINS MEME Software with the default settings, therefore all signals were measured with the parameters specified by the Standard Mode as presented in the table 2.1, transmission speed 50 Hz and accelerometer measurement range ± 2 g. The signals were acquired one at a time, while the participants were given

voice commands as to when to start and finish to remove any additional distractions. Each sample was saved to a csv file with the name representing the exact date of capture. Following, they were all manually sorted into appropriate folders describing the performed action.

2.2.1 Study participants

Altogether, nine subjects (five males and four females) volunteered as study participants. The basic driving activities were carried out by six subjects (four males and two females) that were graduate students in their 20s. The distracting driving activities were performed by four subjects (one male and three females) in the ages ranging from 23 to 57. All participants had normal or corrected-to-normal vision. One subject participated in both experiments.

2.2.2 Scenarios

The tests were made up of such scenarios that represent the basic and distracting driving activities, which are to be distinguished. The primary activity scenarios were chosen as recommended by WORD and are being evaluated as the driving test is being administered. As stated in "Tabela nr 7 Dz.U. 2012 nr 0 poz. 995" [27] these activities include:

- Passing through unprioritised intersections (three- and four-legged),
- Passing through intersections marked with signs establishing priority of passage,
- Drive through intersections with traffic lights,
- Drive through intersections where traffic flows around an island
- Perform one of the following parking maneuvers:
 - perpendicular
 - angle
 - parallel

2.2.2.1 Basic driving activities

This experiment was carried out using the driving simulator. All participants started with a test ride to familiarize with the equipment. Once they felt comfortable, they were given a scenario and were asked to perform the action while wearing the JINS MEME Eyewear. The supervisor was in charge of controlling the software and giving voice commands, so that the participants could focus only on driving. The scenarios were divided into three categories, each performed in suitable environment. Altogether, there were 12 scenarios present in this segment.

The first set of activities were executed at a roundabout. It consisted of taking a first, second or third exit, what represented a left, straight and right turn. The actions are illustrated in the figure 2.6.

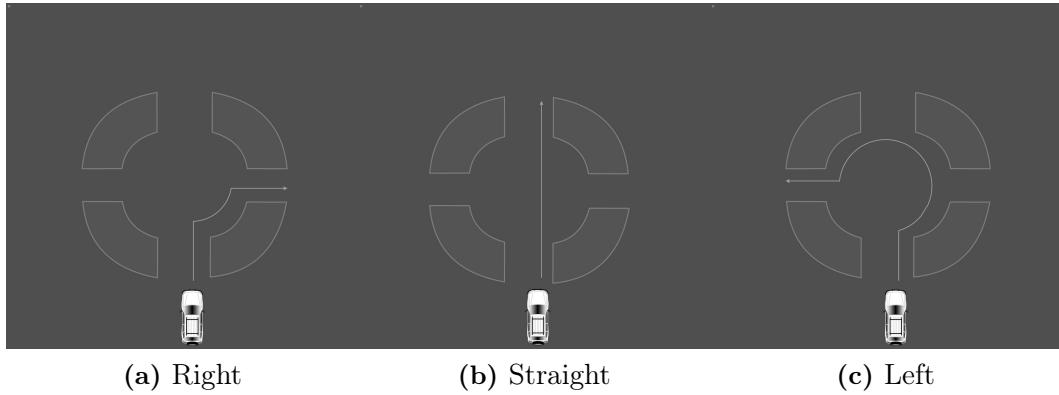


Fig. 2.6: Roundabout scenarios.

The second set of actions were executed at an intersection. The scenarios are similar to the roundabout and are illustrated in the figure 2.7.

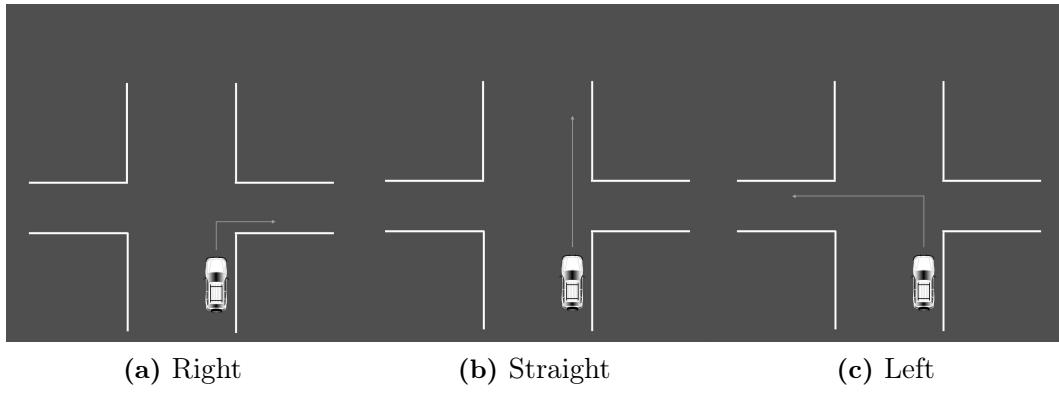
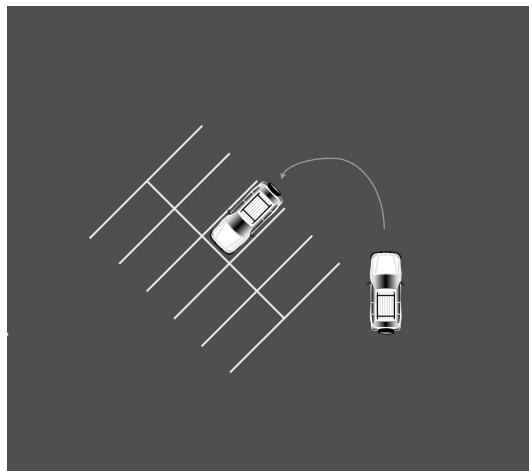
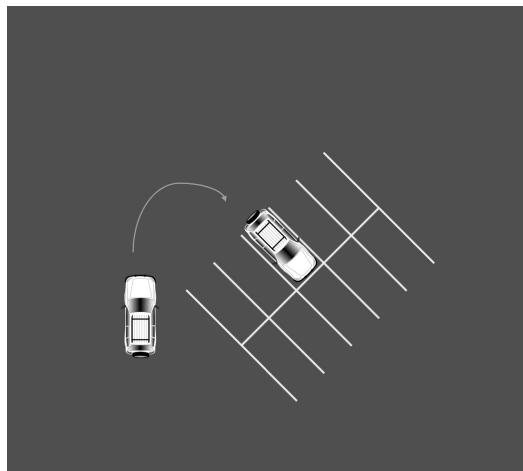


Fig. 2.7: Crossroad scenarios.

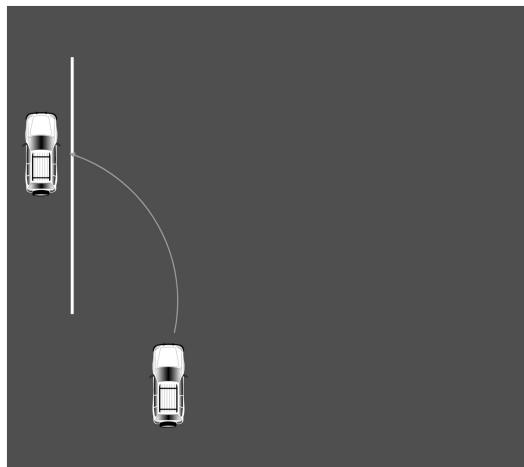
The last set of scenarios involves different kinds of parking techniques. Namely, angle, parallel and perpendicular. Each one of the presented actions was done twice, on either side of the road. All scenarios are illustrated in the figure 2.8.



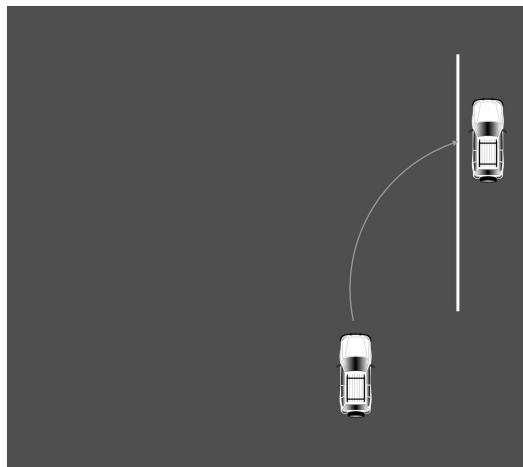
(a) park 1



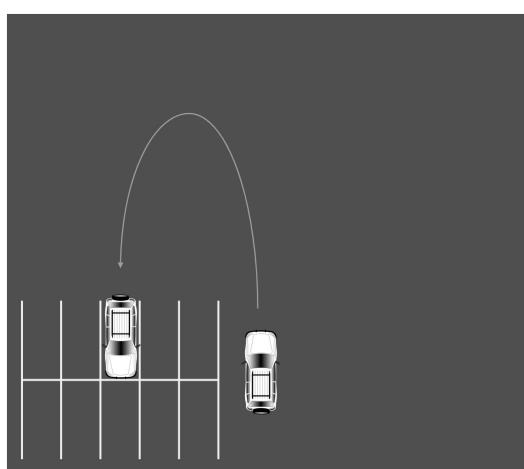
(b) park 2



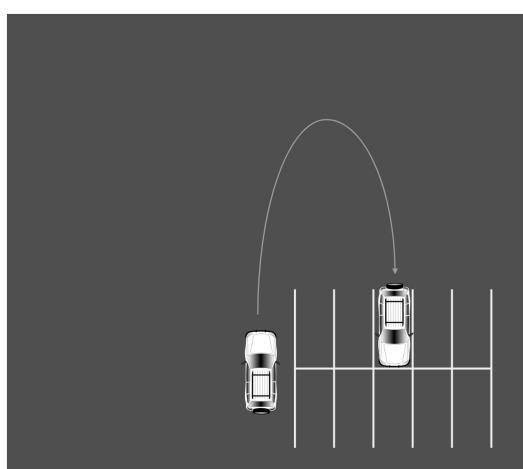
(c) park 3



(d) park 4



(e) park 5



(f) park 6

Fig. 2.8: Parking scenarios.

2.2.2.2 Distracting driving activities

The second experiment concerned the secondary or distracting driving activities. They represent all behaviors that take place behind a wheel, but are not connected to driving itself. They do, however, have an impact on the quality of the performance. Since these actions do not require being in a car, they were performed in a similar environment. In this part of the study, four scenarios called eating, drinking, turning and bending were introduced . The actions are explained in details in the table 2.3.

| Eating | Drinking | Turning | Bending |
|----------------------------|------------------------|--|---|
| Taking a bite of any food. | Taking a sip of water. | Turning back and reaching to a passenger's seat. | Bending and picking up a fallen object. |

Tab. 2.3: Secondary driving activities scenarios.

2.2.3 Obtained data

There are 1200 samples total from the first experiment, evenly distributed among classes that each represent one action. That implies that each folder contains 100 samples acquired from six participants. One subject contributed to half of the samples, with the remaining ones being distributed equally among the other subjects. The second experiment provided 700 samples evenly distributed among classes. One participant provided 100 samples in each class, whereas the other ones 25 each. Altogether, there are 2100 samples collected in this study. The samples are in a form of a csv file containing such information as the chosen mode with specified parameters used for signal acquisition, as well as the actual enumerated data with timestamps and header indicating what is represented by the specific value.

2.3 The algorithm

In this section the algorithm consisting of preprocessing and developing the neural network, is to be presented. The scheme is shown in the figure 2.9.

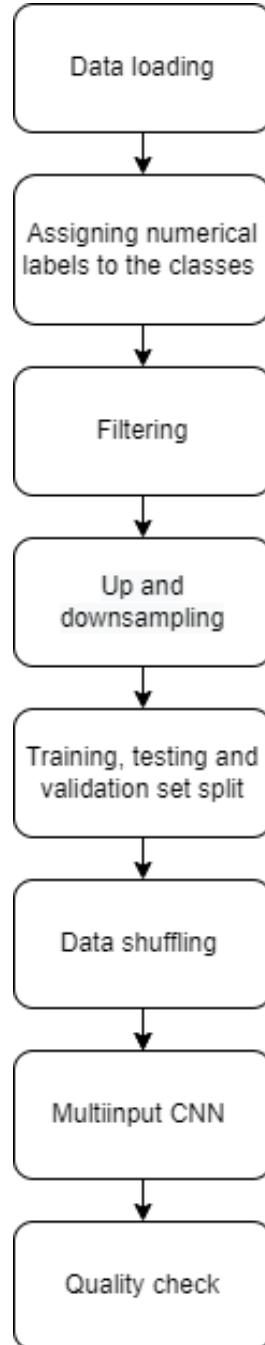


Fig. 2.9: Scheme of the methodology.

2.3.1 Data format, loading and labeling

All the data used in the experiment shall be stored in one place according to the specified path. The folders are loaded one at a time and given a numerical label in accordance with the order of reading. Each file is read into a dataframe and stored collectively.

2.3.1.1 Data format

The format of the data is as follows. First the data acquisition parameters are presented, followed by the header describing the content of each column that contains the sample number, the date in the format: dd.mm.yyyy:hh:mm:ss, then the 3 channel accelerometer components:

ACC_X - acceleration in the X-direction,

ACC_Y - acceleration in the Y-direction,

ACC_Z - acceleration in the Z-direction.

Followed by the EOG sensor components:

EOG_L - left eye raw EOG signal,

EOG_R - right eye raw EOG signal,

EOG_H - the difference between left and right eye potential $EOG_L - EOG_R$,

EOG_V - negative arithmetical mean of the left and right eye potential
 $-(EOG_L + EOG_R)/2$.

| | | | | | | | | | | | | |
|---------------------------------|------|-------|-------|-------|--------|--------|--------|--------|--------|--------|--------|--------|
| Data mode: Standard | | | | | | | | | | | | |
| Data quality: Standard | | | | | | | | | | | | |
| Acceleration sensor's range: 2g | | | | | | | | | | | | |
| sample number | date | ACC_X | ACC_Y | ACC_Z | EOG_L1 | EOG_R1 | EOG_L2 | EOG_R2 | EOG_H1 | EOG_H2 | EOG_V1 | EOG_V2 |
| | | | | | | | | | | | | |

Tab. 2.4: The format of acquired data.

2.3.1.2 Data loading and labeling

The data stored in relevant path is read by folders sequentially and stored in a list of dataframes containing one sample signal. The rows carrying the parameter specifications are omitted and the header is corrected, so that all signals are described correctly. The labels for the associated activities are presented in the table [3.1](#).

| Label | Activity |
|-------|-------------------------------|
| 0 | P_Crossroad_Left |
| 1 | P_Crossroad_Right |
| 2 | P_Crossroad_Straight |
| 3 | P_Parking_Diagonal_Left |
| 4 | P_Parking_Diagonal_Right |
| 5 | P_Parking_Parallel_Left |
| 6 | P_Parking_Parallel_Right |
| 7 | P_Parking_Perpendicular_Left |
| 8 | P_Parking_Perpendicular_Right |
| 9 | P_Roundabout_Left |
| 10 | P_Roundabout_Right |
| 11 | P_Roundabout_Straight |
| 12 | S_Bending |
| 13 | S_Drinking |
| 14 | S_Eating |
| 15 | S_Turning_Back |

Tab. 2.5: Driving activities labels.

2.3.2 Preprocessing

For a more accurate assessment of the raw data extracted from the sensor devices, in the preprocessing step noises and artifacts, such as the power-line noise or the baseline drift, were removed by applying appropriate filters and techniques. Since two different kinds of signals were acquired, the filtering step was performed separately for each of them.

2.3.2.1 Filtering the accelerometer signal

The raw accelerometer data from all three channels is passed through a median and low pass filter. The first type of filter is generally used to remove big spikes a.k.a outliers, that might be caused by electromagnetic interference. The median filter determines whether a sample is representative of its surroundings by considering each sample in the signal individually and examining its neighbors. The median of, in this case, 11 neighboring values is then used to replace the original value. However, high frequency noise can be found at extremely small amplitudes. Therefore with the use of a low pass filter, frequencies above the cut-off frequency, set to 20 Hz, are blocked. Although, this type of filter preserves low frequencies, it produces undesired distortions. Therefore both approaches are integrated in order to fully benefit from them, by first applying the median filter and then the passing the obtained signal through the Hamming window-based low pass filter. The drawback of this method is attenuated values at the signal's edges, however, since they might also have contained noise due to human control, they were trimmed off.

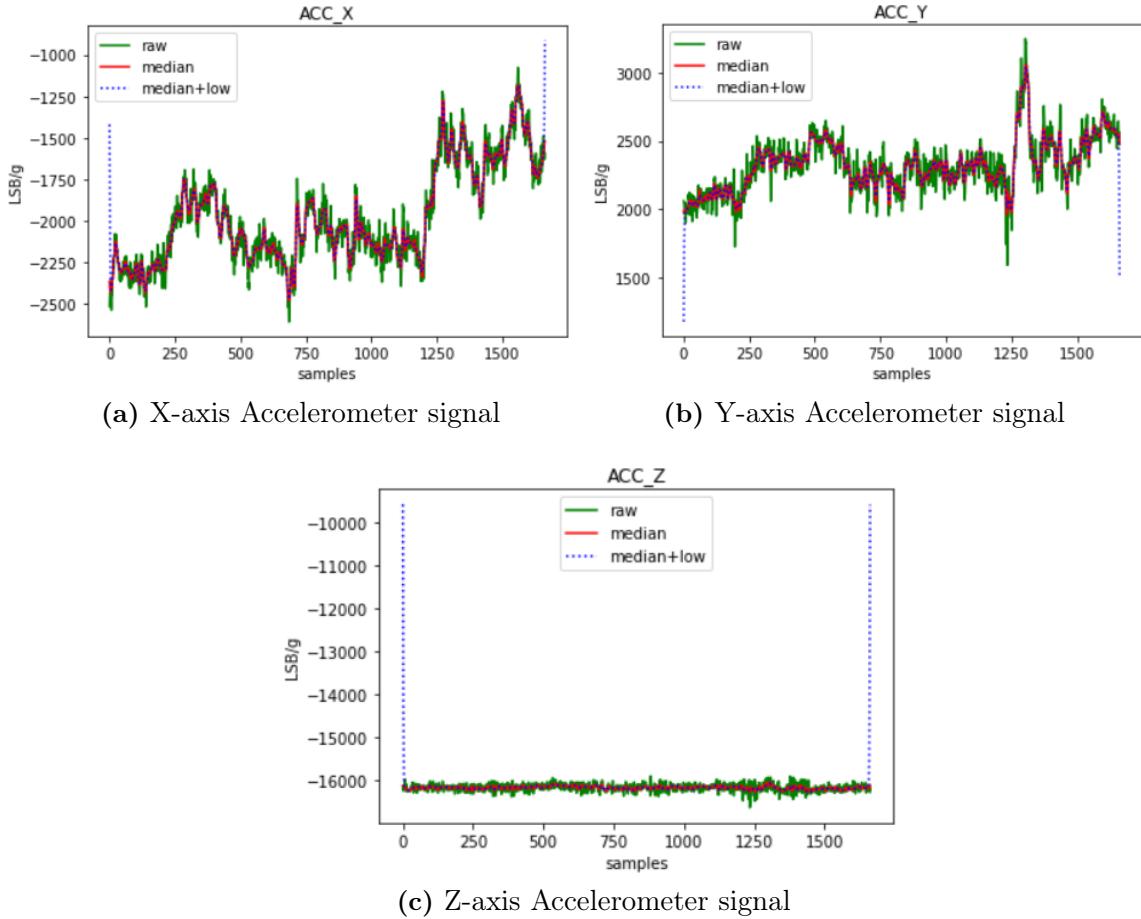


Fig. 2.10: 3-axis accelerometer signal filtering of going straight at a roundabout.

Afterwards, a data normalization technique was applied to the dataset. Z-score normalization was utilized to make the signal independent of any preceding shifts in the data, allowing the readings to significantly deviate from the rest value. The mean and standard deviation of each signal is computed and the values of the samples are replaced by the newly calculated values according to the formula:

$$x' = (x - \mu)/\sigma,$$

where μ is the mean of the signal, σ its standard deviation, x the current value of a sample and x' the new value.

so that the new mean of all the values is 0 and the standard deviation is 1.

2.3.2.2 Baseline drifts removal and filtering the EOG signals

The raw EOG signal presented in figure 2.11 contains different types of artifacts that first need to be filtered out.

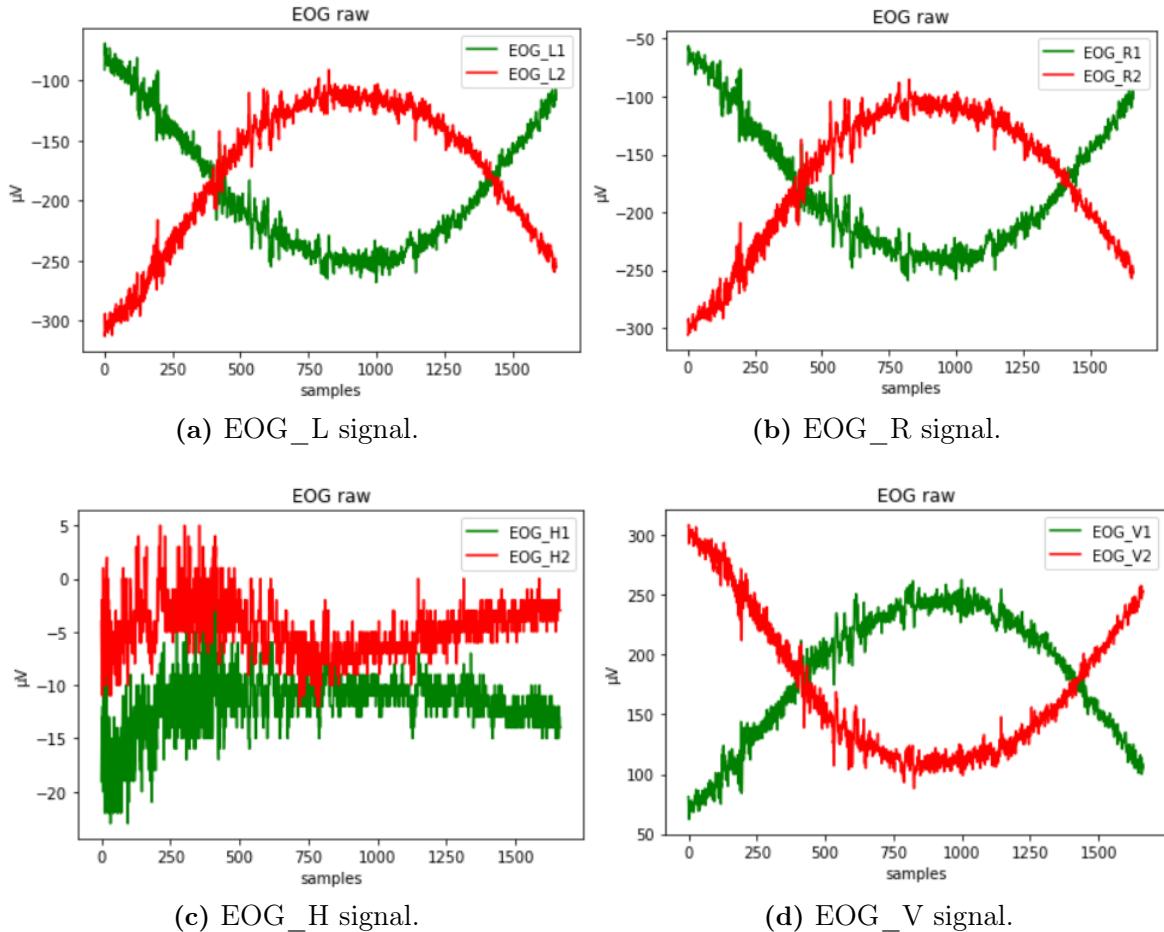


Fig. 2.11: Raw EOG signal of going straight at a roundabout.

To suppress power line noise and other The EOG signal undergoes the process of filtering by utilizing a second order lowpass Butterworth filter, which is then applied twice - forward and backward to the signal. Such filter has zero phase and twice the original filter's order. Afterwards, a baseline drift removal technique is utilized. Baseline drift is a slow, unrelated alteration that is superimposed on the EOG signal. There are numerous potential reasons for it, such as electrode polarization or interfering background signals [13]. Using a technique called detrending by differencing a time series, this objective was accomplished. The new value at the current time step is determined as the difference between the current and previous observations. Similarly to the case of accelerometer data, the EOG data was also normalized using the z-score normalization technique and the edges of the signal were neglected.

2.3.2.3 Resampling

For some tasks to be accomplished, more time was needed than for the others. Also depending on the precision of the driver, the acquired signals had different lengths. The shortest one was obtained for secondary activity Turning back and lasted 103 samples, whereas the longest one for primary activity taking a left turn at a roundabout was 3013 samples long. To be able to train the model, the data needed to be unified. Therefore, in order not to lose important data, all signals were resampled with the rate 3000.

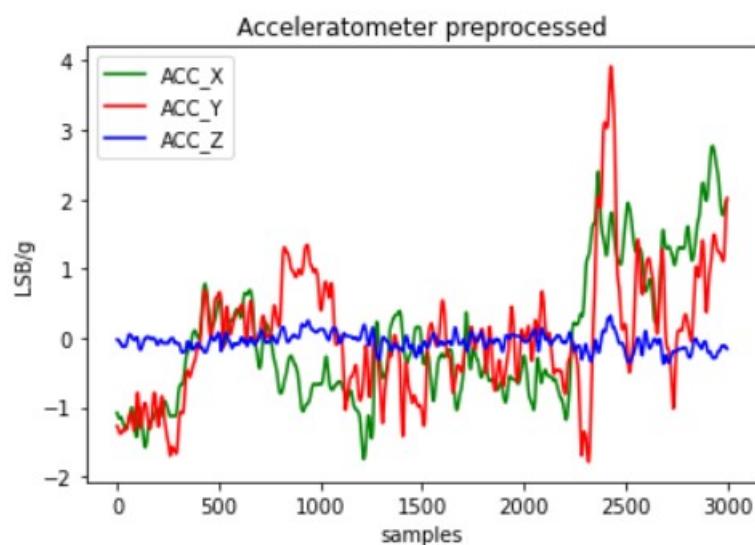


Fig. 2.12: Preprocessed accelerometer signal of going straight at a roundabout.

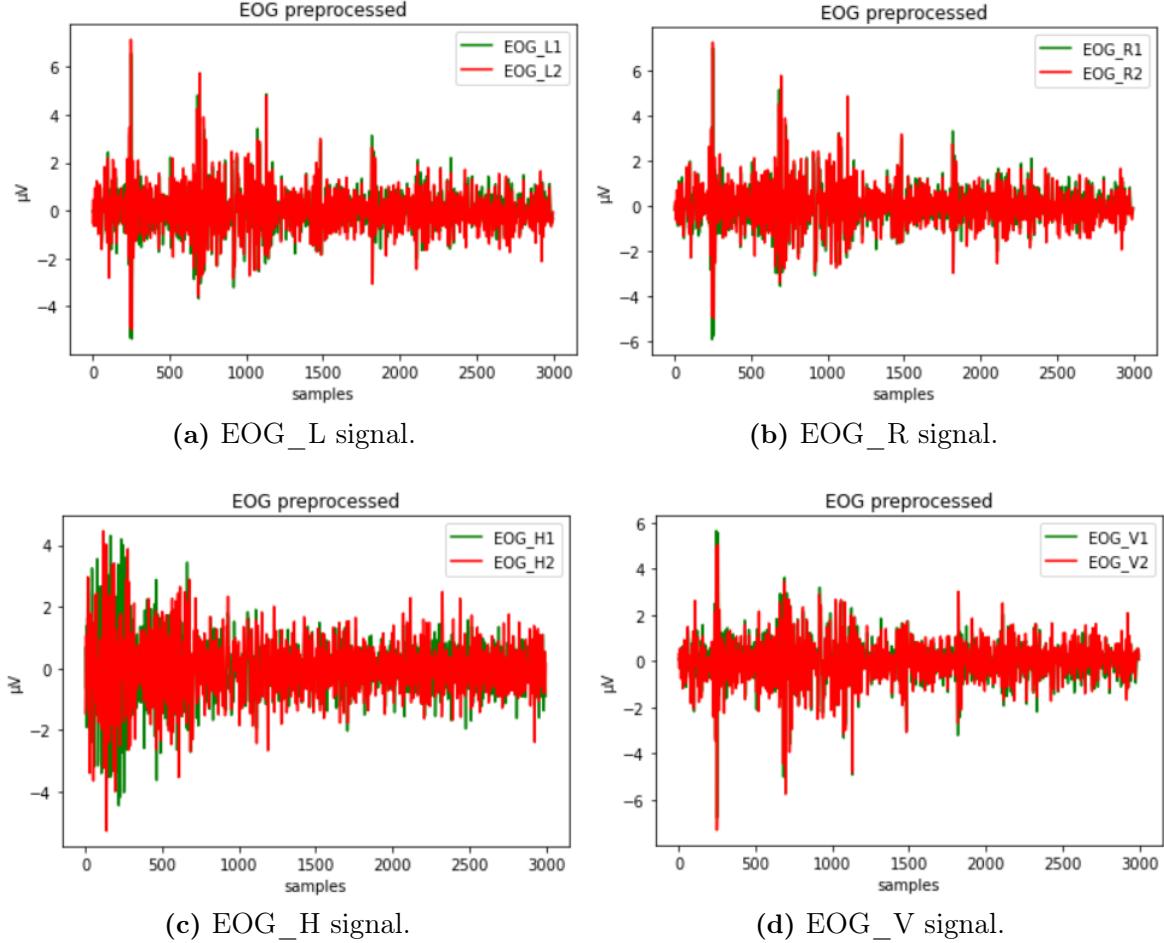


Fig. 2.13: Preprocessed EOG signal of going straight at a roundabout.

2.3.3 Neural Network architecture

The primary algorithm used in this study is 1-Dimensional [Convolutional Neural Network \(CNN\)](#). CNN has the ability to automatically identify significant features from highly dimensional data by several convolutional operators. Convolutions have the benefit of accounting for the data's spatial structure. By such, it is intended that extra information regarding the position in relation to other samples is taken into consideration. Both univariate and multivariate time series can be analyzed using 1D CNN. In the latter case, two or more parallel signals are combined into a single dataset, where each row corresponds to a time step and each column to a distinct time series. However, only samples along the time dimension in 1D CNN have an inherent ordering. Contrary to the most commonly used 2D CNN, the channels for the various variables do not.

The architecture of the network, used for training the model, is shown in the figure [2.14](#).

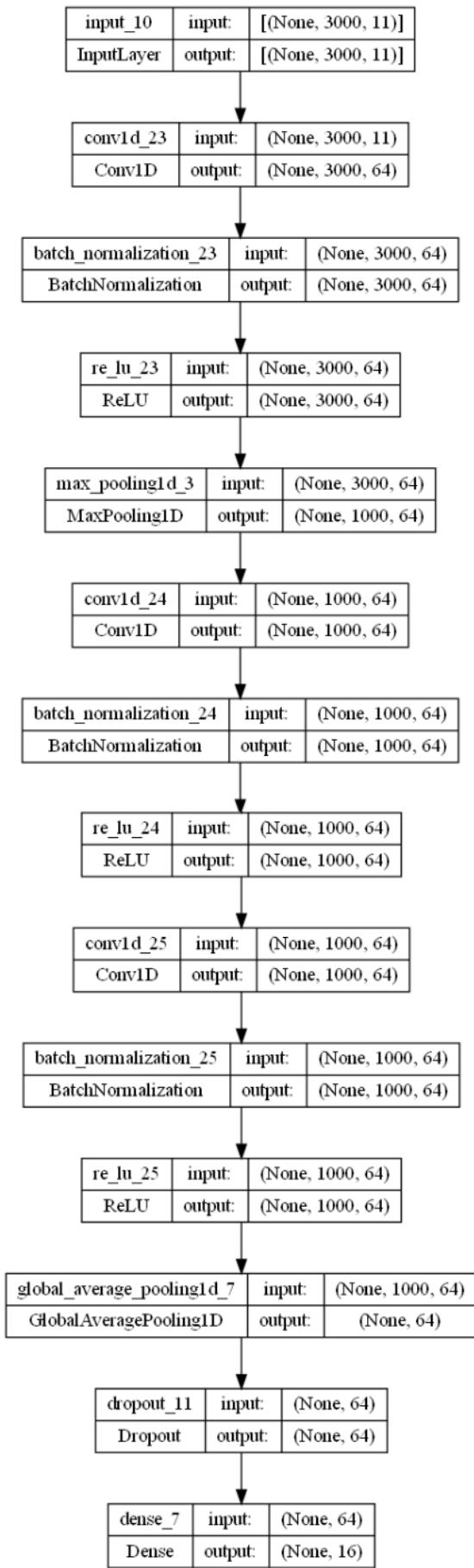


Fig. 2.14: Network architecture.

In terms of functionality, the model can be divided into two parts.

The first part serves as a feature extractor and is typical to this type of network. It utilizes convolution filtering processes to perform template matching. It employs such layers as a convolutional layer, a batch normalization layer, a [Rectified Linear Unit \(ReLU\)](#) correction layer and a pooling layer to produce so called "feature maps". This procedure can be carried out numerous times in order to allow the network to learn higher-level features.

The second part is the classification to one of the output classes. In order to create a new vector at the output, the input vector values are first reshaped using the global average pooling layer, along with an additional dropout layer to prevent the model from overfitting, and a dense layer, which assigns the final label representing the predicted class value by performing a matrix-vector multiplication.

Since it is a multivariate model, eleven features were employed in the input layer representing accelerometer and EOG signals each 3000 samples long.

The network consists of the following layers:

- **1st, 5th, 8th layer** The Convolutional layer

This layer allows for detecting features from a signal. The convolution operation is carried out between the vector and the filter by calculating its dot product, and returns a new vector with as many channels as the number of filters specified. Convolutions are typically performed by moving filters of a given width one element at a time; this is known as the Horizontal stride.

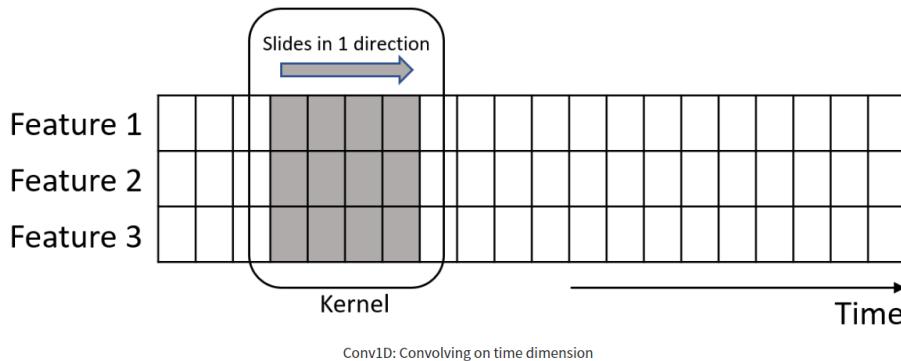


Fig. 2.15: Convolution operation. [26]

- **2nd, 6th, 9th layer** The Batch normalization layer

The batch normalization layer acts as a regularizer, normalizing the inputs throughout the backpropagation process, and is used to improve convergence of the model. It employs a transformation that keeps the output mean and standard deviation close to 0 and 1, respectively.

- **3rd, 7th, 10th layer** The ReLU layer

This activation layer helps capture interactions between certain features and non-linearities. Each value in the tensor is fed via an ReLU activation function described by the formula:

$$\text{ReLU}(z) = \max(0, z)$$

- **4th layer** The Max Pooling layer

The pooling operation entails sliding a two-dimensional filter across each channel of the feature map and summarizing the features that fall inside the filter's coverage zone. Therefore, instead of precisely positioned features produced by the convolution layer, further operations are conducted on summarised features. As a result, the model is more resistant to changes in the features' positions in the input signals.

- **11th layer** The 1D Global Average Pooling layer

The 1D Global Average Pooling layer is used to reduce the dimensionality by down sampling the entire feature map to a single value by computing the maximum of all the values for each of the input channels in a 2-dimensional tensor.

- **12th layer** The Dropout layer

When large neural networks are trained on limited datasets, they have a tendency to overfit the training data. To account for that problem, some nodes are arbitrarily ignored.

- **13th layer** The Dense layer

This is the last layer of the network, in which a matrix-vector multiplication is performed. The output of the previous layer is multiplied by the kernel weights yielding a matrix that allows for assigning to one of the output classes.

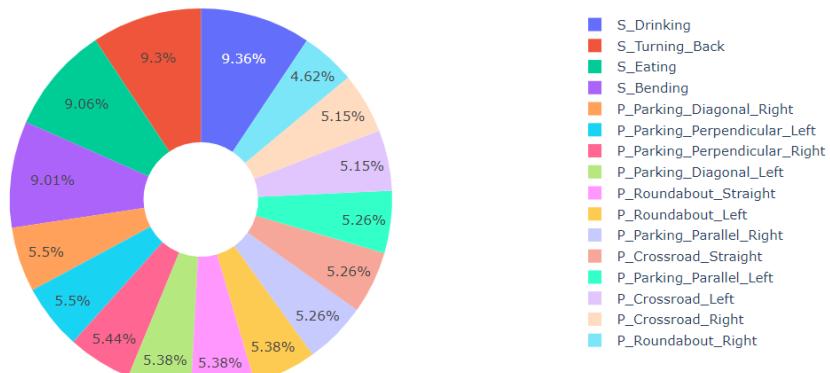
Parameters of specific layers:

- filters: 128,
- kernel size: 5
- pool size: 3
- dropout rate: 0.4
- batch size: 32
- maximum number of epochs: 0.004
- minimum learning rate: 0.0002

3. Results and Discussion

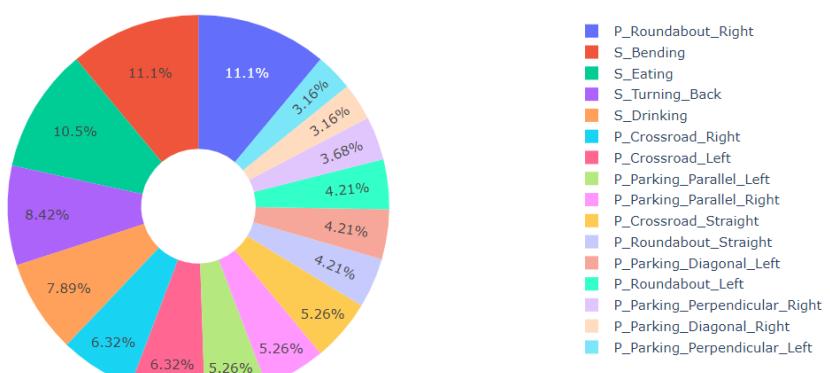
The network was trained and evaluated on the preprocessed accelerometer and EOG signals. The data was split into two subsets for training and testing with a 9:1 ratio. The training set was further divided into training and validation sets with a 8:2 ratio. Since the signals were sorted, to train the model on signals from all potential classes, the data had to be shuffled. The distribution of all classes is shown in the figure 3.1.

Distribution of classes in the training set



(a) Training set.

Distribution of classes in the training set



(b) Testing set.

Fig. 3.1: Distribution of activity samples in the training and testing sets.

3.1 Evaluation methods

The performance of the classifiers was assessed using the following common set of assessment indicators taking into account that the dataset was imbalanced.

Accuracy presents the percentage of correct predictions in respect to all predictions made.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \times 100\%$$

Precision is a metric that identifies the successful predictions out of all predictions made in favor of the event.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall presents the fraction of correctly classified predictions of a particular activity in respect to all predictions made in favor of the activity.

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-score is a harmonic mean of the precision and recall, which in comparison to accuracy should provide a more realistic model assessment in multiclass predictions with unbalanced classes.

$$F1 - score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

True positive (TP) - correctly classified trials

False positive (FP) - incorrectly classified trials

True negative (TN) - correctly classified nonevent trials

False negative (FN) - incorrectly classified nonevent trials

Categorical crossentropy loss measures the model performance by comparing the actual and predicted labels according to the formula:

$$CE = - \sum_{i=1}^N t_i \times \log(p_i)$$

t - true label

p - predicted label

N- the number of scalar values in the model output

All experiments in this study were performed using Python programming language utilizing Keras libraries for the development of neural network and sklearn for measuring its performance.

3.2 Accuracy and Loss while training

Figure 3.2 shows the curve of training and validation accuracy and loss of the model in respect to the number of epochs elapsed. The loss function is categorical crossentropy. When the epoch reached 130, the training accuracy was observed to be greater than 90%, providing a loss value of 0.2. The validation rate was at 80% with 0.6 loss. The model obtained the optimal parameters in 188 epochs.

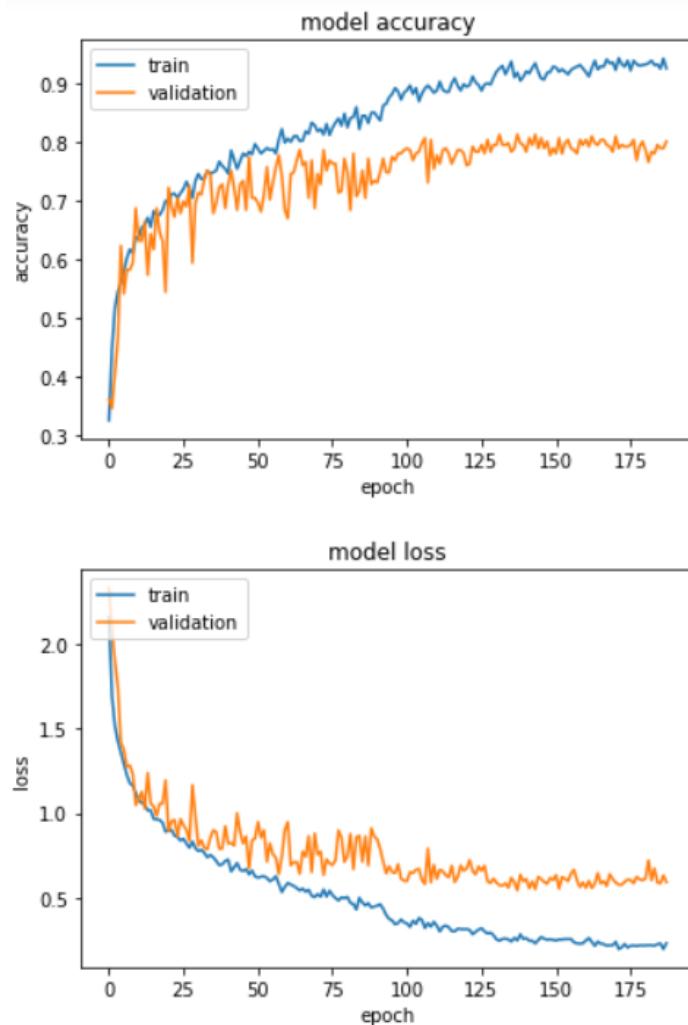
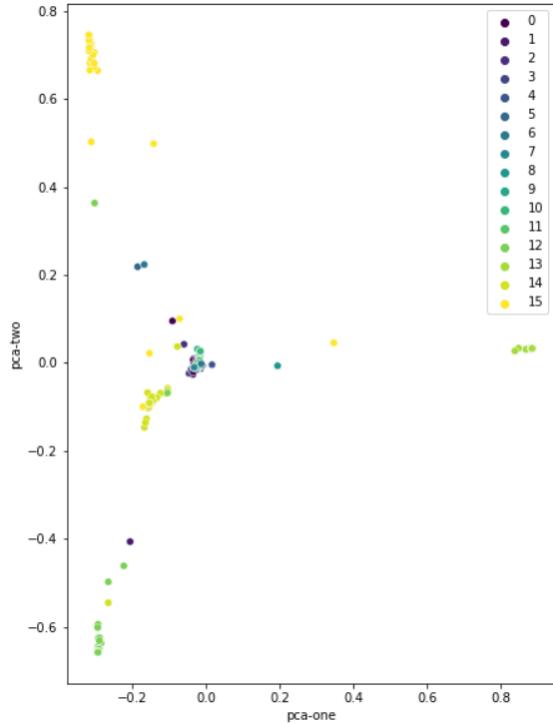
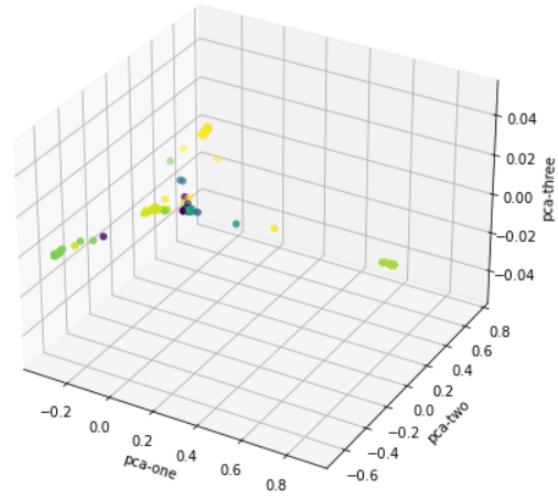


Fig. 3.2: Accuracy and loss graph for validation and training .

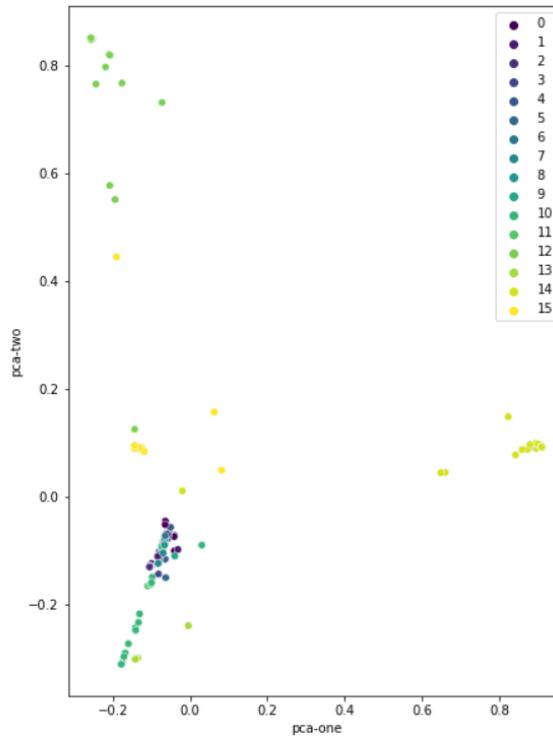
The figure 3.3 illustrates in both 2 and 3 dimensions how well the classes after 188 epochs were separated. For the purpose of visualization, a dimensionality reduction technique called [Principal Component Analysis](#) was utilized. It makes use of the correlation between various dimensions and aims to offer the fewest possible variables while retaining the most variance or details about the distribution of the original data.



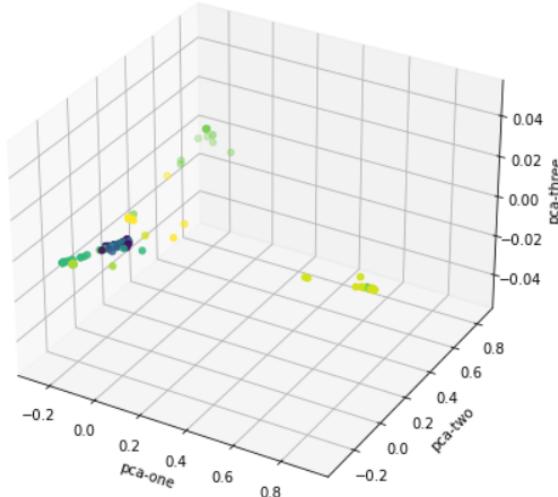
(a) Scatter plot 2d of training set labels.



(b) Scatter plot 3d of training set labels.



(c) Scatter plot 2d of testing set labels.



(d) |Scatter plot 3d of testing set labels.

Fig. 3.3: Scatter plot of the training and testing sets according to predicted labels.

It can be observed that the distinction between primary and secondary driving activities is very apparent. The latter ones are also separated in such a way that they mostly do not overlap. On the other hand, the primary activities cover areas very close to one another, so the greatest misclassifications are anticipated.

| Activity | Precision | Recall | F1-score |
|-------------------------------|-----------|--------|----------|
| P_Crossroad_Left | 1.00 | 0.67 | 0.80 |
| P_Crossroad_Right | 0.75 | 0.75 | 0.75 |
| P_Crossroad_Straight | 0.75 | 0.90 | 0.82 |
| P_Parking_Diagonal_Left | 0.67 | 0.75 | 0.71 |
| P_Parking_Diagonal_Right | 0.50 | 0.50 | 0.50 |
| P_Parking_Parallel_Left | 0.44 | 0.40 | 0.42 |
| P_Parking_Parallel_Right | 0.86 | 0.60 | 0.71 |
| P_Parking_Perpendicular_Left | 0.38 | 1.00 | 0.55 |
| P_Parking_Perpendicular_Right | 1.00 | 0.29 | 0.44 |
| P_Roundabout_Left | 1.00 | 0.88 | 0.93 |
| P_Roundabout_Right | 0.82 | 0.86 | 0.84 |
| P_Roundabout_Straight | 0.86 | 0.75 | 0.80 |
| S_Bending | 0.95 | 0.90 | 0.93 |
| S_Drinking | 1.00 | 1.00 | 1.00 |
| S_Eating | 0.90 | 0.95 | 0.93 |
| S_Turning_Back | 0.88 | 0.94 | 0.91 |

Tab. 3.1: Evaluation scores of driving activities classification.

The weighted average of precision, recall and f1 score of all activity predictions was equal to 0.83, 0.80, 0.80, respectively. The class that is classified the best is secondary activity drinking, whereas the worst primary parallel parking on the left and perpendicular parking on the right. All activity predictions had a weighted average precision, recall, and f1 score of 0.83, 0.80, and 0.80, respectively. The category yielding the best results is secondary activity drinking with all values equal to 1, and the category with the worst performance is primary parking to the left with f1 score 0.42. Overall, the classification of secondary driving activities performed better, all getting f1 score above 0.9.

The prediction outcome of a classification model is summarized in the confusion matrix. It presents details regarding the model's actual and expected classifications. The main diagonal contains the amount of correct predictions, while the values outside the diagonal depict inaccurate classifications. Figure 3.4 shows the confusion matrix for each of the 15 driver's activities as well the number of all expected and predicted samples belonging to a particular class, the percentage of correct and incorrect classifications and the accuracy of all predictions for test data.

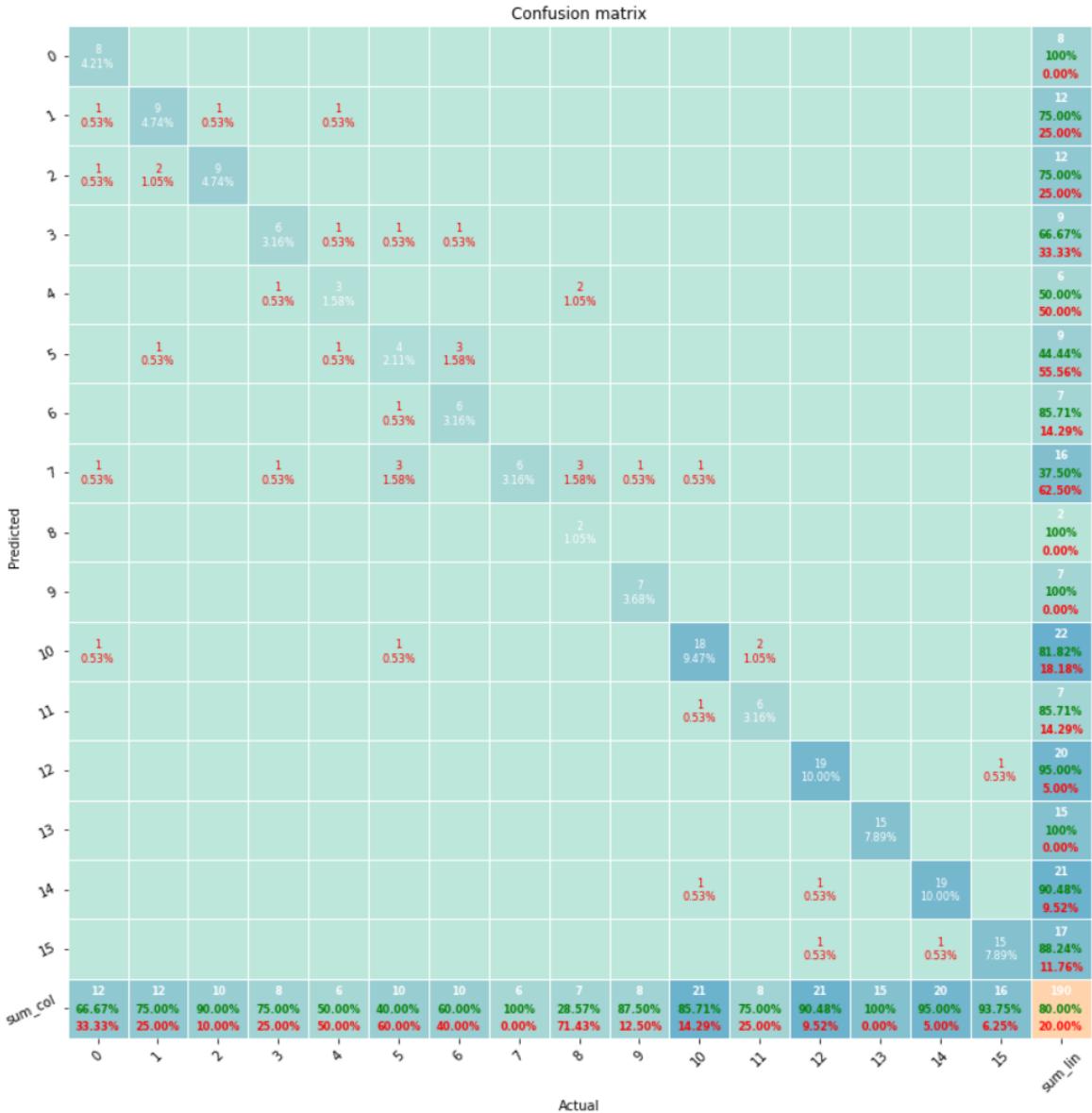


Fig. 3.4: Confusion matrix.

A similar category of activities was where the majority of misclassifications occurred. Parking activities reveal that, due to considerable similarities, the model had the most difficulty distinguishing between the same activity being performed on the

left and right side. An accuracy rate of 99.5 % was achieved for the binary classification between secondary and primary activities. The former was consistently placed in the appropriate group, while the latter was once incorrectly labeled as eating. An analysis of classifications for a specific set of activities, comprising crossroad, parking, roundabout, and secondary activities, produced accuracy values of 97.9%, 96.8%, 97.4% and 99.5%, respectively. These results alongside with the values of precision, recall and f1-score are presented in table 3.2. Despite the fact that these actions were supplied into the network as separate activities, the network was able to identify patterns that differentiate distinct action types.

| Activity | Accuracy [%] | Precision | Recall | F1-score |
|------------|--------------|-----------|--------|----------|
| Crossroad | 97.9 | 0.97 | 0.91 | 0.94 |
| Parking | 96.8 | 0.92 | 0.96 | 0.94 |
| Roundabout | 97.4 | 0.94 | 0.92 | 0.93 |
| Secondary | 99.5 | 0.99 | 1.00 | 0.99 |

Tab. 3.2: Evaluation scores of the type-based classification.

4. Summary

In conclusion, this study explores the possibility of distinguishing driving activities by tracking the muscle activation signal that is detected by EOG electrodes as well as by examining variations in head motion. Using simple light weight preprocessing methods and 1D CNN a series of primary and secondary driving activities were recognized with high accuracy. The model performed outstandingly in respect to the binary classification and yielded very good results when grouping the activities by type. The main goal of the study, recognizing all behaviors primary activities - crossroad turning left, right and going straight, angle parking left and right, parallel parking left and right, perpendicular parking left and right, roundabout turning left, right and going straight and secondary activities - bending, drinking, eating and turning back, showed relatively good performance, accuracy, precision and f1 score at the level of 80%, considering the high complexity of the imbalanced data.

Importantly, the findings demonstrate that the EOG and accelerometer signals are enough to determine the primary and secondary activities of drivers with an accuracy of 80%. Moreover, this detection is feasible using easy to wear commercial eyeglasses with an [IMU](#) affixed to the temples and EOG electrodes affixed to the bridge of the nose. All necessary components are readily available, portable, and small in design, and the program is created entirely from open source parts, demonstrating that it is possible to avoid utilizing pricey proprietary technologies in this situation. It is further proposed to use these experiments to develop algorithms that will contribute to making roads safer with more energy-efficient driving.

Although the preliminary results on activity detection are promising from an application perspective, they still require confirmation in larger investigations. After these successes, future research envisages examining a wider variety of subjects due to the user-dependent characteristics of the signal. Additionally, the model should also be trained and tested in a real-world setting to make it more robust to unstable environment. A broader spectrum of secondary activities might be investigated as well.

Bibliography

- [1] ANH SON LE, TATSUYA SUZUKI, H. A. Evaluating driver cognitive distraction by eye tracking: From simulator to driving. *Transportation Research Interdisciplinary Perspectives* 4 (2020).
- [2] ARNOLD, L.S., B. A. T. B. D. J. L. K. S. . H. W. Effectiveness of distracted driving countermeasures: A review of the literature (research brief). *AAA Foundation for Traffic Safety* (11 2019).
- [3] AT EXSILIO SOLUTIONS), R. J. I. Accuracy, precision, recall f1 score: Interpretation of performance measures, 10 2016.
- [4] BEANLAND, V., FITZHARRIS, M., YOUNG, K. L., AND LENNÉ, M. G. Driver inattention and driver distraction in serious casualty crashes: Data from the austalian national crash in-depth study. *Accident Analysis Prevention* 54 (2013), 99–107.
- [5] CHEN, K., ZHANG, D., YAO, L., GUO, B., YU, Z., AND LIU, Y. Deep learning for sensor-based human activity recognition: Overview, challenges and opportunities, 01 2020.
- [6] CHOI, S., KIM, J., KWAK, D., ANGKITITRAKUL, P., AND HANSEN, J. Analysis and classification of driver behavior using in-vehicle can-bus information. *Bienn. Workshop DSP In-Veh. Mob. Syst.* (01 2007).
- [7] DHULIAWALA, M., LEE, J., SHIMIZU, J., BULLING, A., KUNZE, K., STARNER, T., AND WOO, W. Smooth eye movement interaction using eog glasses. pp. 307–311.
- [8] DÍAZ, D., YEE, N., DAUM, C., STROULIA, E., AND LIU, L. Activity classification in independent living environment with jins meme eyewear. In *2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)* (2018), pp. 1–9.
- [9] DONIEC, R., PIASECZNA, N., LI, F., DURAJ, K., POUR, H. H., GRZEGORZEK, M., MOCNY-PACHOŃSKA, K., AND TKACZ, E. Classification of roads and types of public roads using eog smart glasses and an algorithm based on machine learning while driving a car. *Electronics* 11, 18 (2022).

- [10] EL-RIFAIE, A., AL BARAKEH, Z., NEJI, B., AND GHANDOUR, R. Driver distraction and stress detection systems: A review.
- [11] EUROPEAN ROAD SAFETY OBSERVATORY. BRUSSELS, EUROPEAN COMMISSION, D. G. F. T. European commission (2022) road safety thematic report - driver distraction, month =, 2022.
- [12] FREDRIKSSON, R., LENNÉ, M. G., VAN MONTFORT, S., AND GROVER, C. European ncap program developments to address driver distraction, drowsiness and sudden sickness. *Frontiers in Neuroergonomics* 2 (2021).
- [13] GU, J. J., MENG, M. Q.-H., COOK, A., AND FAULKNER, M. G. A study of natural eye movement detection and ocular implant movement control using processed eog signals. *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164)* 2 (2001), 1555–1560 vol.2.
- [14] HOZHABR POUR, H., LI, F., WEGMETH, L., TRENSE, C., DONIEC, R., GRZEGORZEK, M., AND WISMULLER, R. A machine learning framework for automated accident detection based on multimodal sensors in cars. *Sensors* 22, 10 (2022).
- [15] HUSSAIN, A., ALI, S., ABDULLAH, AND KIM, H.-C. Activity detection for the wellbeing of dogs using wearable sensors based on deep learning. *IEEE Access* 10 (2022), 53153–53163.
- [16] ISHIMARU, S., HOSHIKA, K., KUNZE, K., KISE, K., AND DENGEL, A. Towards reading trackers in the wild: Detecting reading activities by eog glasses and deep neural networks. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers* (2017), Association for Computing Machinery, pp. 704–711.
- [17] JIN, L., NIU, Q., HOU, H., XIAN, H., WANG, Y., AND SHI, D. Driver cognitive distraction detection using driving performance measures. *Discrete Dynamics in Nature and Society* 2012 (11 2012).
- [18] KINNEAR, N., AND STEVENS, A. The battle for attention: Driver distraction - a review of recent research and knowledge.
- [19] KLAUER, S. G., DINGUS, T. A., NEALE, V. L., SUDWEEKS, J., AND RAMSEY, D. J. The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data.
- [20] KOPROWSKI, R., AND WRÓBEL, Z. *Praktyka przetwarzania obrazów w programie MatLab*. Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2004.
- [21] LAB, T. M. P. V. Biological signals acquisition.

- [22] LIN, C.-T., KING, J.-T., BHARADWAJ, P., CHEN, C.-H., GUPTA, A., DING, W., AND PRASAD, M. Eog-based eye movement classification and application on hci baseball game. *IEEE Access* 7 (2019), 96166–96176.
- [23] LTD, J. C. *JINS MEME ES_R Instruction Manual*.
- [24] LTD, J. C. User interface.
- [25] LV, Z., PEI WU, X., LI, M., AND ZHANG, D. A novel eye movement detection algorithm for eog driven human computer interface. *Pattern Recognition Letters* 31, 9 (2010), 1041–1047.
- [26] MACNICA. Convolutional neural network (cnn) for time series classification, 10 2021.
- [27] MIN. TRANSPORTU, B. I. G. M. Rozporządzenie ministra transportu, budownictwa i gospodarki morskiej z dnia 13 lipca 2012 r. w sprawie szkolenia osób ubiegających się o uprawnienia do kierowania pojazdami, instruktorów i wykładowców, dz.u. 2012 poz. 1019, 2012.
<https://isap.sejm.gov.pl/isap.nsf/download.xsp/WDU20120000995/0/D20120995.pdf>.
- [28] MOHD NSIR, H., AMINUDDIN, M., BRAHIN, N., AND MISPLAN, M. S. Hybrid mean fuzzy approach for attention detection. *International Journal of Online and Biomedical Engineering (iJOE)* 17 (06 2021), 58.
- [29] NHTSA. U.S. Department of Transportation distracted driving 2020. <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813309>, May 2022.
- [30] NISAR, M. A., SHIRAHAMA, K., LI, F., HUANG, X., AND GRZEGORZEK, M. Rank pooling approach for wearable sensor-based adls recognition. *Sensors* 20, 12 (2020).
- [31] PURVES D, AUGUSTINE GJ, F. D. E. A. *Neuroscience. Types of Eye Movements and Their Functions*, 2 ed. Sunderland (MA): Sinauer Associates, 2001.
- [32] RIZZO, M., AND HURTIG, R. Looking but not seeing. *Neurology* 37, 10 (1987), 1642–1642.
- [33] VÄGVERKET. Safe traffic : Vision zero on the move. <https://rosap.ntl.bts.gov/view/dot/18273>, March 2006.
- [34] XING, Y., LV, C., WANG, H., CAO, D., VELENIS, E., AND WANG, F.-Y. Driver activity recognition for intelligent vehicles: A deep learning approach. *IEEE Transactions on Vehicular Technology* 68, 6 (2019), 5379–5390.
- [35] YAN, F., LIU, M., DING, C., WANG, Y., AND YAN, L. Driving style recognition based on electroencephalography data from a simulated driving experiment. *Frontiers in Psychology* 10 (2019).