Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

Iselin Johanna Nordstrøm-Hauge
Molly Vassbotn

EEG-Based Alcohol Detection System with AI Techniques

Towards the Design of BCI Systems for Driver Monitoring

Master's thesis in Cybernetics and Robotics
Supervisor: Marta Molinas
Co-supervisor: Andres Soler
June 2023

NTNU
Norwegian University of
Science and Technology

Iselin Johanna Nordstrøm-Hauge
Molly Vassbotn

EEG-Based Alcohol Detection System with AI Techniques

Towards the Design of BCI Systems for Driver Monitoring

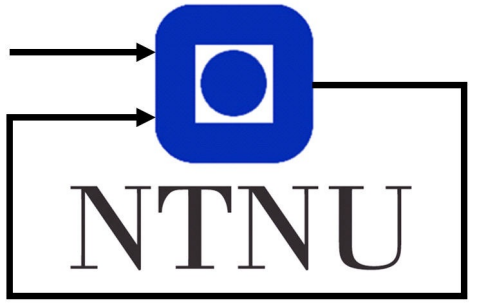
Master's thesis in Cybernetics and Robotics
Supervisor: Marta Molinas
Co-supervisor: Andres Soler
June 2023

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

NTNU
Norwegian University of
Science and Technology

EEG-Based Alcohol Detection System with AI Techniques

Towards the Design of BCI Systems for Driver Monitoring



Authors:
Iselin Johanna Nordstrøm-Hauge
Molly Vassbotn

Supervisor:
Marta Molinas

Co-supervisor:
Andres Soler

Master thesis
Department of Engineering Cybernetics
Norwegian University of Science and Technology

1st June 2023

Preface

This master's thesis was written in the spring of 2023 with the Department of Engineering Cybernetics at the Norwegian University of Science and Technology (NTNU). The project was proposed and supervised by Professor Marta Molinas and Andres Soler.

The thesis is a continuation of the work presented in the semester projects "Design of protocol and collection of data for an EEG-based alcohol detector" [1, 2]. Parts of Chapter 2 and Sections 3.1, 3.2 and 3.3 are updated and extended versions of chapters from these project theses, in which Appendix A and B were also presented. These appendices are included in this thesis with minor modifications. The data set presented in this master's thesis was collected solely during this spring, and no data collected during the semester project were used due to differences in hardware.

The results of the alcohol detector project will be submitted to the IV Latin American Workshop on Computational Neuroscience (IV LAWCN) as M. Vassbotn, I. J. Nordstrøm-Hauge, A. Soler, and M. Molinas, "EEG-Based Alcohol Detection System for Driver Monitoring". This paper can be found in Appendix G.

Before starting this project, neither of us had any knowledge about EEG signals, BCIs or alcohol metabolism. Hopefully, this master's thesis reflects the knowledge and experiences we have gained through working on this project.

Trondheim, 1st June 2023
Molly Vassbotn & Iselin J. Nordstrøm-Hauge

Acknowledgements

First, we would like to thank our supervisor Professor Marta Molinas for proposing this project and making it possible for us to partake in the entire process from designing an experiment, to conducting it and analysing the results. We would also like to thank our co-supervisor Andres Soler for allowing us to come knocking at his door whenever we needed help. We thank them both for always being available to discuss major and minor challenges we have faced during this project, and for giving us input. This master's thesis would not have become what it is today without their passion for the project.

Molly would like to acknowledge her family and friends for support and words of encouragements during this project. She would especially like to thank her boyfriend for keeping her spirits high. Lastly, she would also like to express gratitude towards Iselin for keeping her motivated and for an excellent partnership.

Iselin would like to thank her friends and family for their continuous support and motivation. She would like to give extra acknowledgement to her dad for always providing helpful feedback. Last, she would like to thank co-author Molly for a great partnership and for staying positive, even in the most challenging times of the project.

Finally, we would like to thank the participants for making it possible to complete this thesis. We thank them for spending hours helping us with recording the data for the project, and for staying patient when problems occurred. We are extremely grateful for them agreeing to help us with the data collection.

Abstract

This thesis examines the possibility of detecting the presence of alcohol in electroencephalography (EEG) signals. Driving under the influence of alcohol is a global problem, and the purpose of this work is to contribute with the first steps towards designing an EEG-based alcohol detector to be utilised in an onboard brain-computer interface (BCI) system. This BCI system could help prevent the consequences of driving under the influence of alcohol.

An experiment was designed for collecting EEG signals from 20 participants, both while under the influence of alcohol and not. The data collected during the experiment was recorded using 16 EEG channels, and the participants partook in two recording sessions. In one session, the participants were presented with an alcoholic drink consisting of a mix of 0.45 g/kg ethanol and orange juice. In the other, they were presented with a non-alcoholic drink. During each 66-minute recording session, the participants were instructed to perform the Flanker task and their blood alcohol concentration (BAC) values were measured at specific points in time. This was in addition to 5-minute EEG recordings where the participants were instructed to sit still, while relaxing with their eyes open.

The analysis of the recorded BAC values showed that the BAC values after alcohol consumption increased over time. Most of the participants did not reach their peak BAC, which is the maximum point of BAC before it starts decreasing. To be able to capture the peak in the future, the participants should be instructed to not eat beforehand, or they should be given an undiluted alcoholic drink. A statistical analysis of the Flanker data shows that there is a significant decrease in the response time of the participants while under the influence of alcohol. This is likely caused by the disinhibition typically observed in people under the influence of alcohol. These results show that the participants were affected by the alcohol, and, therefore, their EEG signals are most likely affected as well.

To investigate the effect of alcohol on EEG signals, three classification models were developed to classify the collected signals. Two individual models were implemented. One is a random individual model which randomly splits the collected data into training and test sets. The other is an individual model across sessions which splits full 5-minute recordings into either the training or test set. The average accuracies of these models were 100% and 90.7%, respectively. Although the random individual model has a better accuracy, the individual model across sessions is seen as the best model since the implementation is more realistic. This makes it more applicable as a real-world alcohol detector.

The last classifier is a general model. This model trains on 19 participants and tests on the last, unseen participant. The average accuracy of this model was 62.9%. As indicated by low precision and recall values, the model has difficulties with classifying alcohol samples correctly.

Sammendrag

Denne masteroppgaven undersøker muligheten for å detektere tilstedeværelsen av alkohol i elektroencefalografi (EEG)-signaler. I dag er alkoholpåvirket kjøring et verdensomspennende problem. Målet med dette prosjektet er å danne grunnlaget for en EEG-basert alkoholdetektor som videre kan brukes i et hjerne-datamaskin-grensesnitt system. Dette systemet kan potensielt hjelpe med å bekjempe problemet med alkoholpåvirket kjøring.

Et eksperiment ble designet hvor EEG-data fra 20 deltagere ble samlet inn. Deltagerne var både påvirket og upåvirket av alkohol under innsamlingen. Dataene ble samlet inn ved å bruke 16 EEG-kanaler, og hver deltager deltok på 2 innsamlingstimer. I den ene innsamlingstimen ble deltageren servert en alkoholholdig drikk med 0.45 g/kg etanol og appelsinjuice. I den andre timen fikk deltageren en tilsvarende drikke uten alkohol. Hver innsamlingstime varte i omtrent 66 minutter. I løpet av timen ble det samlet inn 5-minutters EEG-opptak, i tillegg til at deltageren ble bedt om å gjennomføre Flanker-testen. Promillen ble også målt. Under EEG-opptakene fikk deltageren beskjed om å sitte stille med åpne øyne, å slappe av, og å bevege seg så lite som mulig.

De innsamlede promilledataene ble analysert. Analysen indikerte at promilletoppen ikke ble nådd for de fleste av deltagerne. For at denne skal inkluderes burde deltagerne enten bli servert en konsentrert alkoholholdig drikke, eller så må de unngå å spise før innsamlingstimene. En statistisk analyse av Flanker-dataene ble gjennomført. Denne analysen viser at det er en signifikant nedgang i responstiden for alle deltagere etter alkoholinntak. Denne nedgangen er sannsynligvis forårsaket av impulsiviteten en person kan oppleve etter å ha drukket alkohol. Disse resultatene viser at deltagerne ble påvirket av den konsumerte alkoholen, og derfor er de innsamlede EEG-signalene deres mest sannsynlig også påvirket.

Tre klassifikatorer ble utviklet for å undersøke alkoholens effekt på EEG-signaler. To av disse var individuelle modeller. Den ene individuelle modellen delte EEG-dataene tilfeldig inn i et treningssett og testsett. Den andre modellen, en individuell modell på tvers av datainnsamlinger, sørget for å dele data fra samme 5-minutters opptak inn i enten treningssettet eller testsettet. De gjennomsnittlige nøyaktighetene for disse modellene var henholdsvis 100% og 90.7%. Selv om den tilfeldige individuelle modellen virker bedre basert på nøyaktighet alene, er den individuelle modellen på tvers av datainnsamlinger likevel sett på som den beste av disse. Det er fordi implementeringen av denne modellen er mer realistisk, og dette gjør modellen bedret egnet til bruk i det virkelige liv.

Den siste klassifikatoren er en generell modell. Denne modellen trente på data fra 19 deltagere, og ble testet på dataene fra den siste, usette deltageren. Den gjennomsnittlige nøyaktigheten til denne modellen var på 62.9%. I tillegg hadde modellen en lav presisjon og en lav gjenkalling. Dette tyder på at modellen har problemer med å klassifisere signaler som er påvirket av alkohol riktig.

Table of Contents

List of Tables

List of Figures

Abbreviations

Abbreviation	Description
BAC	Blood Alcohol Concentration
BCI	Brain Computer Interface
BMI	Body Mass Index
CNN	Convolutional Neural Network
DeepLIFT	Deep Learning Important Features
EEG	Electroencephalography
ERPs	Event-Related Potentials
FN	False Negative
FP	False Positive
HD	High Dose
LD	Low Dose
LSL	Lab Streaming Layer
RT	Response Time
SD	Standard Deviation
TN	True Negative
TP	True Positive

1

Introduction

Today, alcohol drinking frequently accompanies socialising as a routine activity in various groups of society. 84.0% of individuals aged 18 and above in the United States have drunk alcohol at some point in their life [3]. Similarly, 81.7% of Norwegians in the age group 16 to 79 have drunk alcohol in 2021 [4]. Around the world, 40.7% of all 20 to 24-year-olds consume alcohol, making it the most commonly used substance among the young [5]. When alcohol is consumed, it reaches the brain within minutes. It can affect the brain in several ways, such as slurred speech, blurred vision and slowed reaction time [6].

1.1 Motivation

Activity in the brain generates electrical signals. Electroencephalography (EEG) is a technique used for capturing these signals by placing electrodes on the scalp. Today, EEG is a commonly used technique for studying the brain. Many studies [7, 8, 9] have been investigating if it is possible to detect alcoholism using EEG data. These studies have been successful, resulting in high classification accuracy. However, there has been less focus on using EEG data for detecting alcohol in a healthy body.

Driving under the influence of alcohol is a worldwide problem. It is estimated to cause the death of at least 273 000 road users every year, although the actual number is believed to be higher [10]. The legal blood alcohol concentration (BAC) for driving varies for different countries, but, for most, the BAC limit is within the range of 0.2‰ to 0.8‰ [11].

The legal BAC limit in Norway is 0.2‰ [12]. Here, 25% of drivers who died in traffic between 2001 and 2010 had a BAC above the legal limit. 1 out of 10 drivers injured while driving is also above this limit [13]. In Canada and the United States, the legal BAC limit is 0.8‰ [14]. In the United States, 31% of all traffic crash fatalities involve people driving under the influence of alcohol [15]. Similarly, the involvement of alcohol was a contributing factor in approximately 20% of all fatal collisions in Canada in 2018 [16].

To decrease the number of injuries and deaths, it is important to stop people from driving under the influence of alcohol before the accidents happen. Today, using breathalysers is the most common way of detecting if a person is under the influence of alcohol. The breathalyser estimates the BAC by using a single breath sample. It is the police's preferred tool to use when they suspect a person is driving while intoxicated.

Although using breathalysers is a quick and inexpensive way of measuring the BAC, it has some disadvantages. Using breath samples is an indirect way of measuring the amount of alcohol in the blood, and incorrect measurements can occur. Residual alcohol in the mouth can result in a higher measured BAC than the actual value. If a person has been drinking juice few minutes before a breathalyser test, the breathalyser can show a BAC value higher than zero, resulting in it falsely detecting the presence of alcohol. Factors such as temperature and humidity can also affect the accuracy of the breathalyser [17].

Several studies [18, 19, 20] have been proposing an in-vehicle alcohol detector that can inform drivers if their level of alcohol intoxication is above the legal limit or if their level of drowsiness is considered too high for driving. These factors can be determined by using several different methods. [18] is using a series of low-cost alcohol MQ3 sensors together with machine learning, while [19] is monitoring biological signals like the body-trunk plethysmogram and respiration of the driver. [20] is using a neural network for image processing of the face of the driver.

Creating a similar in-vehicle alcohol detector using a brain-computer interface (BCI) with EEG signals could be of great interest. Using EEG signals instead of breath samples could potentially result in fewer sources of errors, as brain signals could be less affected by the external factors that affect the breathalyser. This could lead to a system with higher precision and better accuracy than the breathalyser.

1.1.1 Problem Description

The purpose of this thesis is to implement an alcohol detector using EEG signals. These EEG signals were collected from 20 participants. Each participant takes part in two separate recording sessions, where they are served an alcohol-based drink in one session and a non-alcoholic substitute in the other session. The resulting EEG dataset was used as input in a deep learning model. The purpose of the model is to classify whether features of alcohol drinking are present or not in the EEG signals, and by this create an alcohol detector. Figure 1.1 shows the entire alcohol detector project from start to finish.

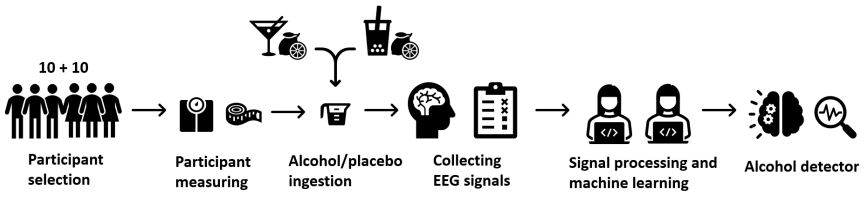


Figure 1.1: A graphical representation of the steps in the design of an alcohol detector [1, 2].

1.1.2 Related Work

[21] **Ziya Ekşi, Akif Akgül and Mehmet Recep Bozkurt. "The Classification of EEG Signals Recorded in Drunk and Non-drunk People." In: *International Journal of Computer Applications* Vol. 68, No. 10, April 2013**
This article aimed to design a system that can detect if a participant is drunk or not based on their EEG signals. The Yule-Walker method was applied as preprocessing to the data, and the classification was done by using an artificial neural network. The data set consisted of EEG data from 50 drunk people and 50 non-drunk people. Here, the state of drunk was not defined clearly. The best results achieved by the network were obtained with 900 epochs of an unknown length, with a classification accuracy of 95%. With 300 and 1500 epochs, the accuracy was 80% in both cases.

[7] **Leila Farsi et al. "Classification of alcoholic EEG signals using a deep learning method." In: *IEEE Sensors Journal* Vol. 21, No. 3, February 1, 2021**
This paper introduces and compares two algorithms that are used for the detection of alcoholism using EEG signals. Algorithm 1 applied feature extraction methods to a data set and input this data to an artificial neural network. The performance of this algorithm was compared to Algorithm 2, which is a deep learning algorithm. Raw EEG data were used as the input for this algorithm. Algorithm 2 achieved an average classification accuracy of 93%, which is 7% higher than the average accuracy of Algorithm 1. In addition, the paper also indicates that Algorithm 2 outperformed the state-of-the-art algorithms in the literature.

[8] **Hamid Mukhtar, Saeed Mian Qaisar and Atef Zaguia. "Deep Convolutional Neural Network Regularization for Alcoholism Detection Using EEG Signals." In: *Sensors* Vol. 21, No. 16, August 2021**
In this paper, raw, normalised EEG signals were given as input to a convolutional neural network to detect alcoholism. The architecture of the network is compact, consisting of only four convolutional layers, with normalisation and pooling operations. To avoid overfitting, regularisation techniques such as dropout, batch normalisation and L1/L2 regularisation were added to the network. To evaluate the performance of the model, k-fold cross-validation was used, with $k = \{3, 5, 10\}$. Using $k = 10$ lead to the best test accuracy of 98%.

[22] **Jones, A.W. "Biochemical and Physiological Research on the Disposition and Fate of Ethanol in the Body." In: *Medicolegal aspects of alcohol*, 5th Edition, 2008**
In this research, Jones presents, among other topics, how factors such as gender, body mass index (BMI), body composition, type of alcohol, drinking duration, food intake and biological differences can affect the alcohol metabolism in the human body. Through this, the BAC evolution is also influenced. Another focus of the research is how different methods for measuring the alcohol content influence the measured BAC value. The BAC may vary depending on where in the body it is measured (i.e., the blood, the breath or in different tissues) and on what kinds of instruments are used. Sometimes, irregular spikes in the BAC curves appear when consecutive measurements are taken in intervals of 1-2 minutes. These spikes could be due to short bursts in the alcohol absorption rate caused

by a ”sudden and unpredictable opening and closing of the pyloric sphincter”. Jones mentions that the spikes can be more apparent when using a breathalyser compared to other instruments. This could be attributable to factors such as breathalyser characteristics and breathing patterns, among other suggestions.

1.1.3 What Remains to Be Done?

As presented in the literature above, a high emphasis has been placed on the health-damaging effects of alcohol by classifying people suffering from alcoholism. Hardly any work regarding the effects of mild drinking on healthy people has been done, especially concerning low levels of alcohol consumed during social activities. Only [21] has been investigating the classification of alcohol-influenced EEG signals from healthy people, by using an artificial neural network. This study also applied preprocessing techniques to the data set. Using deep learning methods with raw EEG signals as input has not been tried yet, although applying such methods has provided successful results when classifying alcoholic and non-alcoholic EEG signals.

1.2 Objectives

The overall aim of this thesis is to establish the basis of a BCI system for alcohol detection based on EEG signals, and to investigate the effect that a mild alcohol dose may have on selective attention, accuracy and time response to stimuli. The investigation is divided into two main objectives, and one sub-objective:

O1 Design of an experiment that can provide input information for the design of an alcohol detector system based on EEG signals and perform data collection

O1.1 Perform an analysis of the BAC and behavioural data collected during the experiment

O2 Evaluate the classification of alcoholic and non-alcoholic EEG signals using deep learning techniques on individual and general models

1.3 Approach

The central element of this work is the analysis of human EEG signals with and without the influence of alcohol. In this work, the EEG data of 20 healthy participants were recorded during an experiment where both alcoholic and non-alcoholic drinks were ingested. Before the data collection experiment was performed, extensive research was conducted to design the recording protocol. The first part of this research was done while working on the project theses [1, 2] which predates the work in this master's thesis. The articles presented in Section 3.1.1, as well as the articles mentioned in Section 2.3 and Section 2.4, were reviewed. Based on these, as well as the knowledge gained during the work with the project theses, adjustments were made to improve the designed protocol. This improved protocol was designed as explained in Section 3.1. The collected BAC and behavioural

data were analysed by using the software described in Section 3.3.2.

The second part of the research was conducted by reviewing related works for the design of the alcohol detector. The most relevant results of this research are presented in Section 1.1.2. After this research was done, the gathered information was used to implement and adapt the convolutional neural network (CNN) EEGNet as described in Section 3.5.2 and Section 3.6. Due to some technical issues with the EEG equipment, some of the collected data were not at full length. Adjustments were made in order to use as much of this data as possible in the training and testing of the models. The metrics presented in Section 3.7 were used to evaluate the performance of the models.

1.4 Contribution

The following are the contributions of this thesis:

- A gender-balanced EEG data set based on the data of 20 healthy people with two sessions, where one session is with alcohol and the other is a non-alcoholic session
- A participant-based classifier which can detect the presence of alcohol with an average accuracy of 90.7%
- An alcohol detector which can classify the presence of alcohol through EEG signals across participants with an average accuracy of 62.9%

1.5 Limitations

The duration of one data collection session is approximately 66 minutes. Due to individual differences in alcohol metabolism, each participant will most likely reach different stages of intoxication during this period [23]. Therefore, it is impossible to guarantee that all participants will reach the peak BAC value during the experiment. This leads to differences across participants in the data set, where some participants are more intoxicated than others.

The equipment used to collect the data set, Unicorn Hybrid Black, does not provide software to check the scalp-electrode impedance. When using consumer grade EEG recording equipment, impedances up to 10 kΩ are acceptable [24]. With no possibility of checking the impedance before or during the recordings, it is not possible to confirm whether the collected data are within the acceptable value range or not. Consequently, the quality of the collected EEG signals might not be as good as desired.

1.6 Outline

This thesis consists of five chapters. Chapter 2 presents the theoretical background material. Here, brain signals, the generation of EEG signals and the effect of alcohol on the brain are presented. Chapter 3 describes the protocol used for data collection, the resulting

data set and the methods used for classification. The BAC values, the Flanker task and the classification results are presented in Chapter 4. These results are discussed in Chapter 5. The conclusion and future work of the thesis are provided in Chapter 6. In Appendix A, B and C, the questionnaire, consent form and information letter to the participants are provided. An overview of the collected data set is presented in Appendix D. Additional results from the Flanker task are presented in Appendix E. In Appendix F, additional plots of confusion matrices from the general model are presented. Finally, a version of the submitted paper derived from the work of this thesis can be found in Appendix G.

2

Background

This chapter introduces brain signals and how these can be captured by EEG. Further, the chapter explores how ethanol, hereby referred to as alcohol, is metabolised in the human body, and its effects on the human brain. An introduction to sample sizes in EEG studies is presented, as well as the optimal channels for the detection of alcoholism. This chapter aims to present the necessary information needed for understanding the designed protocol for data collection and to understand how alcohol affects the brain. This chapter's purpose is also to provide an understanding of how the BAC evolves after alcohol ingestion.

2.1 Brain Signals

The human brain can be divided into four areas, as presented in Figure 2.1. Each of these areas, called lobes, are responsible for different functions. The frontal lobe, for instance, is responsible for decision-making, planning and movement control [25].

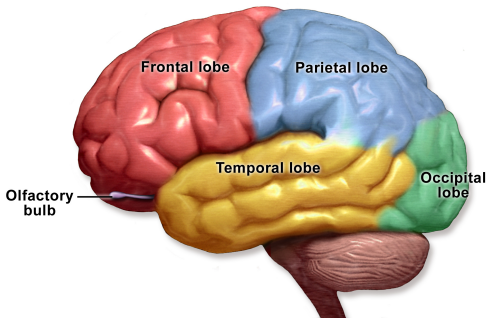


Figure 2.1: The four lobes of the human brain [26].

2.1.1 Neural Activity

There are two types of cells in the brain. These are called glial cells and nerve cells, or neurons. A neuron consists of dendrites, an axon and a cell body, as illustrated in Figure 2.2. The neuron can transmit electrical impulses, and brain signals, from the dendrites and along the axon. When the signal has reached the end of the axon, it triggers the release of a chemical transmitter known as a neurotransmitter. The neurotransmitter travels from the axon terminal bundle of one neuron to the dendrites of another neuron. This gap is known as a synapse. The travelling across the synapse stimulates the creation of a new electrical impulse [27].

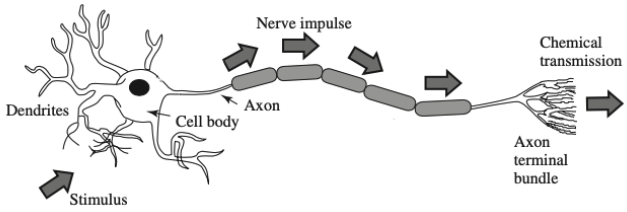


Figure 2.2: A neuron's structure. Adapted from [28].

2.1.2 EEG Signals

EEG is a non-invasive method used for studying the electrical activity generated in the brain by the activity of thousands of neurons in the cortical areas [28]. This activity is captured by electrodes placed on the surface of the scalp. The International 10-10 System, presented in Figure 2.3, is a system standard which describes the placement of the electrodes. It provides guidelines for the placing of up to 85 electrodes [29].

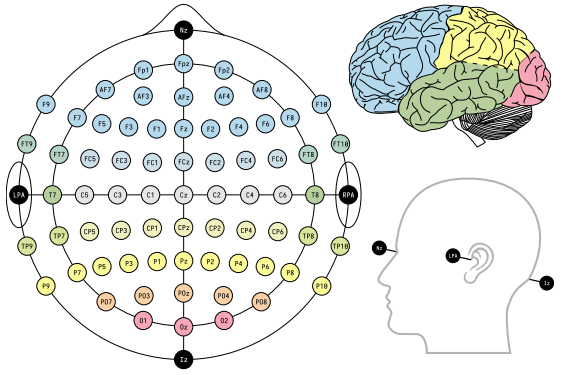


Figure 2.3: The 10-10 system for placement of electrodes [30].

EEG can be used to measure event-related potentials (ERPs), which are voltage changes

locked to a specific event or stimulus [31]. Such events or stimuli are typically related to cognitive, sensory or motor tasks [32].

EEG signals are characterised by low voltages, making them vulnerable to noise [33]. Noise that can be attributed to a specific source is called artifacts. Muscle contractions, heartbeats, blinking, movements of the eyes and power line interference can all be sources of artifacts [34]. In addition, a common problem in EEG studies is that EEG signals vary from participant to participant. The signals can also vary in recording sessions of a single participant performed at different times. These problems can be caused by a misalignment of electrodes or different head shapes across participants. The scalp-electrode impedance might also change over time [35].

2.1.3 Frequency Bands

EEG signals can be divided into five frequency bands, as shown in Table 2.1. Each frequency band is distinguished by a frequency range and is associated with a state of the brain. The alpha band can be divided into the slow alpha band (7.5-9 Hz) and the fast alpha band (9-12 Hz). A low dose of alcohol is known for increasing the slow alpha activity. Moderate doses of alcohol increase the activity in both the slow alpha band and the theta band. The effects of alcohol on the beta band are more ambiguous [36].

Table 2.1: The name of each frequency band together with their frequency range and the associated state of the brain. Adapted from [1].

Name	Frequency band [Hz]	Associated state
Delta (δ)	$0.5 - 4$	Deep sleep
Theta (θ)	$4 - 7$	Drowsiness
Alpha (α)	$7.5 - 12$	Relaxed awareness
Beta (β)	$12 - 30$	Active thinking
Gamma (γ)	> 30	Peak concentration

2.1.4 Brain-Computer Interface

A BCI is a system that allows direct communication between a user and an external device, such as a computer. To enable this communication, the brain signals from the user are recorded and interpreted by the BCI. The signals are then translated into commands which can be used to interact with the external device [37]. Until today, BCIs have mainly been used to help paralysed or disabled patients interact with their environment [38].

2.2 Alcohol and the Brain

After alcohol is ingested orally, the body begins to absorb it within 10 minutes. The amount of alcohol in a person's bloodstream is indicated by the BAC value. After alcohol

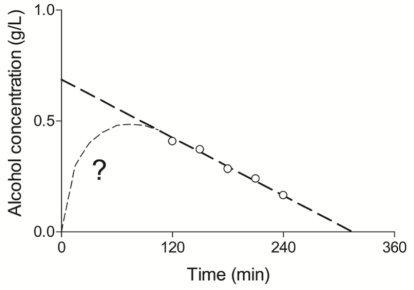
is consumed, the BAC value rises rapidly. The peak is usually reached 30-90 minutes after the consumption [39].

2.2.1 Alcohol Metabolism

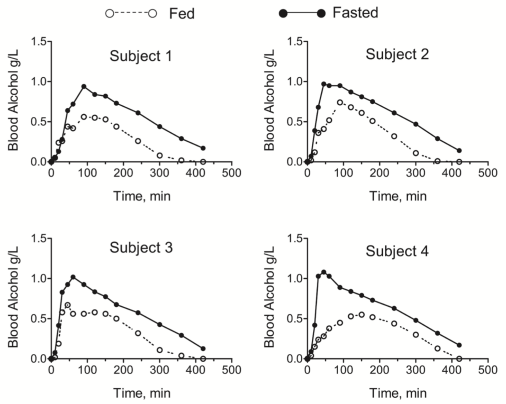
Alcohol metabolism is the body's ability to convert ingested alcohol into other compounds. Through this process, the alcohol is detoxified and eliminated from the bloodstream. This prevents it from damaging organs and cells within the body.

Several factors affect the metabolism of the body and, by that, the BAC [23]. One of these factors is how fast the alcoholic drink is ingested. Slow ingestion of alcohol leads to a lower BAC peak, while rapid consumption leads to a quickly rising BAC value and a higher BAC peak. The gender of the person consuming the drink is also affecting the metabolism. Even when the alcohol dose is adjusted for body weight, women tend to reach a higher peak BAC value compared to men [40]. This is because women tend to have a higher amount of body fat and a smaller amount of body water compared to men of the same weight [41]. When food is present in the stomach, alcohol is absorbed more slowly. This causes a lower peak BAC value [22].

Figure 2.4a depicts the typical evolution of a BAC curve. The first part of the curve, the absorption phase, is the most unpredictable part of the BAC curve. As stated above, it can depend on factors such as gender, age and BMI [22]. Figure 2.4b shows how BAC curves can vary between individuals who have received the same dose of alcohol, both with and without food.



(a) A hypothetical BAC curve showing the unpredictable nature of the absorption phase. Adapted from [22].



(b) The BAC profile curves of four individuals after ingesting 0.8 g/kg ethanol on an empty and full stomach [22].

Figure 2.4: Illustrations of how the BAC curves of individuals can vary after receiving the same dose of alcohol.

2.2.2 The Effect of Alcohol on the Brain

One of the main impacts of alcohol on the brain is its ability to alter the synapses of neurons [42]. In other words, alcohol affects the transmission of signals between neurons in the brain. This affects how the lobes in the brain communicate. This disturbance can cause several problems. One example is the slowing down of the central nervous system, which can cause a person to speak, think and move slower than usual [43].

The impact of alcohol can also be observed in the frontal lobe of the brain. As mentioned in Section 2.1, this lobe is responsible for decision-making, planning and self-control. Alcohol consumption can impair these functions, leading to impulsive actions and decisions [25, 44]. Findings in [45] indicate that alcohol consumption can result in a reduced attention span. While these symptoms are typically associated with alcohol intoxication, research has shown that individuals may experience different reactions to alcohol [46].

2.3 Sample Size in EEG Studies

The replication of a scientific study is crucial when building confidence in results, and it is regarded as one of the fundamental characteristics of a study [47, 48]. An insufficiently large sample size is a reason for replication failure, and it may lead to the significance of the study being overestimated [49]. The sample size in EEG studies is often small, as the collection and processing of EEG data often are resource intensive [50]. When investigating the sample size in 150 ERP articles, [51] found that the average number of participants is 21.

2.4 Optimal EEG Channels for Detection of Alcoholism

As mentioned in Section 2.1.2, the 10-10 system provides up to 85 positions for electrode placements. While using all of these would provide information about the electrical activity in all brain regions, it would also result in both a resource-intensive data collection and a computationally expensive analysis. [52] performed a discrete harmony search to find which positions in the 10-10 systems were optimal for alcoholism detection. This search resulted in 12 optimal channels, presented in Figure 2.5. Using these 12 channels reduced the accuracy of alcoholism detection by only 2.63% compared to using 61 channels.

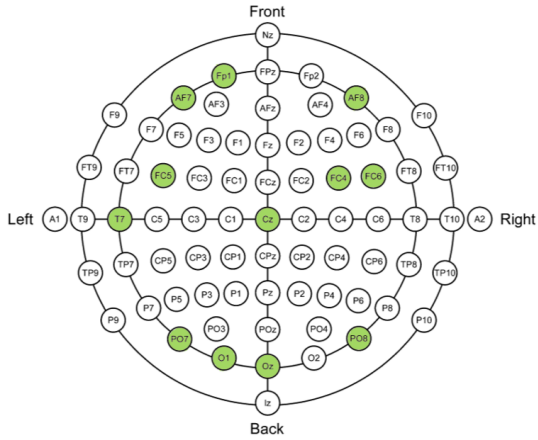


Figure 2.5: The 12 optimal channels for detection of alcoholism. The optimal channels are marked in green. Adapted from [53].

3

Materials and Methods

This chapter presents the designed protocol for the collection of EEG signals that are both affected and not affected by alcohol. First, the choices made in the design of the protocol are presented. Then, the resulting protocol is described in detail, as well as the utilised hardware and software. Next, the chapter presents the resulting data set and the preprocessing of this. The neural network used for the classification of the EEG data, EEGNet, is introduced, and the alterations made to EEGNet are also explained. Last, the training and testing of the implemented models are mentioned, as well as the metrics used to evaluate their performance.

3.1 Experimental Design

3.1.1 Pre-Design Research

Before the protocol for the data collection was designed, the following related literature was reviewed. In all the presented studies, young men were given alcohol while their EEG signals were measured. All studies found changes in the slow alpha, theta, or beta frequency bands after alcohol consumption. As mentioned in Section 2.1.3, alcohol is known for causing changes in these frequency bands. Thus, these studies were advised when creating the protocol, as the given doses are documented to affect the EEG signals. All following doses are given in either g/kg or ml/kg. g or ml refers to the amount of ethanol, while kg refers to the participant's weight.

[36] describes an experiment where 24 young males were administered low doses (LD) of alcohol and a placebo drink. The dose of alcohol was set to 0.75 ml/kg. 95% ethanol was mixed with a "sugar-free, non-caffeinated, carbonated beverage" to obtain a 20%-by-volume drink. The placebo drink was made by mixing the same drink mixer with 1 ml of ethanol. This ethanol, called a primer, was added to make the participants unsure whether their drink was alcoholic or not. The experiment lasted for 3 hours.

In [54], 37 healthy and young men received either an alcoholic consisting of 0.5 g/kg ethanol, or a non-alcoholic drink. The alcoholic drinks consisted of vodka and orange juice. The placebo drinks had a placebo cocktail mix instead of vodka. All participants were right-handed. Before the experiment, each participant was served breakfast. Two baseline EEG recordings of 4 minutes were completed before the drink was served. 45 minutes after the drink, new EEG recordings were done. The participants were then served another glass of the drink, and the waiting and recording of EEG signals were repeated. The BAC of each participant was measured before and after each EEG recording. The total time of the experiment was 2.5 hours.

[55] presents an experiment where 21 healthy, young men were given a low dose (LD), high dose (HD) and a placebo drink in a random order. The LD was set to 0.5 ml/kg. The HD was set to 0.8 ml/kg. Only one type of drink was administered per session, and there was at least one day between each of the three sessions. For the placebo condition, the participant was given a volume of ginger ale equal to 2.0 ml/kg. The alcoholic drinks consisted of either one part LD or HD ethanol and three parts ginger ale. Each session lasted 140 minutes, giving a total experiment time of approximately 7 hours per participant.

3.1.2 Key Protocol Parameters

Based on the related literature presented in Section 3.1.1, and the available time and equipment, the protocol was designed.

Participant Selection

In Section 3.1.1, all studies had male participants in their twenties or thirties. All participants were healthy and neither abstained from alcohol nor suffered from alcohol use disorders. The average number of participants for these studies was approximately 27. In Section 2.3, the average sample size for 150 ERP studies was found to be 21 participants. Based on these findings, the decision was made to set the total number of participants to 20.

The goal of this project is to create an alcohol detector that works on all drivers. Therefore, both males and females were included. To get a gender-balanced data set, the decision was made to choose 10 males and 10 females for the group of 20 participants.

To make the data from the participants easily comparable, a set of selection and exclusion criteria for the participants was set. These criteria are seen in Table 3.1. The criteria were chosen based on the information presented in the articles in Section 3.1.1. The participants who fit the criteria were chosen based on the results of the questionnaire presented in Appendix A. The chosen participants were also instructed to read and sign the project's consent form, seen in Appendix B.

Alcohol Dose

In all experiments in Section 3.1.1, the alcoholic drinks were based on either vodka or pure ethanol. The alcohol was mixed with either a juice or a carbonated beverage. With

Table 3.1: A summary of the selection and exclusion criteria for the participants of the project.

Selection criteria	Exclusion criteria
20-30 years old	No history of drug or alcohol abuse
Right-handed	No history of drug or alcohol abuse in close family
Social drinker	No major medical issues or history of psychiatric problems

these articles in mind, the decision was made to serve a drink consisting of orange juice and standard 40%-by-volume Smirnoff vodka. The Smirnoff vodka was chosen as pure ethanol was not possible to obtain. For the non-alcoholic drink, the vodka was replaced by an equal volume of a non-alcoholic, vodka-flavoured mixer diluted with water. The drinks were mixed with a ratio of one part vodka or vodka mixer to three parts orange juice.

To choose the alcohol dose, the previously mentioned articles and Table 23.1 and Table 23.2 in [56] were consulted. [56] reviews and summarises alcohol-related EEG studies. The doses presented in these tables were used to calculate the mean and median for an LD, HD and a single dose option, and the results are presented in Table 3.2.

Table 3.2: Calculated mean and median doses from 35 EEG and ERP studies. Adapted from [2].

	Single dose	LD	HD	Mean
Mean	0.452 g/kg	0.348 g/kg	0.604 g/kg	0.468 g/kg
Median	0.400 g/kg	0.300 g/kg	0.660 g/kg	0.453 g/kg

Based on these results, and the doses described in Section 3.1.1, the alcohol dose was set to 0.45g/kg alcohol. The selected drink doses for both the alcoholic and the non-alcoholic drink are presented in Table 3.3.

Table 3.3: A summary of the drink doses given to the participants. The volume of alcohol served was calculated based on the ethanol content per ml vodka stated on the Smirnoff bottle.

Drink dose	Alcoholic drink	Non-alcoholic drink
Ethanol	0.45 g/kg	0.00 g/kg
Vodka mixer and water	0.00 g/kg	0.45 g/kg
Orange juice	$3.00 \cdot (0.45$ g/kg)	$3.00 \cdot (0.45$ g/kg)

Chosen EEG Channels

The 12 optimal channels for the detection of alcoholism mentioned in Section 2.4 were chosen for the data collection. As the available equipment has 16 channels, this placement resulted in 4 leftover electrodes. These 4 electrodes were placed to make the electrode placements symmetrical. The final placement of the electrodes is shown in Figure 3.1.

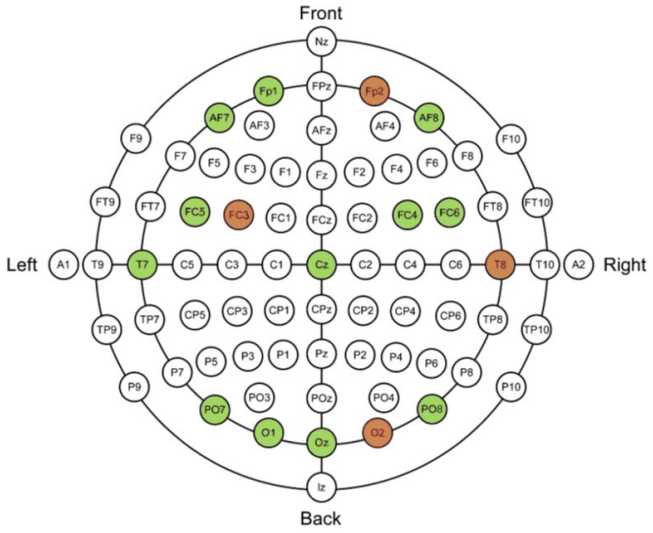


Figure 3.1: The 16 channels used for the recording of EEG signals. The red channels are the four leftover channels added to the optimal channels for symmetry. Adapted from [53].

The Flanker Task

The participants were instructed to perform the Flanker task [57] during the recordings. The task measures the selective attention, accuracy and response time (RT) of the participants. This is of interest as alcohol is known for affecting both the attention span and the RT of intoxicated people [45]. The Flanker task also tests a person's ability to ignore irrelevant stimuli around a focal point. This can be compared with how a driver needs to keep their focus on the road, while still retaining an overview of the surroundings.

During this version of the Flanker task, the participant is presented with five letters. They are instructed to press either the A-key or the L-key on the keyboard depending on the middle letter. If the middle letter is either X or C, they should press A. If the middle letter is V or B, the correct response is L. The middle letter is flanked by four identical letters. These are all either X, C, V or B. Consequently, 16 combinations of letters can be shown to the participant. If the flanked letters and the middle letter correspond to the same response, the trial type is called congruent. If not, the trial type is called incongruent.

The participants are presented with all combinations presented in Table 3.4 in random order. This sequence is called a block. Between each combination of letters, a cross is presented for two seconds before a new combination appears. There are a total of 6 blocks in the task, meaning the participants are presented with 96 sequences of letters in total. 48 of these are congruent and 48 are incongruent. After each block is presented, there is a seven-second break. Among other parameters, the RT and the response of the participant are recorded.

Table 3.4: The stimuli presented during one block in the Flanker task.

Congruent	Incongruent
XXXXX	XXVXX
XXCXX	XXBXX
CCCCC	CCVCC
CCXCC	CCBCC
VVVVV	VVXVV
VVBVV	VVCVV
BBBBB	BBXBB
BBVBB	BBCBB

Recording Time

The articles in Section 3.1.1 describe a total recording time of 2.5-7 hours per participant. With 20 participants, this would lead to a total recording time of 50-140 hours with 2 sessions per participant. This does not include the time needed before each session for preparation and after each session for clean-up and debriefing. Considering this, the total recording time was set to about 2 hours per participant, giving about 1 hour per session.

To ensure enough EEG data would be collected, each EEG recording was set to 5 minutes. The drink ingestion period was set to 10 minutes. Making the participants drink during such a short period would cause the BAC to rise faster and therefore also reach the peak faster, as mentioned in Section 2.2.1. The Flanker task lasted approximately 7 minutes. After each part of the session, there was a break of either 2 or 5 minutes. The BAC of the participant was measured before the first EEG recording, as well as in each 5-minute break starting 15 minutes after the drink ingestion period ended.

Since the desired session time was approximately 1 hour, the decision was made to perform a total of four EEG recordings during a session. This gave a total session time of approximately 66 minutes, meaning the total experiment time for each participant was just above 2 hours.

3.2 The Protocol and Data Collection

The general outline of the data collection session is visualised in Figure 3.2. The first part, the pre-experiment recordings, took place before the participant was given either an alcoholic or a non-alcoholic drink. After the drink ingestion period of 10 minutes, the post-drink recordings commenced. The following sections describe each part of the session, as well as the necessary preparations, in more detail.

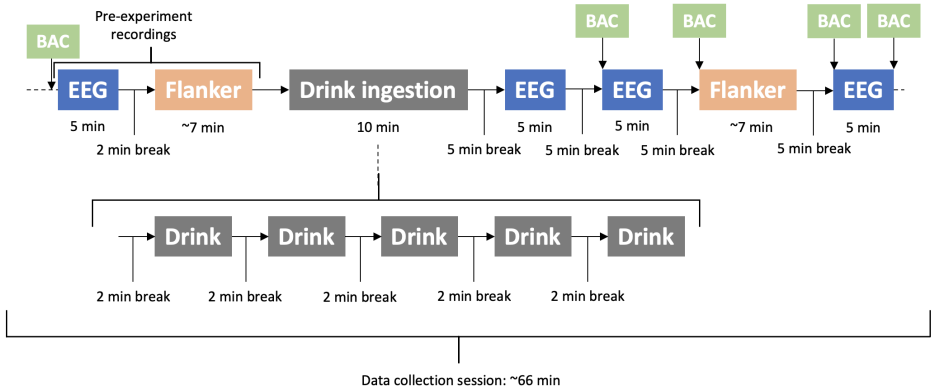


Figure 3.2: A detailed overview of the data collection session.

3.2.1 Preparation

Before the data collection sessions, the participants were given the information letter presented in Appendix C. Here, they were instructed to not drink caffeine on the same day as the recording. They were also instructed to abstain from alcohol the evening before and the morning of the recording.

Since the drink dose was determined individually based on each participant's body weight, the weight of each participant was measured before their first session. The height and the head size of the participant were also measured.

3.2.2 Pre-Experiment Recordings

The pre-experiment recordings are illustrated in Figure 3.3. At the beginning, a breathalyser was used to confirm that the participant was not affected by alcohol. After this, the recording session started with a 5-minute EEG recording. During these five minutes, the participant sat in a chair while relaxing with their eyes open. The participant was instructed to focus their eyes at one point and to sit as still as possible. They were not allowed to talk during the EEG recording.

After this recording, the participant was given two minutes to read the instructions of the Flanker task. After the two minutes, they performed the Flanker task as described in Section 3.1.2. The EEG signals of the participant were recorded while performing the task.

3.2.3 Drink Ingestion

After the pre-experiment recordings, the drink ingestion began. During a period of 10 minutes, the participant was given their drink divided equally into five cups. One cup was presented every two minutes. The participant was instructed to drink each cup as fast

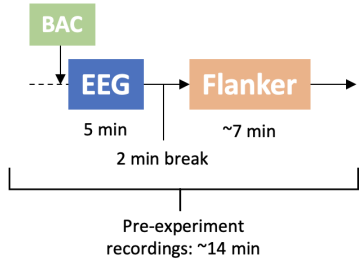


Figure 3.3: The pre-experiment recordings.

as possible. This was done to better control when the alcohol entered the body of the participant. Figure 3.4 summarises the drink ingestion part of the session. During these 10 minutes, no EEG signals were recorded.

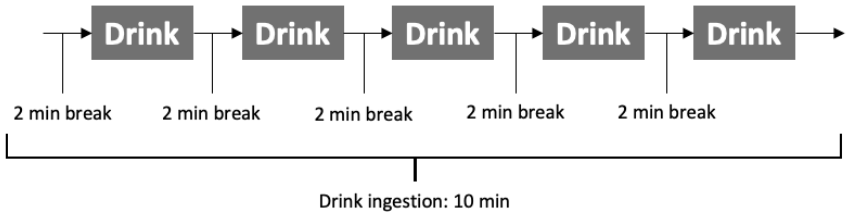


Figure 3.4: An overview of the drink ingestion period. Every two minutes the participants are instructed to drink a cup as fast as they can. The drink is divided equally into 5 cups, giving a total drink ingestion time of 10 minutes.

3.2.4 Post-Drink Recordings

As seen in Figure 3.5, the post-drink recordings consisted of three 5-minute EEG recordings and one Flanker recording. All recordings were performed as described in Section 3.2.2. Between each recording, there was a 5-minute break. BAC recordings were also performed. For these recordings, the breathalyser Alcoscan ALC-1 was used. Due to limitations in this breathalyser, as described in Section 3.3.1, the first BAC value was not recorded until 15 minutes after the drink ingestion.

The BAC measurements were done regardless of the type of drink the participant received. This was done to make sure the participant did not know which drink they were given. When they were given alcohol, the mouthpiece was changed between each recording to make sure there was no residual alcohol affecting the recordings. The mouthpiece was not changed between each recording when the participant received the non-alcoholic drink.

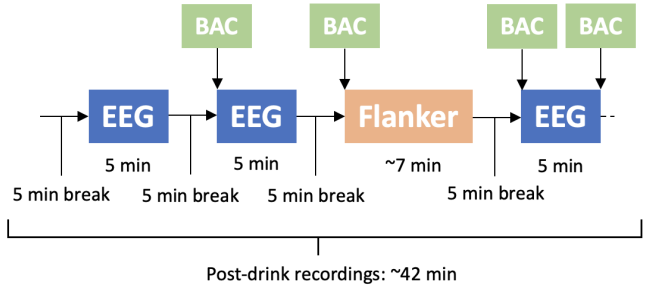


Figure 3.5: A visualisation of the recordings after the drink is ingested. The BAC values are recorded using a breathalyser.

3.2.5 Deviation From the Protocol

Originally, the order of the two sessions was to be random for all participants, meaning some would be served the alcoholic drink in their first session and the others would be served the non-alcohol drink first. This was done to make sure the participants would not know which drink they were given. If they knew they were given alcohol, they might alter their behaviour due to this, and not due to the effect of the alcohol itself. The first participant, P1, who was assigned alcohol in the first session, was convinced the drink contained alcohol. Due to this, the decision was made to serve each of the remaining participants the non-alcohol drink in the first recording session. In this drink, the vodka-flavoured mixer would add an alcohol-like flavour, but the flavour would not be strong enough to give away whether the drink did contain alcohol or not. Consequently, 19 out of 20 participants received the non-alcohol drink in the first recording session and an alcoholic drink in their second recording session.

3.3 Equipment

3.3.1 Hardware

Unicorn Hybrid Black

The available EEG recording equipment was the Unicorn Hybrid Black headset, seen in Figure 3.6. As one Unicorn Hybrid Black headset only has eight channels available, two headsets were used simultaneously to use the 12 optimal channels presented in Section 2.4. These were combined with the g.GAMMAcap from g.tec to increase the number of available electrode placements, as the cap from Unicorn Hybrid Black only has eight available positions.

(a) The Unicorn Hybrid Black electrodes with the g.GAMMAcap from g.tec.

(b) The experimental setup of the two Unicorn Hybrid Black devices.

Figure 3.6: The Unicorn Hybrid Black headsets with the chosen electrode placements. The 3D printed earclips secure the four earlobe references, two from each headset. The two headsets are secured to each other by using hair elastics and to the cap by using velcro.

Alcoscan ALC-1

The breathalyser Alcoscan ALC-1, seen in Figure 3.7, was used to measure the BAC of the participants. The accuracy of the breathalyser is $\pm 0.05‰$ at 1‰, giving a precision of 95%. Alcoscan ALC-1 cannot be used to measure the BAC value in the first 15 minutes after alcohol ingestion.

Figure 3.7: Alcoscan ALC-1, the breathalyser used to measure the BAC of the participants [1].

KERN MPE

KERN MPE is the device used to measure the height and weight of the participants before the data collection. It is depicted in Figure 3.8

(a) The display of the KERN MPE machine shows the participant's weight.

(b) The vertical measuring bar is used to measure the participant's height.

Figure 3.8: Pictures of the KERN MPE machine which measures the weight and height of the participants [1].

3.3.2 Software

Code

The code written for this project is found in this GitHub[1]. As of June 2023, the repository is private. Access can be provided upon request.

Python Libraries

The Expyriment library was used to create the Flanker task. The results were analysed using the Pandas and Matplotlib libraries. To perform the statistical analysis of the Flanker results, the `ttest_ind` function from the SciPy library was used. The MNE library was used for inspection of the collected data. When creating the classifier, the NumPy library was used for data loading. The scikit-learn and TensorFlow libraries were used for normalisation of the data and for implementing the classifiers.

[1] https://github.com/wavesresearch/Alcohol_Detection_Project_2022-2023

Lab Streaming Layer

Lab Streaming Layer (LSL) is an open-source software system. It enables synchronous streams and recordings of neural, physiological and behavioural data. For this project, LSL was utilised in the recording of the EEG data and the Flanker task data. Each time there was an event in the Flanker task, a marker was sent to the LSL stream. The events and the corresponding markers can be seen in Table 3.5. For both the recording of the EEG data and the Flanker data, time stamps were also sent in the stream.

Table 3.5: The marker sent to the LSL stream with the corresponding event in the Flanker task. Adapted from [1].

Event	Marker
Cross presented	'1'
Congruent stimulus presented	'2'
Incongruent stimulus presented	'3'
Break between blocks	'4'
Flanker task started	'5'
Flanker task ended	'6'
A-key pressed	'97'
L-key pressed	'108'
No key pressed	'None'

LabRecorder

LabRecorder[2] is the program that records the LSL streams. It allows the combination of all LSL streams into a single file. All combined streams are time synchronised using the time stamps and saved in an XDF file.

3.4 Data Set

The key parameters of the 20 participants who took part in the data collection are shown in Table 3.6. The BMI of each participant was calculated using Equation 3.1.

$$BMI = \frac{Weight\,[kg]}{Height\,[m]^2} \tag{3.1}$$

Each EEG recording was planned to last for five minutes, but due to trouble with the recording devices, some recordings were shortened, or no data were recorded at all. An overview of the length of each recording for all participants can be seen in Appendix D.

[2] https://github.com/labstreaminglayer/App-LabRecorder

Table 3.6: Overview of all participants and their relevant parameters.

Participant	Gender	Age	Weight [kg]	Height [cm]	BMI
P01	Female	24	67,9	174	22.4
P02	Female	23	85,5	161	33.0
P03	Female	24	84,1	174	27.8
P04	Female	23	72,8	165	26.7
P05	Female	25	60,6	165	22.3
P06	Male	23	64,1	182	19.4
P07	Female	24	70,3	169	24.6
P08	Male	24	72,5	177	23.1
P09	Male	24	73,5	182	22.2
P10	Female	23	61,4	162	23.4
P11	Male	25	83,6	188	23.7
P12	Male	25	88,5	191	24.3
P13	Female	24	57,1	166	20.7
P14	Male	23	93,2	187	26.7
P15	Female	28	70,2	178	22.2
P16	Female	25	81,6	165	30.0
P17	Male	23	92,2	191	25.3
P18	Male	23	81,5	184	24.1
P19	Male	24	95,1	194	25.3
P20	Male	23	72,4	175	23.6

3.4.1 Preprocessing of Data

Before the collected data set was used as input to the classifiers, some preprocessing was applied. First, the 5-minute recordings were split into epochs. After this, the data were split into a training and a test set. The length of the epochs and how the data were split into training and test sets for each classifier is presented in Section 3.6.

The data set was not subjected to any artifact removal or filtering procedures; only normalisation was performed. [58] suggests that normalisation of the data before the training of a neural network is crucial for obtaining good results. By applying normalisation, the wide ranges in the raw EEG signals are reduced [8].

The standard scaler was used to normalise the data by removing the mean and scaling to unit variance. The standard score z, of a sample x, is calculated by using Equation 3.2. μ is the mean of the training samples, and σ is the standard deviation (SD) of the samples. An overview of the preprocessing of the data can be seen in Figure 3.9.

$$z = \frac{x - \mu}{\sigma} \tag{3.2}$$

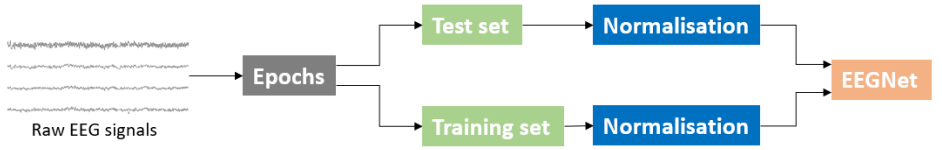


Figure 3.9: A visualisation of the preprocessing of the data. After the data is split into epochs, it is split into a training and a test set. The two sets are normalised separately before the data is fed into the model, EEGNet.

3.5 EEGNet

To create models for the classification of EEG signals, a CNN was used. In this thesis, the CNN architecture EEGNet was chosen to create the classification model for alcohol detection. EEGNet was made specifically for the classification and interpretation of EEG signals. It is known for performing well on different types of EEG signals, even when data is very limited. EEGNet has been shown to perform as well as other more paradigm-specific EEG CNN models, but EEGNet has 10^2 fewer parameters than these models. This makes EEGNet more computationally efficient [59].

3.5.1 Architecture

Figure 3.10 shows the architecture of EEGNet. EEGNet uses temporal convolution in the first layer to learn frequency filters. After this, EEGNet uses depthwise convolution. The purpose of this is to provide a direct way of learning spatial filters for each temporal filter. This enables efficient extraction of frequency-specific spatial filters. The regularisation technique dropout is applied to prevent overfitting.

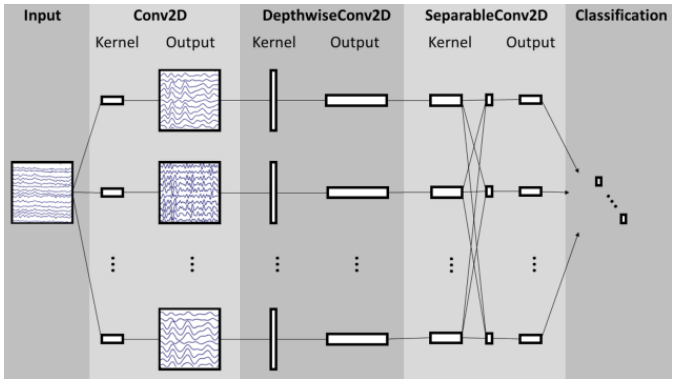


Figure 3.10: A visualisation of the architecture of EEGNet [59].

Separable convolution is used to separate the process of learning to summarise individual feature maps over time, from the process of optimally combining them. Finally, EEGNet

has a fully connected layer with softmax activation. This produces a probability distribution over the possible class labels. EEGNet has several hyperparameters, including F_1, D and F_2. These hyperparameters control the number of temporal filters, the depth of the network and the number of pointwise filters, respectively.

3.5.2 Alterations

The creators of EEGNet have provided a Keras implementation of EEGNet. The implementation can be found in this Github[3] repository. This implementation has been used as a foundation for the classifiers developed in this thesis. Some alterations have been applied to the original implementation as there are differences in the used data sets. The original implementation is made for data sampled at 128 Hz, while the data collected in this thesis were sampled at 250 Hz. Because of this, the length of the temporal convolution in the first layer is set to 125, as it is intended to be half the sampling rate.

The original implementation is built for multi-class classification. The objective of this thesis is to perform binary classification for classifying whether a signal is affected by alcohol or not. Therefore, the activation function in the output layer is chosen to be sigmoid instead of softmax and the number of classes is set to one.

3.6 Implemented Models

Three classifiers were made to detect alcohol-affected EEG signals. All were implemented using EEGNet with the alterations described in Section 3.5.2. All models were optimised by using the Adam algorithm [60] with a learning rate of 0.001. The used loss function was *binary cross entropy*. The hyperparameters of EEGNet were chosen to be their default values; $F_1 = 8$, $D = 2$, $F_2 = 16$. Two of the implemented classifiers are individual models, and the third classifier is a general model. All three models are presented in the following sections.

3.6.1 The Individual Models

Both individual models were trained and tested on data from the same participant.

The Random Individual Model

In the random individual model, the data from each participant was split into epochs of 5, 10, 20 and 30 seconds. After splitting, 20% of the data were used as the test set, while the remaining 80% of the data were used as the training set. The splitting of test and training data was done randomly, meaning epochs from the same 5-minute recording can be present in both the training and the test set. The same epoch is not present in both sets.

[3] https://github.com/vlawhern/arl-eegmodels

As the data set for each participant is quite small, 5-fold cross-validation was used to train and validate the model. The cross-validation was performed by splitting the training data for each participant into five subsets. Four of the five subsets were used for training, while the last was used for validation. This was repeated five times. Each time, the accuracy of the validation set was stored, and in the end, an average of all the accuracies was calculated and used as an estimate for the overall model accuracy.

This estimate of the model accuracy was then used to determine which epoch length lead to the best performance. This epoch length was then used on the test set to get a final evaluation of the performance of the random individual model. The best-performing epoch length for this model was also used for the two other models.

The Individual Model Across Sessions

In this model, all epochs from the same 5-minute recordings were placed in the same set, instead of splitting epochs randomly into the training and test set. One non-alcoholic recording and one alcoholic recording were chosen randomly to be in the test set of each participant. The remaining data were used in 3-fold cross-validation, as seen in Figure 3.11. The test set was used for the final evaluation of the model.

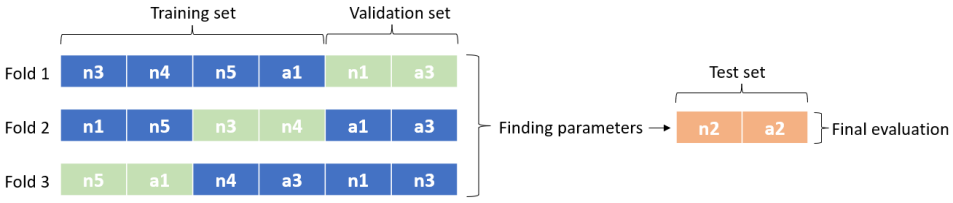


Figure 3.11: An example of how the data were divided for the individual model across sessions. The blue boxes represent training data and the green boxes represent validation data. The orange boxes represent the test data used for the final evaluation of the model. 3-fold cross-validation was used to find the best-performing hyperparameters for the model. a1-a3 are the alcohol-affected recordings, while n1-n5 are the 5-minute recordings not affected by alcohol.

3.6.2 The General Model

For the general model, EEGNet was trained using data from 19 of the 20 participants and then tested on the last, unseen participant. The hyperparameters of the model were chosen by using cross-validation.

3.7 Evaluation

To get an unbiased evaluation of the performance of the models, several evaluation metrics were used. These metrics describe the performance of the model on unseen data. A confusion matrix is a visual representation of the performance of the model. It displays the true

negative (TN) and true positive (TP) predictions on the diagonal. The anti-diagonal shows the number of false negative (FN) and false positive (FP) predictions.

By using Equation 3.3, the confusion matrix can be used to calculate the accuracy of the model. Although accuracy is commonly used, it can give an inaccurate description of the performance if the data set is imbalanced.

$$\text{Accuracy} = \frac{TN + TP}{TN + FN + TP + FP} \tag{3.3}$$

Precision is a measure of how well the model avoids false positive predictions. Recall describes the model's ability to correctly identify all positive cases. They are calculated by using Equation 3.4 and Equation 3.5, respectively.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.4}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3.5}$$

The F1 score, calculated by using Equation 3.6, is a performance metric that combines both the precision and the recall to provide a comprehensive evaluation of the overall performance of the model.

$$\text{F1 score} = 2 \cdot \frac{precision \cdot recall}{precision + recall} = 2 \cdot \frac{TP}{2TP + FP + FN} \tag{3.6}$$

The last metric used is specificity, which describes how well the model avoids predicting false negatives by using Equation 3.7.

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{3.7}$$

4

Results

In this chapter, the BAC values and the behavioural data from the Flanker task are presented. The classification results of the alcohol detectors are summarised. The average accuracy, precision, recall, F1 score and specificity for both the individual models and the general model are presented. Selected confusion matrices are also presented.

4.1 About the Results

For the Flanker task, all presented results are derived based on the average values of all participants. Individual results and other comparisons of the averages from all participants can be found in Appendix E. The significance of the average results presented in this chapter was analysed by performing the t-test to obtain the relevant p-values. Here, a p-value less than $p = 0.1$ indicates a tendency. A p-value less than $p = 0.05$ indicates significance.

The two individual models presented in this chapter were trained and tested as described in Section 3.6.1. The general model was trained and tested as described in Section 3.6.2. To evaluate the performance of these three classifiers, the five metrics presented in Section 3.7 were used. All metrics are given as a value in the range of 0 to 1, where 1 equals perfect classification within that metric. To ensure the evaluation of these models was unbiased, the test set used for the final evaluations of the models was never present in the training sets.

4.2 BAC Evolution

The measured BAC values after ingestion of alcohol for each participant are presented in Table 4.1. Before the recordings, all participants had a BAC value of 0.000‰. In the non-alcoholic recording session, all participants had a BAC value of 0.000‰ throughout the session. Figure 4.1 shows the average BAC values for males, females and all participants at each BAC measuring point during the alcoholic recording session.

Table 4.1: The BACs of all participants at approximately 15, 25, 37, and 42 minutes after (m.a.) alcohol ingestion.

Participant	Gender	15 m.a.	25 m.a.	37 m.a.	42 m.a.
P01	Female	0.450‰	0.440‰	0.430‰	0.450‰
P02	Female	0.270‰	0.330‰	0.380‰	0.400‰
P03	Female	0.420‰	0.420‰	0.500‰	0.540‰
P04	Female	0.440‰	0.430‰	0.490‰	0.450‰
P05	Female	0.440‰	0.450‰	0.470‰	0.490‰
P06	Male	0.310‰	0.390‰	0.400‰	0.430‰
P07	Female	0.110‰	0.140‰	0.190‰	0.210‰
P08	Male	0.250‰	0.260‰	0.290‰	0.320‰
P09	Male	0.420‰	0.430‰	0.440‰	0.450‰
P10	Female	0.340‰	0.260‰	0.260‰	0.360‰
P11	Male	0.550‰	0.530‰	0.540‰	0.500‰
P12	Male	0.390‰	0.370‰	0.460‰	0.480‰
P13	Female	0.520‰	0.470‰	0.490‰	*
P14	Male	0.430‰	0.470‰	0.480‰	0.480‰
P15	Female	0.530‰	0.480‰	0.570‰	0.560‰
P16	Female	0.480‰	0.460‰	0.580‰	0.540‰
P17	Male	0.250‰	0.280‰	0.290‰	0.380‰
P18	Male	0.480‰	0.380‰	0.470‰	0.520‰
P19	Male	0.320‰	0.310‰	0.340‰	0.400‰
P20	Male	0.630‰	0.470‰	0.480‰	0.490‰

* No measurement due to technical issue with the breathalyser

Figure 4.2 shows the BAC evolution of P11 (male) compared to the average BAC evolutions, both for all males and for all participants. As the figure shows, P11 had a different BAC curve than average with a clear drop from the penultimate to the last measurement. The averages for males and all participants show an increase in BAC value in the same interval.

Figure 4.3 shows the BMI of each participant plotted against that participant's average BAC values.

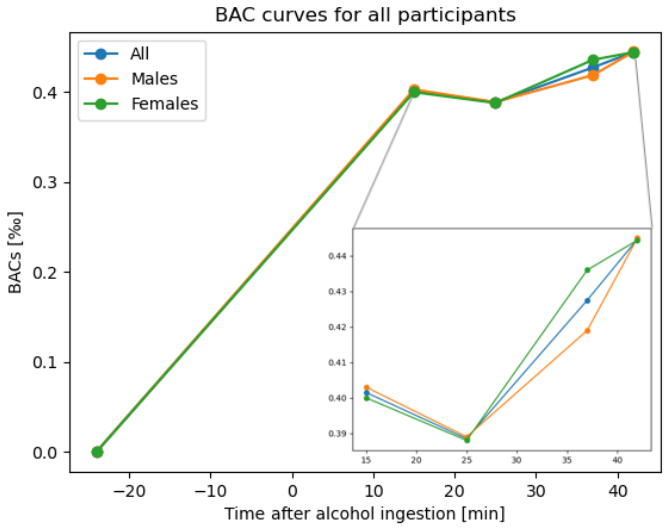


Figure 4.1: Average BAC values for males, females and all participants during the alcoholic recording session. The inserted window shows an enlarged version of the most relevant area.

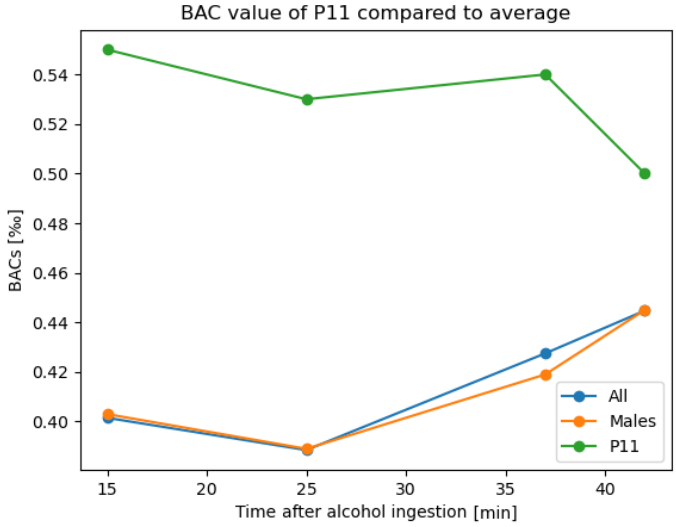


Figure 4.2: The measured BAC value of P11 at each BAC measuring point compared to the average of all participants and the average values of the males.

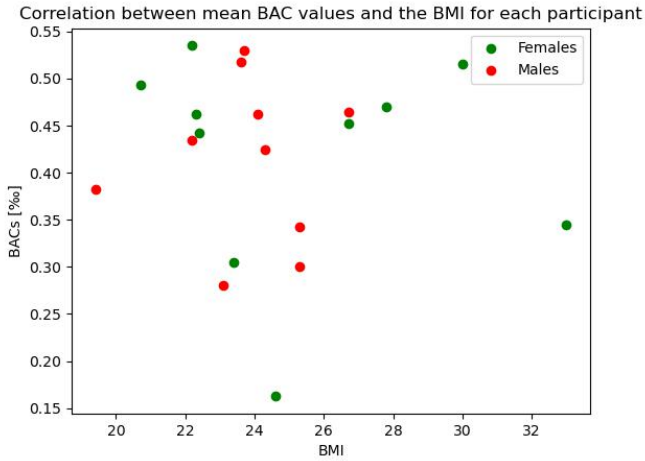
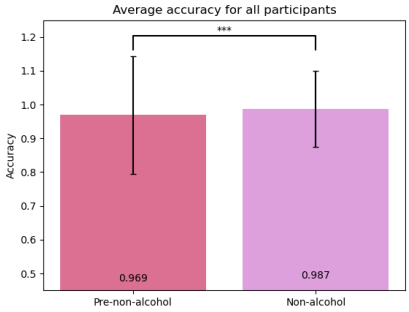


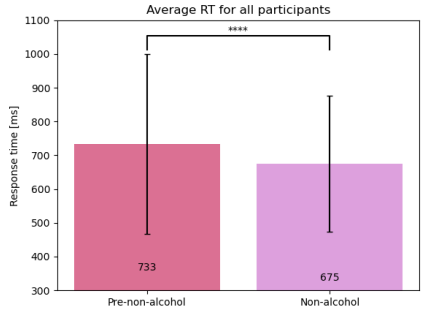
Figure 4.3: Average BAC values plotted against the BMI for each participant.

4.3 Behavioural Data From the Flanker Task

Figure 4.4 shows the average accuracy and RT before and after the ingestion of the non-alcohol drink. As Figure 4.4a shows, the average accuracy increases with a value of 0.018 from before to after the ingestion of the non-alcohol drink. With a p-value of $p = 0.00012$, this change is significant. For the average RT, seen in Figure 4.4b the value decreases by 58 ms. between the two Flanker tasks. The p-value of $p = 3.27 \cdot 10^{-14}$ indicates that this change is also significant.



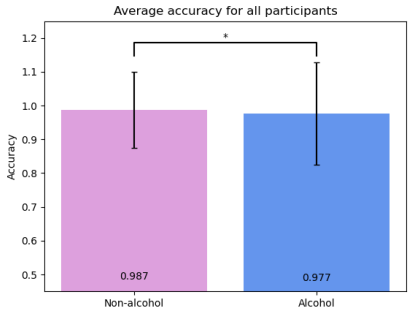
(a) Average accuracy before and after ingestion of the non-alcohol drink. Significance: *** $p = 0.00012 < 0.0005$.



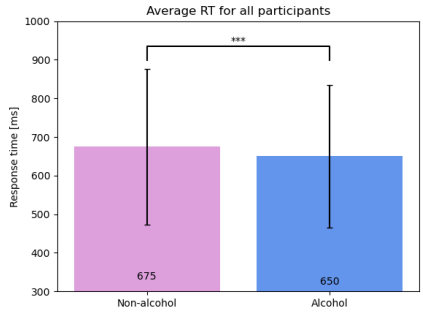
(b) Average RT before and after ingestion of the non-alcohol drink. Significance: **** $p = 3.27 \cdot 10^{-14} < 0.00005$.

Figure 4.4: Comparison of average Flanker task values before and after ingestion of the non-alcohol drink.

Figure 4.5 presents the average results from the Flanker tasks performed after drink ingestion. As seen in Figure 4.5a, the average accuracy for all participants decreases with a value of 0.01 from 0.987 to 0.977 when alcohol is present. The difference in accuracy is significant with a p-value of $p = 0.015836$. Figure 4.5b shows the average RT for all participants. After ingestion of alcohol, the RT decreased by 25 ms., from 675 to 650 ms. The difference in RT is significant with a p-value of $p = 0.000091$.



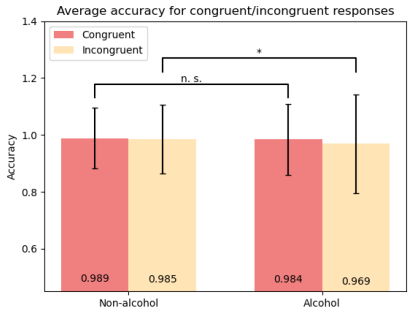
(a) Average accuracy. Significance: * $p = 0.015836 < 0.05$.



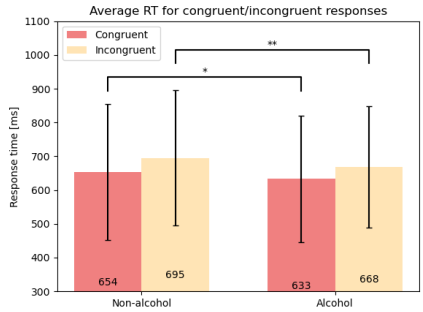
(b) Average RT. Significance: *** $p = 0.000091 < 0.0005$.

Figure 4.5: Average accuracies and RTs for the Flanker task.

Figure 4.6 shows the average congruent and incongruent results. Figure 4.6a presents the average accuracies for the congruent and incongruent responses. Figure 4.6b presents the average RTs.



(a) Average accuracy for congruent and incongruent responses. Significance: n.s. $p = 0.429890$, * $p = 0.014666 < 0.05$.



(b) Average RTs for congruent and incongruent responses. Significance: * $p = 0.016503 < 0.05$, ** $p = 0.001529 < 0.005$.

Figure 4.6: Average congruent and incongruent responses for the Flanker task.

For the congruent responses, the average accuracy decreases with a value of 0.005, from

0.989 to 0.984 from the non-alcohol to the alcohol-influenced Flanker task. The corresponding average RT decreased from 654 to 633 ms., giving a difference of 21 ms. For the incongruent responses, the average accuracy decreases from 0.985 to 0.969. Thus, the difference between the non-alcohol and alcohol-influenced Flanker task is 0.016. The average RT decreases from 695 to 668 ms., giving a difference of 27 ms. All changes were found to be significant, except for the change in congruent accuracy. Here, the p-value was $p = 0.429890$. This value is also too high to indicate a tendency ($p < 0.1$).

4.4 Detection of Alcohol Presence

4.4.1 The Individual Models

The Random Individual Model

Table 4.2 presents the average accuracies with SDs from the cross-validation of the random individual model. The average accuracy and SD are shown for epochs of length 5, 10, 20 and 30 seconds. As seen in Table 4.2 and Figure 4.7, the random individual model achieves the highest accuracies when the epoch length is 5 seconds.

As the model has the highest accuracy when the epochs are 5 seconds long, this epoch length was used when evaluating the performance of the random individual model on the test set. The performance on the test set for each participant is presented in Table 4.3. As seen, the maximum score is achieved for all participants, across all metrics.

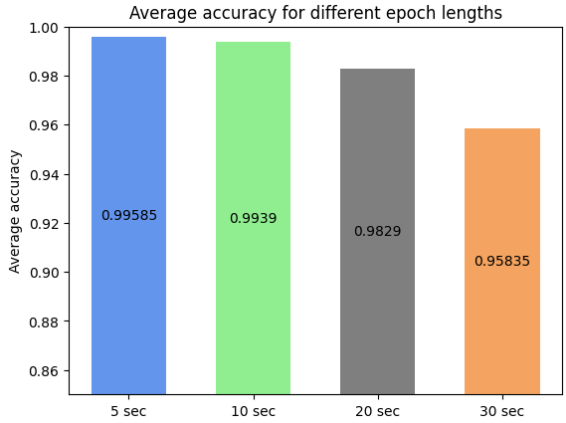


Figure 4.7: The average accuracies of the individual model for different epoch lengths. The accuracies presented are averages across all participants.

Table 4.2: Validation set accuracies of the random individual model for different epoch lengths when using 5-fold cross-validation.

Length of epochs	**5 seconds**	**10 seconds**	**20 seconds**	**30 seconds**
Accuracy (SD)	$\mu(\sigma)$	$\mu(\sigma)$	$\mu(\sigma)$	$\mu(\sigma)$
P01	1.000 (0.000)	0.945 (0.082)	1.000 (0.000)	0.928 (0.067)
P02	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
P03	1.000 (0.000)	1.000 (0.000)	0.968 (0.063)	0.967 (0.039)
P04	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	0.952 (0.062)
P05	1.000 (0.000)	0.988 (0.023)	1.000 (0.000)	0.880 (0.153)
P06	1.000 (0.000)	1.000 (0.000)	0.866 (0.146)	1.000 (0.000)
P07	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
P08	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	0.953 (0.092)
P09	1.000 (0.000)	0.969 (0.061)	0.968 (0.063)	1.000 (0.000)
P10	1.000 (0.000)	0.988 (0.022)	1.000 (0.000)	0.950 (0.099)
P11	0.985 (0.029)	0.988 (0.023)	0.951 (0.059)	0.819 (0.083)
P12	0.932 (0.135)	1.000 (0.000)	1.000 (0.000)	0.954 (0.061)
P13	1.000 (0.000)	1.000 (0.000)	0.969 (0.042)	0.907 (0.113)
P14	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
P15	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
P16	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
P17	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	0.981 (0.036)
P18	1.000 (0.000)	1.000 (0.000)	0.989 (0.021)	0.938 (0.089)
P19	1.000 (0.000)	1.000 (0.000)	0.958 (0.039)	0.938 (0.123)
P20	1.000 (0.000)	1.000 (0.000)	0.989 (0.021)	1.000 (0.000)

Table 4.3: The performance of the random individual model on the test set, when using epochs of length 5 seconds.

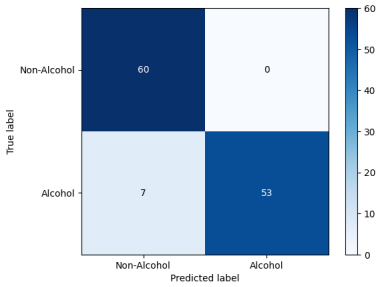
Participant	Accuracy	Precision	Recall	F1 score	Specificity
P01	1.000	1.000	1.000	1.000	1.000
P02	1.000	1.000	1.000	1.000	1.000
P03	1.000	1.000	1.000	1.000	1.000
P04	1.000	1.000	1.000	1.000	1.000
P05	1.000	1.000	1.000	1.000	1.000
P06	1.000	1.000	1.000	1.000	1.000
P07	1.000	1.000	1.000	1.000	1.000
P08	1.000	1.000	1.000	1.000	1.000
P09	1.000	1.000	1.000	1.000	1.000
P10	1.000	1.000	1.000	1.000	1.000
P11	1.000	1.000	1.000	1.000	1.000
P12	1.000	1.000	1.000	1.000	1.000
P13	1.000	1.000	1.000	1.000	1.000
P14	1.000	1.000	1.000	1.000	1.000
P15	1.000	1.000	1.000	1.000	1.000
P16	1.000	1.000	1.000	1.000	1.000
P17	1.000	1.000	1.000	1.000	1.000
P18	1.000	1.000	1.000	1.000	1.000
P19	1.000	1.000	1.000	1.000	1.000
P20	1.000	1.000	1.000	1.000	1.000
Average	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

The Individual Model Across Sessions

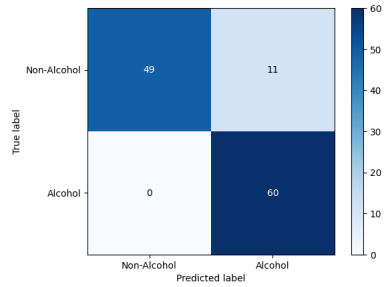
The results of the test set for the individual model across sessions are shown in Table 4.4. 13 of the participants scored perfectly across all metrics. The confusion matrices of the seven participants who did not score perfectly are presented in Figure 4.8. The lowest accuracy was obtained for P14, with an accuracy of 50.8%. The best-performing metric is precision, with an average of 99.2%, closely followed by specificity with 99.1%.

Table 4.4: The performance of the individual model across sessions on the test set, when the data is split into epochs of 5 seconds. Results marked in green are above the average value of that metric, and those marked in red are below average.

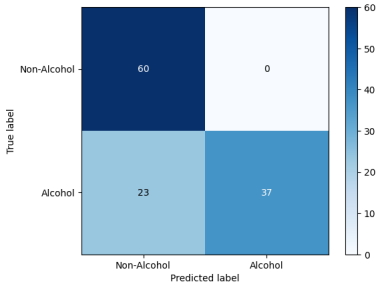
Participant	Accuracy	Precision	Recall	F1 score	Specificity
P01	1.000	1.000	1.000	1.000	1.000
P02	1.000	1.000	1.000	1.000	1.000
P03	1.000	1.000	1.000	1.000	1.000
P04	1.000	1.000	1.000	1.000	1.000
P06	1.000	1.000	1.000	1.000	1.000
P09	1.000	1.000	1.000	1.000	1.000
P10	1.000	1.000	1.000	1.000	1.000
P12	1.000	1.000	1.000	1.000	1.000
P13	1.000	1.000	1.000	1.000	1.000
P15	1.000	1.000	1.000	1.000	1.000
P17	1.000	1.000	1.000	1.000	1.000
P18	1.000	1.000	1.000	1.000	1.000
P20	1.000	1.000	1.000	1.000	1.000
P08	0.941	1.000	0.883	0.938	1.000
P11	0.908	0.845	1.000	0.916	0.816
P07	0.808	1.000	0.616	0.762	1.000
P16	0.808	1.000	0.616	0.762	1.000
P19	0.616	1.000	0.233	0.378	1.000
P05	0.555	1.000	0.466	0.635	1.000
P14	0.508	1.000	0.016	0.032	1.000
Average	**0.907**	**0.992**	**0.842**	**0.871**	**0.991**



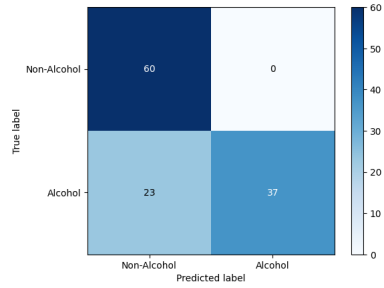
(a) P8



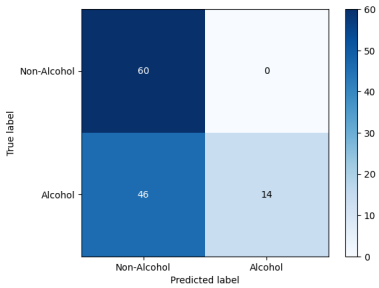
(b) P11



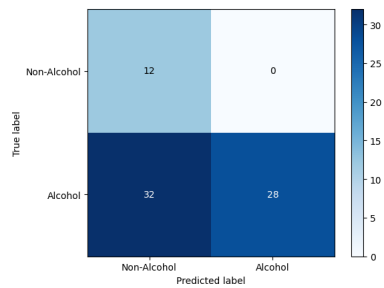
(c) P7



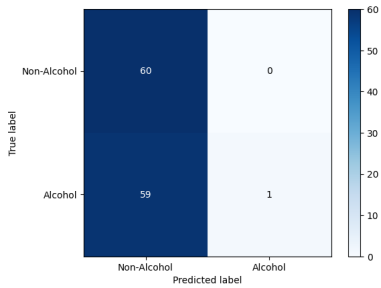
(d) P16



(e) P19



(f) P5



(g) P14

Figure 4.8: The confusion matrices of the participants from the individual model across sessions who did not achieve an accuracy of 100%.

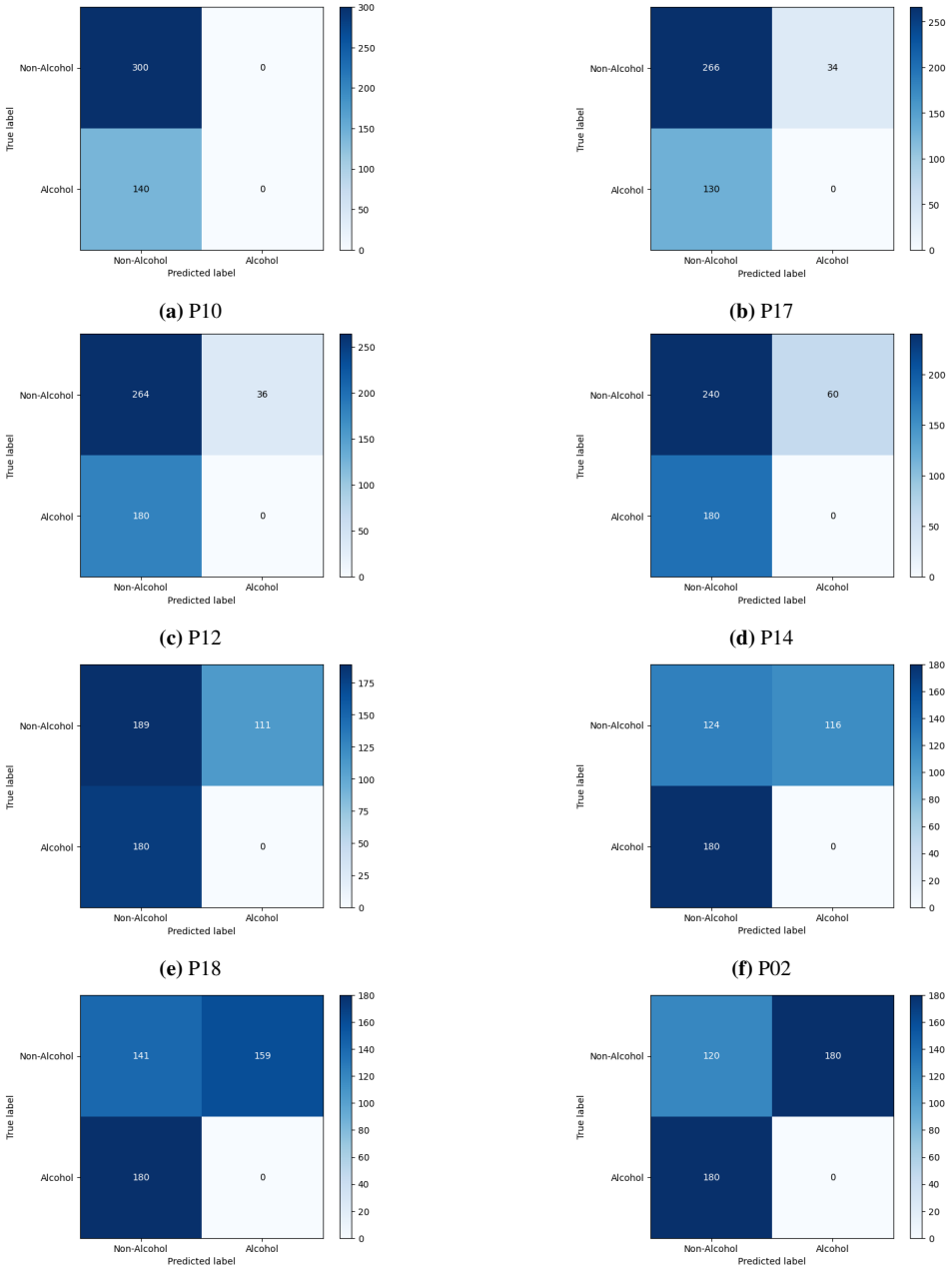
4.4.2 The General Model

Table 4.5 shows the results of the general model on the test set. The participant column shows which participant was used in the test set for each model run. The average accuracy across all participants is 62.9%. The best-performing metric is specificity, with a value of 81.5%. The best accuracies were achieved with P15 and P09 in the test set, with an accuracy of 94.1% and 90.2%, respectively. The lowest accuracies were obtained using P04, P19 and P02, resulting in accuracies of 25.0%, 29.4% and 29.5%, respectively.

Table 4.5: The performance of the general model on the test set. Results marked in green are above the average value for that metric, and those marked in red are below the average.

Participant	Accuracy	Precision	Recall	F1 score	Specificity
P15	0.941	1.000	0.856	0.922	1.000
P09	0.902	0.814	1.000	0.897	0.829
P13	0.875	0.750	1.000	0.857	0.800
P01	0.857	1.000	0.666	0.800	1.000
P16	0.760	1.000	0.361	0.531	1.000
P20	0.727	0.645	0.606	0.625	0.800
P05	0.694	1.000	0.267	0.431	1.000
P06	0.682	1.000	0.333	0.500	1.000
P10	0.682	0.000	0.000	0.000	1.000
P11	0.664	0.621	0.528	0.571	0.764
P08	0.642	1.000	0.044	0.085	1.000
P07	0.625	0.500	0.333	0.400	0.800
P03	0.625	0.500	0.666	0.571	0.600
P17	0.619	0.000	0.000	0.000	0.887
P12	0.550	0.000	0.000	0.000	0.880
P14	0.500	0.000	0.000	0.000	0.800
P18	0.394	0.000	0.000	0.000	0.630
P02	0.295	0.000	0.000	0.000	0.615
P19	0.294	0.000	0.000	0.000	0.470
P04	0.250	0.000	0.000	0.000	0.400
Average	**0.629**	**0.492**	**0.333**	**0.360**	**0.814**

Figure 4.9 shows the confusion matrices of the participants where none of the alcohol samples were classified correctly by the model. The remainder of the confusion matrices for the general model can be found in Appendix F.



(a) P10 **(b)** P17

(c) P12 **(d)** P14

(e) P18 **(f)** P02

(g) P19 **(h)** P04

Figure 4.9: The confusion matrices of the participants for which the general model did not classify any alcohol samples correctly.

5

Discussion

The purpose of this chapter is to discuss the results presented in Chapter 4. The BAC evolution, the Flanker results and the limitations of the EEG data are discussed, as well as the epoch length. Then, the individual models are summarised and compared. Last, the performance of the general model is evaluated.

5.1 BAC Evolution

As Figure 4.1 shows, the average BAC value for all participants decreases from the first to the second measurement after drink ingestion. This trend is also visible in the averages for males and females separately. The consecutive BAC measurements show increased average BAC values for all participants. Consequently, the peak BAC value was not reached at the second measurement 25 minutes after alcohol ingestion. Even though the shape of the BAC curve can vary highly during the absorption phase, as seen in Figure 2.4b, a decrease in the BAC normally indicates that the peak value has been reached and that the absorption phase is over. Therefore, this dip in value is most likely not due to the unpredictable nature of the absorption phase.

An explanation for this decrease can be the sudden opening and closing of the pyloric sphincter described in Section 1.1.2. This opening and closing can cause spikes in the BAC profile of a participant. It may seem unlikely that this spike occurred for all participants at the same time, but Figure 4.1 only shows average BAC values at each point of time. Table 4.1 shows that there were several participants (P02, P05-P09, P14, P17) who experienced an increase in BAC at 25 minutes after alcohol ingestion instead of a decrease. This means the pyloric sphincter may have opened at a later time for these participants than for the rest, and the pyloric spasms are still a plausible explanation for the dip in BAC value. As also mentioned in Section 1.1.2, the spikes in BAC values can be more apparent when using a breathalyser compared to using other methods for measuring the BAC. Therefore, the average dip in BAC value should not be emphasised. To avoid measuring a decrease like this, the BAC could be measured by using another instrument

than a breathalyser, or it could be measured in intervals of more than 10-12 minutes.

Considering the increasing trend of the BAC curves seen in Figure 4.1, most participants did not reach the peak BAC value. As described in Section 2.2, the peak is usually reached 30-90 minutes after the end of ingestion. The preferred outcome of the experiment would be to have the participants reach the BAC peak during the alcoholic recording session. This is to enable analyses of the behaviour of the participants while under peak BAC influence. As the two last BAC measurements were at 37 and 42 minutes after alcohol ingestion, the duration of each session was believed to be long enough to capture the peak. The participants were told to eat beforehand, and the chosen alcoholic drink was diluted with orange juice. These are both factors which can prolong the time until the BAC peak is reached. To increase the chances of the participants reaching the peak BAC, they could have been instructed to not eat beforehand, or they could have been served an undiluted alcoholic drink.

As seen in Figure 4.2, the BAC curve of P11 shows a distinct decline at the end compared to the rest of the participants. This participant seems to have reached the peak BAC value. Given that the conditions for the drink ingestion and BAC measurements were the same for all participants, the difference in BAC curves is probably caused by biological differences affecting the alcohol metabolism. After the end of the experiment, the above-average and declining BAC values of the participant were mentioned to the participant himself. Upon learning this, P11 informed the authors that he normally gets quite drunk when he drinks, but also that he sobers up quite fast. This indicates that P11 does have a high alcohol metabolism.

Figure 4.3 shows each participant's average BAC plotted against their corresponding BMI. There are several areas where some of the participants can be said to be placed linearly. For instance, the three females with a BMI in the range of 27 to 30 may indicate a linear correlation between BMI and BAC in that area. On the other hand, the number of samples in this plot is relatively small and, therefore, no conclusions can be drawn based on the presented results.

5.2 The Flanker Task

Figure 4.4 presents the changes in average accuracy and RT from the pre-non-alcohol to the non-alcohol Flanker tasks. Here, the accuracy increases after the non-alcohol drink ingestion, and the change is significant. The average RT decreases significantly for the same Flanker tasks. Since the presented drink was the non-alcohol drink, the expected change would not be a significant increase in accuracy. As mentioned in Section 3.2.5, the pre-non-alcohol Flanker task was the first time 19 out of 20 participants performed the Flanker task during the experiment. When performing a new test, it is normal to be nervous. This nervousness could cause the participant to answer slower than if they were familiar with the test, or to answer inaccurately. Since the changes between the two Flanker tasks are quite significant for both parameters, this nervousness is the most likely explanation for the observed differences.

Before the non-alcohol and alcohol Flanker tasks were performed by the participants, they had already performed the Flanker task once in the pre-experiment recordings. By making sure the participants had performed the Flanker task once before the counting Flanker tasks, the nervous element mentioned above was hopefully reduced as the participants knew what to expect of the task. This means that the changes in RT and accuracy from the non-alcohol to the alcohol task should only be caused by the introduction of alcohol.

Both the average accuracy, depicted in Figure 4.5a and the average RT, seen in Figure 4.5b decrease from the non-alcohol to the alcohol Flanker task. With p-values of p=0.015836 and p=0.000091, respectively, these changes are significant. Both decreases can be explained by the effect caused by alcohol entering the body. As described in Section 2.2.2, consuming alcohol causes a person to think and move slower than normal. It can also lead to impulsive actions and decisions. As the RT decreases after alcohol consumption, it is probably the impulsiveness caused by alcohol that has affected the participants. This would also explain the decrease in accuracy. When the participants make more impulsive decisions, they answer the tasks faster. This may lead to decreased accuracy as the participants may not have realised what the correct answer is before pressing a key.

As Figure 4.6 shows, the changes described above are also present for both the congruent and the incongruent responses. All average accuracies and average RTs decrease from the non-alcohol to the alcohol Flanker task. Here, all changes are significant except for the change in congruent average accuracy. The congruent tasks are the tasks where the flanked letters presented are compatible with the middle letter. Due to this, it is probably easier for the participants to react correctly to these stimuli than incongruent stimuli. This can explain why the decrease in average congruent accuracy is not-significant. Since the RT is longer for the incongruent stimuli than the congruent stimuli, both for the non-alcohol and alcohol-influenced Flanker tasks, the participants are probably affected by the irrelevant stimuli.

The Flanker task was chosen as a part of this experiment to test the participants' ability to filter the relevant information from the irrelevant. As described in Section 3.1.2, this can be compared to how a driver needs to be aware of both the road they are driving on and their surroundings. As the results above show, the consumption of alcohol seems to affect a person's ability to make the right decision as fast as needed. While the changes presented are small, they are still significant. And, as discussed in Section 5.1, these results are obtained before most participants have reached their peak BAC value, where they are most affected by the alcohol. Table 4.1 shows that the highest achieved BAC across all participants was P20's BAC of 0.630‰. While this is over the legal BAC driving limit of 0.2‰ in Norway, it is well below the limit of 0.8‰ set in a lot of other countries. This highlights how important it is to include tests such as the Flanker task in alcohol-related experiments to get an idea of how alcohol affects the selective attention and inhibitory function of the person consuming it.

5.3 Limitations of the EEG Data

As mentioned in Section 1.5, the scalp-electrode impedance was not possible to measure while collecting the data. Consequently, the impedance could be higher than the desired value for EEG signals, and the quality of the data could be lower than preferred. This could have led to the data being more challenging to classify than it could have been with a lower impedance.

During the collection of the data, construction work was performed outside the data collection room. At times, the construction noise was quite noticeable inside the room. A few participants completed their sessions without the presence of the construction noise, but most did experience it at some point during the experiment. This noise is, therefore, most likely affecting the collected EEG signals. Consequently, the noise interference could have affected the classification results negatively.

5.4 Epoch Length

As seen in Table 4.2 and Figure 4.7, the random individual model performs best for the epoch length of 5 seconds, which is the shortest length. A shorter epoch length results in more data points, and increasing the number of data points can, in general, lead to better CNN performance. Thus, the model performance increasing when using the shortest epoch length is as expected.

5.5 Comparison of the Individual Models

The random individual model performs better than the individual model across sessions. The two models have an average accuracy of 100% (Table 4.3) and 90.7% (Table 4.4), respectively. A reason for the worse performance of the latter model can be differences in the EEG signals across the recordings. These differences could have occurred due to the cap not being placed at the same spots on the two recording days. Even for the sessions recorded on the same day, the impedance between the electrodes and the scalp could have varied between the recordings, or the electrodes could have moved positions slightly.

An explanation for the exceptionally good performance of the random individual model could be that it has epochs from the same 5-minute recordings in both the training and the testing set. Thus, the model could have learned the differences between the recordings during the training. Even though the individual data samples from the same 5-minute recordings are not completely identical, they may cause some form of data leakage when divided into both the training and test sets. This is because the samples from the 5-minute recordings may have similar traits, and this could be the reason why the model achieved a perfect score.

The individual model across sessions, on the other hand, is evaluated on a test set with full 5-minute recordings. These recordings are not seen in any part of the training data. This

means it could be more difficult to classify the test data if there are any differences present, such as varying electrode placements and electrode-scalp impedances.

Although the random individual model is performing better, the individual model across sessions provides a more realistic implementation of an alcohol detector. If an alcohol detector were to be calibrated for individuals before use, the EEG signals used to detect alcohol might differ slightly from the signals used to calibrate the device. This is comparable to how the data is split in the individual model across sessions. Therefore, this model is viewed as the best of these models. This is despite it performing slightly worse than the other model, but the realistic use of data outweighs the better performance.

The confusion matrices in Figure 4.8 show the results from the individual model across sessions for the participants who did not achieve an accuracy of 100%. As seen, the model classifies almost all non-alcohol samples correctly. This means the lower accuracies for these participants are caused mostly by alcohol samples being misclassified. The majority of these samples are false negative classifications. From a real-world perspective, this is the worst kind of error for an alcohol detector since this means alcohol-affected drivers would have been classified as non-affected. While the model's average accuracy of 90.7% is quite good, the number of false positives affects the average precision, recall and F1 score negatively. Therefore, it is important to also use these metrics as an indicator of whether the model can be applied in the real world or not.

5.6 Classification Results of the General Model

As presented in Table 4.5, the average accuracy of the general model is 62.9%, but there are large differences in each participant's performance. The accuracies are in the range of 25.0% to 94.1%. These results are as expected since there might be differences in the EEG signals across participants, as described in Section 2.1.2. These differences indicate that training on a set of participants and testing on an unseen participant is challenging. This could be the reason for the lower performance of this model compared to the individual models, where the lowest accuracy obtained was 50.8%.

The general model is struggling to correctly classify alcohol samples, as indicated by the low precision and recall values and the low F1 score presented in Table 4.5. For eight of the participants, the model was not able to correctly classify any alcohol samples, as seen in the confusion matrices in Figure 4.9. For four of these participants (P02, P10, P17, P19), the BAC values were all noteworthy lower than the average values presented in Table 4.1 and Figure 4.1. Low BAC values could result in less clear alcohol features as the brain is less affected by alcohol. This can make the classification of alcohol-affected signals more difficult.

Another possible explanation for the low results across participants is differences in head shapes. The head shape affects how good the connections between the electrodes and the scalp are. During the data collection, two caps were used. One was a size medium, and one was a size large. The medium cap was a bit large for some of the female participants,

leading to little contact between the electrodes and scalp, especially at the back of the head. This may have been part of the reason for the poor classification accuracies of P02 and P04, as these participants were both females with smaller heads than the other participants.

The results obtained in this thesis are not higher than the 95% classification accuracy achieved by [21], which investigated the detection of alcohol-influenced EEG signals. A reason for this might be the limitations in the data set described in Section 5.3. However, with no information about either the demographics of the data set or the channels used in [21], it is difficult to perform a direct comparison between the results of that study and the results presented in this thesis.

As Table 4.5 shows, P09 and P15 achieved accuracies above 90%, and P01 and P13 achieved accuracies of 85.7% and 87.5%, respectively. The precision, recall and F1 scores of all these participants are also quite high. As discussed above, these metrics should also be taken into consideration when judging the performance of an alcohol detector. Based on these results, these four participants provide the belief that an EEG-based alcohol detector could eventually become a helpful tool in the prevention of drunk driving. However, the performance, in general, must be significantly improved to use this approach as an alternative to a breathalyser.

6

Conclusion and Future Work

In this chapter, the work of the thesis is concluded in light of the objectives presented in Section 1.2 and the previously presented discussion. In the end, the suggestions for future work are presented.

6.1 Conclusion

Section 1.2 presented the three objectives of this thesis. Objective **O1** was to design an experiment that provides input data for the alcohol detector. First, related works were reviewed, and key protocol parameters were discussed and decided. Then, the experiment was designed, resulting in a data set. The experiment is summarised in Figure 3.2.

Objective **O1.1** has been met by analysing the BAC data presented in Section 4.2 and the behavioural data presented in Section 4.3. As discussed in Section 5.1, the BAC peak value was probably not reached for most of the participants, apart from P11. Section 5.2 discusses how the participants performed on the non-alcohol Flanker task compared to the alcohol Flanker task. All average accuracies and RTs decreased, and all changes were found to be significant, except for the decrease in average congruent accuracy. The results of the Flanker task show that even small amounts of alcohol can affect the performance of the participants in terms of selective attention and inhibitory function.

The last objective, **O2**, was to evaluate the classification of alcohol and non-alcohol EEG signals by using deep learning techniques. As discussed in Section 5.5, the individual model across sessions was able to identify which data were affected by alcohol and which data were not affected, resulting in an average accuracy of 90.7%. For the general model, discussed in Section 5.6, the average accuracy was 62.9%, meaning it was not able to differentiate the signals as well as the individual model. Both accuracies are still higher than the one achieved by guessing, which is 50%.

To conclude, it is possible to differentiate alcohol-affected EEG signals from those that are not affected. The Flanker results indicated that the participants were affected by the alcohol, which again implies that the EEG signals should be affected as well. These results are supported by the performance of the classifiers, especially the individual ones. The high accuracies indicate that EEGNet can extract features which characterise alcohol-affected signals. The performance of the general model is not as good, and it struggles to correctly classify alcohol-affected signals. There could be numerous reasons for this, and improving the performance should be explored further. Still, the models presented in this thesis could be the first step towards creating an EEG-based alcohol detector utilised in a BCI system.

6.2 Future Work

In this thesis, the foundation for a binary alcohol detector has been presented. Although some results are promising, further work is encouraged to make the system feasible and to develop a prototype for real-life application.

Performing channel optimisation could improve the performance of the classifier by removing noisy or redundant channels. The channels used to obtain the results in this thesis are optimal for alcoholism detection. Therefore, finding channels optimal for alcohol detection could result in better performance. This can, for instance, be done by using Deep Learning Important Features (DeepLIFT) [61], which can identify the contribution of each channel to the output of the classifier. It can also be done by using optimisation algorithms comparable to the one described in [52].

Another way to improve the performance of the model is by using high-density EEG. The channels used in the experiment are placed across the entire surface of the participant's head. Since there are only 16 channels, there are large distances between several of the channels. This can potentially result in relevant information loss. By increasing the number of electrode placements, more EEG information can be recorded. This means more data can be used as input in the model, which could help better the model performance.

Since the accuracy of the individual model increased as the epoch length decreased would it be of interest to explore the performance of the classifier on even shorter epoch lengths. This may result in improved accuracy. Removing power line interference and other artifacts can also improve the performance, as it could improve the signal-to-noise ratio. This can make it easier for the classifier to identify neural activity caused by alcohol intoxication.

Expanding the classifier from a binary classifier to a multi-class classifier is of great interest as it would provide an estimate of how intoxicated a person is. This could take the alcohol detector a step closer to real-world use, where it can identify if a person is above the legal BAC limit in their country.

No quantitative feature analysis has been performed in the project since no known features have been extracted from the data. Such features can be found by utilising feature-based

machine learning techniques. When the relevant features have been identified, qualitative analysis such as using topographic maps can be performed to analyse which features are the strongest in which frequency bands.

In addition to the adjustments suggested above, the performance of the alcohol detector can be improved by improving the protocol. This can, for instance, be done by extending the experiment duration to include the peak BAC. This extension would be in agreement with the duration of the experiments described in Section 3.1.1 and in [22]. Another way to improve the protocol could be to add a third recording session, in which the participants are not served any drink, to establish a baseline. A third way to improve the protocol could be to introduce a habituation Flanker task for all participants, in all sessions. This would further ensure the removal of the element of nervousness, which increases the chance of the Flanker task results being unaffected by this.

To summarise, the proposed future work for the alcohol detector project is the following:

FW1 Perform channel optimisation to improve classifier performance

FW2 Use high-density EEG to increase the amount of recorded information

FW3 Explore the epoch length further and improve the data set

FW4 Expand the classifier from binary to multi-class

FW5 Perform feature analysis by utilising machine learning, and perform qualitative and quantitative analysis on the extracted features

FW6 Improve the protocol by extending experiment length, introducing a baseline session without a drink, or by adding one more Flanker task in each session as a habituation task

Bibliography

[1] M. Vassbotn, "Design of protocol and collection of data for an EEG based alcohol detector." doi:10.13140/RG.2.2.15013.37600, 12 2022.

[2] I. J. Nordstrøm-Hauge, "Design of protocol and collection of data for an EEG based alcohol detector." doi:10.13140/RG.2.2.36378.11205, 12 2022.

[3] National Institute on Alcohol Abuse and Alcoholism US, "Alcohol use in the united states: Age groups and demographic characteristics." Available at https://www.niaaa.nih.gov/alcohols-effects-health/alcohol-topics/alcohol-facts-and-statistics/alcohol-use-united-states-age-groups-and-demographic-characteristics, 2023. Accessed 16.05.23.

[4] E. Bye, "Alkoholbruk i den voksne befolkningen," *Norwegian Institute of public Health, Webpublication*, vol. 9, 2018.

[5] W. H. Organization, *Global status report on alcohol and health 2018*. World Health Organization, 2019.

[6] National Institute on Alcohol Abuse and Alcoholism US, "Alcohol metabolism," *Alcohol Alert*, vol. 63, 2004.

[7] L. Farsi, S. Siuly, E. Kabir, and H. Wang, "Classification of alcoholic EEG signals using a deep learning method," *IEEE Sensors Journal*, vol. 21, no. 3, pp. 3552–3560, 2020.

[8] H. Mukhtar, S. M. Qaisar, and A. Zaguia, "Deep convolutional neural network regularization for alcoholism detection using EEG signals," *Sensors*, vol. 21, no. 16, p. 5456, 2021.

[9] V. Singhal, J. Mathew, R. K. Behera, *et al.*, "Detection of alcoholism using EEG signals and a CNN-LSTM-ATTN network," *Computers in biology and medicine*, vol. 138, p. 104940, 2021.

[10] L. Vissers, S. Houwing, and F. Wegman, "Alcohol-related road casualties in official crash statistics," 2018.

[11] World Health Organization, "BAC limit data by country." Available at https://apps.who.int/gho/data/node.main.A1002?lang=en, 2020. Accessed 28.04.23.

[12] Politiet, "Rus og ruskontroll."

[13] I. S. Moan, "Promillekjøring i veitrafikken i norge," *Alkohol i Norge*, 2018.

[14] World Health Organization, "Legal blood alcohol concentration (BAC) limits." Available at https://www.who.int/data/gho/data/indicators/indicator-details/GHO/legal-blood-alcohol-concentration-(bac)-limits. Accessed 20.05.23.

[15] National Highway Traffic Safety Administration, "Drunk driving." Available at https://www.nhtsa.gov/risky-driving/drunk-driving. Accessed 20.05.23.

[16] International transport forum, "Road Safety Annural Report 2020: Canada." Available at https://www.itf-oecd.org/sites/default/files/canada-road-safety.pdf, 2020. Accessed 20.05.23.

[17] Testhelsen, *Alkometer Alcoscan Alc-1*. Sentech Korea Corp, 2022.

[18] J. M. Celaya-Padilla, J. S. Romero-González, C. E. Galvan-Tejada, J. I. Galvan-Tejada, H. Luna-García, J. G. Arceo-Olague, N. K. Gamboa-Rosales, C. Sifuentes-Gallardo, A. Martinez-Torteya, J. I. De la Rosa, *et al.*, "In-vehicle alcohol detection using low-cost sensors and genetic algorithms to aid in the drinking and driving detection," *Sensors*, vol. 21, no. 22, p. 7752, 2021.

[19] K. Murata, E. Fujita, S. Kojima, S. Maeda, Y. Ogura, T. Kamei, T. Tsuji, S. Kaneko, M. Yoshizumi, and N. Suzuki, "Noninvasive biological sensor system for detection of drunk driving," *IEEE transactions on information technology in biomedicine*, vol. 15, no. 1, pp. 19–25, 2010.

[20] V. Vijayan and E. Sherly, "Real time detection system of driver drowsiness based on representation learning using deep neural networks," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 3, pp. 1977–1985, 2019.

[21] Z. Ek, A. Akg, and M. R. Bozkurt, "The classificaton of EEG signals recorded in drunk and non-drunk people," *International Journal of Computer Applications*, vol. 68, no. 10, 2013.

[22] A. Jones, "Biochemical and physiological research on the disposition and fate of ethanol in the body," *Medicolegal aspects of alcohol, 5th edition, Lawyers and Judges publishing company, Tucson*, pp. 47–128, 2008.

[23] S. Zakhari, "Overview: how is alcohol metabolized by the body?," *Alcohol research & health*, vol. 29, no. 4, p. 245, 2006.

[24] S. R. Sinha, L. R. Sullivan, D. Sabau, D. S. J. Orta, K. E. Dombrowski, J. J. Halford, A. J. Hani, F. W. Drislane, and M. M. Stecker, "American clinical neurophysiology society guideline 1: minimum technical requirements for performing clinical electroencephalography," *The Neurodiagnostic Journal*, vol. 56, no. 4, pp. 235–244, 2016.

[25] R. B. Firat, "Opening the "black box": functions of the frontal lobes and their implications for sociology," *Frontiers in Sociology*, vol. 4, p. 3, 2019.

[26] Blausen Medical, "Medical gallery of blausen medical 2014," *WikiJournal of Medicine*, vol. 1, no. 2, pp. 1–79, 2014.

[27] C. Watson, M. Kirkcaldie, and G. Paxinos, "Chapter 1 - nerve cells and synapses," in *The Brain* (C. Watson, M. Kirkcaldie, and G. Paxinos, eds.), pp. 1–10, San Diego: Academic Press, 2010. doi:10.1016/B978-0-12-373889-9.50001-2.

[28] S. Sanei and J. A. Chambers, *EEG signal processing*, pp. 1–30. John Wiley & Sons, 2013.

[29] V. Jurcak, D. Tsuzuki, and I. Dan, "10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems," *Neuroimage*, vol. 34, no. 4, pp. 1600–1611, 2007.

[30] L. R. Krol, "File: EEG 10-10 system with additional information.svg." Available at https://commons.wikimedia.org/wiki/File:EEG_10-10_system_with_additional_information.svg. Accessed 15.12.22.

[31] S. Sur and V. K. Sinha, "Event-related potential: An overview," *Industrial psychiatry journal*, vol. 18, no. 1, p. 70, 2009.

[32] T. C. Handy, *Event-related potentials: A methods handbook*. MIT press, 2005.

[33] Y. Xie and S. Oniga, "A review of processing methods and classification algorithm for EEG signal," *Carpathian Journal of Electronic and Computer Engineering*, vol. 13, no. 1, pp. 23–29, 2020.

[34] M. R. Lakshmi, T. Prasad, and D. V. C. Prakash, "Survey on EEG signal processing methods," *International journal of advanced research in computer science and software engineering*, vol. 4, no. 1, 2014.

[35] H. Morioka, A. Kanemura, J.-i. Hirayama, M. Shikauchi, T. Ogawa, S. Ikeda, M. Kawanabe, and S. Ishii, "Learning a common dictionary for subject-transfer decoding with resting calibration," *NeuroImage*, vol. 111, pp. 167–178, 2015.

[36] C. L. Ehlers, T. L. Wall, and M. A. Schuckit, "EEG spectral characteristics following ethanol administration in young men," *Electroencephalography and clinical neurophysiology*, vol. 73, no. 3, pp. 179–187, 1989.

[37] D. J. McFarland and J. R. Wolpaw, "Brain-computer interfaces for communication and control," *Communications of the ACM*, vol. 54, no. 5, pp. 60–66, 2011.

[38] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday, "Brain–computer interfaces for communication and rehabilitation," *Nature Reviews Neurology*, vol. 12, no. 9, pp. 513–525, 2016.

[39] Y.-C. Jung and K. Namkoong, "Alcohol: intoxication and poisoning–diagnosis and treatment," *Handbook of clinical neurology*, vol. 125, pp. 115–121, 2014.

[40] M. Frezza, C. di Padova, G. Pozzato, M. Terpin, E. Baraona, and C. S. Lieber, "High blood alcohol levels in women: the role of decreased gastric alcohol dehydrogenase activity and first-pass metabolism," *New England Journal of Medicine*, vol. 322, no. 2, pp. 95–99, 1990.

[41] M. S. Mumenthaler, J. L. Taylor, R. O'Hara, and J. A. Yesavage, "Gender differences in moderate drinking effects," *Alcohol Research & Health*, vol. 23, no. 1, p. 55, 1999.

[42] D. M. Lovinger and M. Roberto, "Synaptic effects induced by alcohol," *Behavioral neurobiology of alcohol addiction*, pp. 31–86, 2010.

[43] S. Mukherjee, "Alcoholism and its effects on the central nervous system," *Current neurovascular research*, vol. 10, no. 3, pp. 256–262, 2013.

[44] K. Abernathy, L. J. Chandler, and J. J. Woodward, "Alcohol and the prefrontal cortex," *International review of neurobiology*, vol. 91, pp. 289–320, 2010.

[45] C. M. Steele and R. A. Josephs, "Alcohol myopia: Its prized and dangerous effects.," *American psychologist*, vol. 45, no. 8, p. 921, 1990.

[46] B. D. Bartholow, M. Pearson, K. J. Sher, L. C. Wieman, M. Fabiani, and G. Gratton, "Effects of alcohol consumption and alcohol susceptibility on cognition: a psychophysiological examination," *Biological psychology*, vol. 64, no. 1-2, pp. 167–190, 2003.

[47] E. National Academies of Sciences, Medicine, *et al.*, *Reproducibility and replicability in science*. National Academies Press, 2019.

[48] P. Patil, R. D. Peng, and J. T. Leek, "A statistical definition for reproducibility and replicability," *BioRxiv*, p. 066803, 2016.

[49] K. S. Button, J. Ioannidis, C. Mokrysz, B. A. Nosek, J. Flint, E. S. Robinson, and M. R. Munafò, "Power failure: why small sample size undermines the reliability of neuroscience," *Nature reviews neuroscience*, vol. 14, no. 5, pp. 365–376, 2013.

[50] Y. G. Pavlov, N. Adamian, S. Appelhoff, M. Arvaneh, C. S. Benwell, C. Beste, A. R. Bland, D. E. Bradford, F. Bublatzky, N. A. Busch, *et al.*, "#EEGManyLabs: Investigating the replicability of influential EEG experiments," *cortex*, vol. 144, pp. 213–229, 2021.

[51] P. E. Clayson, K. A. Carbine, S. A. Baldwin, and M. J. Larson, "Methodological reporting behavior, sample sizes, and statistical power in studies of event-related potentials: Barriers to reproducibility and replicability," *Psychophysiology*, vol. 56, no. 11, p. e13437, 2019.

[52] S. Bavkar, B. Iyer, and S. Deosarkar, "Optimal EEG channels selection for alcoholism screening using EMD domain statistical features and harmony search algorithm," *Biocybernetics and Biomedical Engineering*, vol. 41, no. 1, pp. 83–96, 2021.

[53] L. Hu and Z. Zhang, *EEG signal processing and feature extraction*. Springer, 2019.

[54] G. Stenberg, M. Sano, I. Rosén, and D. H. Ingvar, "EEG topography of acute ethanol effects in resting and activated normals.," *Journal of studies on alcohol*, vol. 55, no. 6, pp. 645–656, 1994.

[55] H. L. Cohen, B. Porjesz, and H. Begleiter, "Ethanol-induced alterations in electroencephalographic activity in adult males," *Neuropsychopharmacology*, vol. 8, no. 4, pp. 365–370, 1993.

[56] M. Rangaswamy and B. Porjesz, "Understanding alcohol use disorders with neuroelectrophysiology," *Handbook of clinical neurology*, vol. 125, pp. 383–414, 2014.

[57] B. A. Eriksen and C. W. Eriksen, "Effects of noise letters upon the identification of a target letter in a nonsearch task," *Perception & psychophysics*, vol. 16, no. 1, pp. 143–149, 1974.

[58] J. Sola and J. Sevilla, "Importance of input data normalization for the application of neural networks to complex industrial problems," *IEEE Transactions on nuclear science*, vol. 44, no. 3, pp. 1464–1468, 1997.

[59] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," *Journal of neural engineering*, vol. 15, no. 5, p. 056013, 2018.

[60] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[61] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning important features through propagating activation differences," in *International conference on machine learning*, pp. 3145–3153, PMLR, 2017.

Appendices

A Participant Selection Questionnaire

Questionnaire for Participants of the Alcohol Detector Project

Name:
Age:
Weight in kg:
Height in cm:
Average of weekly units of alcohol consumed:
Approximate head circumference in cm:

Screening Questions

Please answer the questions below by marking the box that is correct for you. A first-degree relative is someone who shares 50% of your genes, and a second-degree relative is someone who shares 25% of your genes.

Alcohol use disorder (AUD) is a medical condition characterised by an impaired ability to stop or control alcohol use despite adverse social, occupational, or health consequences. If you want to read more about AUD please click the following link: https://www.niaaa.nih.gov/publications/brochures-and-fact-sheets/understanding-alcohol-use-disorder.

Do you have a history of alcohol use disorder in your first or second-degree relatives?
- ☐ Yes
- ☐ No

Do you have a history of drug abuse in your first or second-degree relatives?
- ☐ Yes
- ☐ No

Do you have major medical problems that affect the nervous system?
- ☐ Yes
- ☐ No

Do you have a history of alcohol use disorder?
- ☐ Yes
- ☐ No

Do you have a history of drug abuse?

☐ Yes

☐ No

Do you have a history of psychiatric problems?

☐ Yes

☐ No

Are you an abstainer from ethanol use?

☐ Yes

☐ No

Are you allergic to alcohol?

☐ Yes

☐ No

Do you have sensitive skin? This gel will come in contact with your scalp.

☐ Yes

☐ No

Do you have a citrus allergy?

☐ Yes

☐ No

Edinburgh Handedness Inventory

The purpose of the Edinburgh Handedness Inventory is to objectively ascertain the handedness of a participant. It is the most commonly used screening tool for handedness. Below are 20 activities. Please specify the side you prefer to perform the given activity.

If you prefer either the left or the right side for the given activities, then mark one box on the column of that side. If the preference for a particular side is so strong that you would not use the other hand unless forced, then mark both the boxes on that side. If there is no preference for any side, then mark one box on both sides of the activity.

For items that involve a bimanual task such as striking a match, the hand involved in the usage of the key item (i.e. the match) is considered the preferred side. If you have no experience with a given task, leave that item unmarked.

☐ ☐ Writing ☐ ☐

☐ ☐ Drawing ☐ ☐

☐ ☐ Throwing ☐ ☐

☐ ☐ Scissors ☐ ☐

☐ ☐ Comb ☐ ☐

Toothbrush

Knife (without fork)

Spoon

Hammer

Screwdriver

Tennis racket

Knife with fork

Cricket bat

Golf club

Broom

Rake

Striking a match

Opening a box (lid)

Dealing cards

Threading a needle

B Consent Form

Informed Consent Form for Research Involving Human Subjects

You are being invited to participate in a research study, which the Norwegian Center for Research Data (NSD) has reviewed and approved for conduction by the investigators named here. This form is designed to provide you - as a human subject - with information about this study. The investigator or his/her representative will describe this study to you and answer any of your questions. You are entitled to a copy of this form. If you have any questions or complaints about the informed consent process of this research study or your rights as a subject, please contact the PI or Co-PI (marta.molinas@ntnu.no, +47 94287670, andres.f.soler.guevara@ntnu.no).

Project Title: FlexEEG based Alcohol Detector

Principal Investigators: Marta Molinas

Co-investigator: Andres Soler

Thank you for agreeing to participate in this research project. This study involves research aimed at detecting the presence of alcohol in the human body based on the analysis of EEG signals. You will participate in two separate data collection sessions. Before the first session, you will be invited to a 5-minute control measuring session where your height, weight and head circumference will be measured. The two sessions are separated by at least one day. During one of the sessions, you will be presented with a non-alcoholic drink consisting of orange juice and a cocktail-mix taste additive. During the other session you will be presented with 0.45g/kg ethanol mixed with orange juice. You will not be given information about which session involves alcohol, as this may affect the results. Each session will last about 81 minutes. 66 of these minutes are for collection of EEG signals from your brain using the Unicorn Hybrid Black system (www.unicorn-bi.com). Before and after ingestion of the drink you will be asked to perform a cognitive task. You will also be asked to use a breathalyzer during the session to keep track of your blood alcohol level. Before starting collection of EEG signals, 15 minutes will be used to place the EEG headset on your head and give you an introduction to the cognitive task.

Participation in this study will take approximately 160 minutes of your time. We do not anticipate you to experience negative feelings when responding to items in this study. Your participation in this study is completely voluntary. Should you decide to discontinue participation or decline to answer any specific part of the study, you may do so without penalty.

Your participation in this study may help you understand the manifestations of alcohol ingestion on EEG signals. We are not asking you to place your name anywhere on the experimental booklet, so your participation is anonymous. None of your answers can be directly traced back to you. Should you have any further questions, please feel free to contact the study's principal investigator or co-PI, Marta Molinas and Andres Soler at the Department of Engineering Cybernetics. Her office is located in Elektro D+B2 room D244, her phone number is +47 94287670, and her e-mail address is marta.molinas@ntnu.no.

CONSENT STATEMENT: I, _____, hereby give my consent to participate in the research study entitled "FlexEEG based Alcohol Detector." I have read the above information and am aware of the potential risks and complications. I fully understand that I may withdraw from this research project at any time without prejudice or effect on my standing with NTNU. I also understand that I am free to ask questions about techniques or procedures that

will be undertaken. I will sign and return this consent form and receive a copy of the form in case I need to refer back to it. Finally, I understand that information obtained about me during the course of the study will be kept anonymous and cannot be traced back to me.

_____ _____

Participant's signature (20+ years of age) Date

I hereby certify that I have given an explanation to the above individual of the contemplated study and its risks and potential complications.

_____Marta Molinas_____ ___27.10.2022

Principal Investigator's signature Date

C Information Letter

Information letter for Alcohol detector participation

Thank you for agreeing to participate in the Alcohol Detector Project.

Before your participation, you must do the following:

- Do not drink alcohol the night before or the same day as the session.
- Do not ingest any caffeine before your session
- Remember to eat before your session

Not following these guidelines will affect the results.

You do not need to show up with newly washed hair. We recommend that you bring a water bottle as you will be allowed to drink some water after the drink ingestion if you want to. Since each session lasts approximately 1 hour, we recommend visiting the toilet beforehand.

Thank you for your participation!
Molly & Iselin

D Length of the Recorded EEG Data

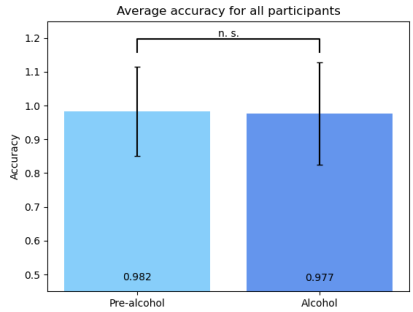
Table D.1: An overview of how much EEG data was recorded in each run for all participants. The planned amount of recorded data for each run was 5 minutes. n1-n5 are the non-alcohol-affected recordings. a1-a3 are the alcohol-affected recordings.

Participant	n1	n2	n3	n4	n5	a1	a2	a3
P01	5 min	5 min	5 min	5 min	*	5 min	5 min	5 min
P02	5 min	*	5 min	5 min	5 min	5 min	5 min	5 min
P03	5 min	5 min	5 min	5 min	5 min	5 min	5 min	5 min
P04	5 min	5 min	5 min	5 min	5 min	5 min	5 min	5 min
P05	5 min	5 min	72 sec	5 min	5 min	5 min	5 min	5 min
P06	5 min	80 sec	34 sec	5 min	5 min	5 min	5 min	5 min
P07	5 min	5 min	5 min	5 min	5 min	5 min	5 min	5 min
P08	5 min	5 min	5 min	5 min	5 min	5 min	5 min	5 min
P09	5 min	5 min	5 min	5 min	5 min	5 min	5 min	5 min
P10	5 min	5 min	5 min	5 min	5 min	5 min	114 sec	5 min
P11	5 min	40 sec	5 min	5 min	5 min	5 min	5 min	5 min
P12	5 min	5 min	5 min	5 min	5 min	5 min	5 min	5 min
P13	5 min	5 min	5 min	5 min	5 min	5 min	5 min	5 min
P14	5 min	5 min	5 min	5 min	5 min	5 min	5 min	5 min
P15	5 min	5 min	5 min	100 sec	5 min	5 min	5 min	5 min
P16	5 min	5 min	5 min	5 min	5 min	5 min	5 min	5 min
P17	5 min	5 min	5 min	5 min	5 min	60 sec	5 min	5 min
P18	5 min	5 min	5 min	5 min	5 min	5 min	5 min	5 min
P19	5 min	5 min	5 min	5 min	5 min	5 min	5 min	5 min
P20	5 min	5 min	5 min	5 min	5 min	5 min	5 min	5 min

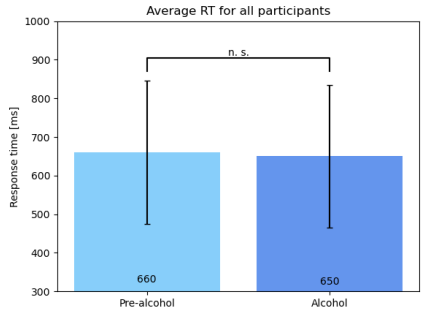
* No measurement due to technical issue with the EEG equipment

E Additional Flanker Results

Flanker Figures with Pre-Experiment Results

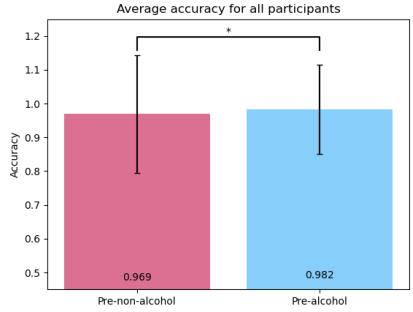


(a) Average accuracy before and after ingestion of the alcoholic drink. Significance: n.s. p=0.2112.

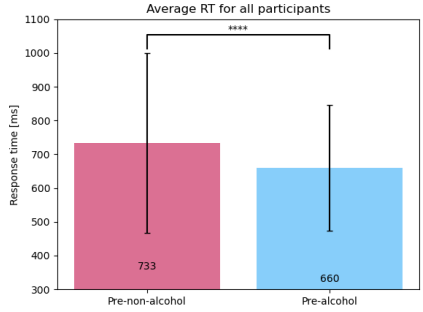


(b) Average response time before and after ingestion of the alcoholic drink. Significance: n.s. p=0.1077.

Figure E.1: Comparison of average Flanker task values before and after ingestion of the alcoholic drink.

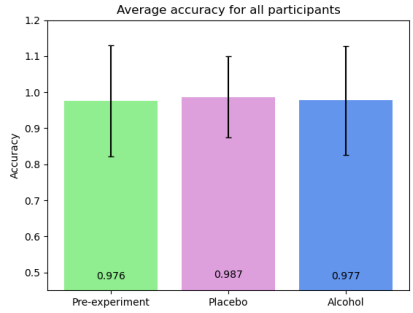


(a) Average accuracy before ingestion of the non-alcoholic drink and before the alcoholic drink. Significance: * p=0.066 $<$ 0.05.

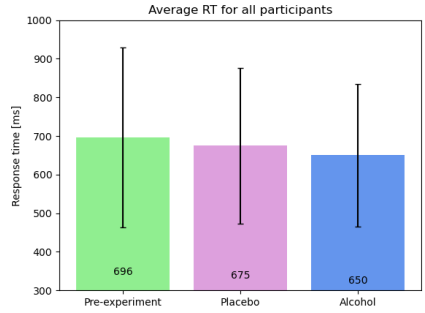


(b) Average response time before ingestion of the non-alcoholic drink and before the alcoholic drink. Significance: **** p=$1.65 \cdot 10^{-22}$ $<$ 0.00005.

Figure E.2: Comparison of average Flanker task values before ingestion of the non-alcoholic drink and before the alcoholic drink.



(a) Average Flanker task accuracy with SD for all participants.



(b) Average Flanker task RTs with SD for all participants.

Figure E.3: Additional group level Flanker results with the pre-experiment Flanker tasks added. The pre-experiment values are derived from the pre-alcoholic and the pre-non-alcoholic Flanker tasks.

Individual Flanker Results

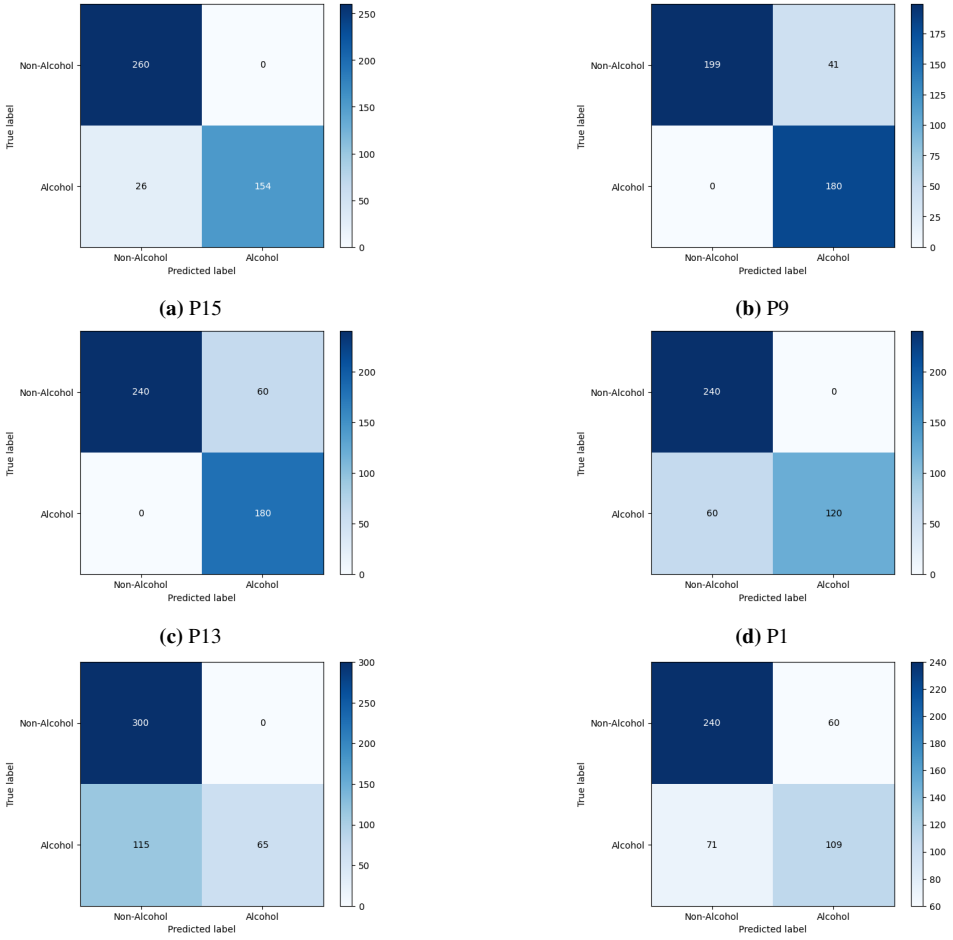
Table E.1: Average individual accuracies with SD from the non-alcoholic and alcoholic Flanker tasks.

	Accuracy (σ)	
	Non-alcohol	**Alcohol**
P01	0.990 (0.102)	1.000 (0.000)
P02	0.990 (0.102)	1.000 (0.000)
P03	0.990 (0.102)	0.979 (0.143)
P04	0.990 (0.102)	0.979 (0.143)
P05	0.990 (0.102)	0.979 (0.143)
P06	0.990 (0.102)	1.000 (0.000)
P07	0.969 (0.174)	0.979 (0.143)
P08	0.979 (0.143)	0.945 (0.222)
P09	0.990 (0.102)	0.979 (0.143)
P10	0.945 (0.222)	0.927 (0.260)
P11	1.000 (0.000)	0.969 (0.174)
P12	0.967 (0.174)	0.956 (0.200)
P13	0.990 (0.102)	1.000 (0.000)
P14	0.990 (0.102)	0.948 (0.222)
P15	0.990 (0.102)	0.990 (0.102)
P16	0.990 (0.102)	0.969 (0.174)
P17	1.000 (0.000)	0.968 (0.174)
P18	1.000 (0.000)	0.979 (0.143)
P19	0.990 (0.102)	1.000 (0.000)
P20	1.000 (0.000)	0.979 (0.143)

Table E.2: Average individual RTs with SD from the non-alcoholic and alcoholic Flanker tasks.

	RT [ms] (σ)	
	Non-alcohol	**Alcohol**
P01	559 (96)	614 (111)
P02	882 (317)	735 (229)
P03	623 (131)	596 (78)
P04	594 (132)	655 (169)
P05	722 (204)	819 (268)
P06	780 (142)	682 (102)
P07	632 (117)	685 (149)
P08	559 (87)	526 (82)
P09	688 (157)	627 (89)
P10	874 (254)	863 (193)
P11	655 (159)	660 (159)
P12	800 (197)	691 (163)
P13	796 (190)	742 (171)
P14	820 (280)	848 (230)
P15	585 (105)	609 (150)
P16	666 (139)	585 (82)
P17	638 (105)	563 (88)
P18	587 (111)	544 (72)
P19	521 (83)	504 (67)
P20	516 (72)	461 (72)

F Confusion Matrices for the General Model



(a) P15

(b) P9

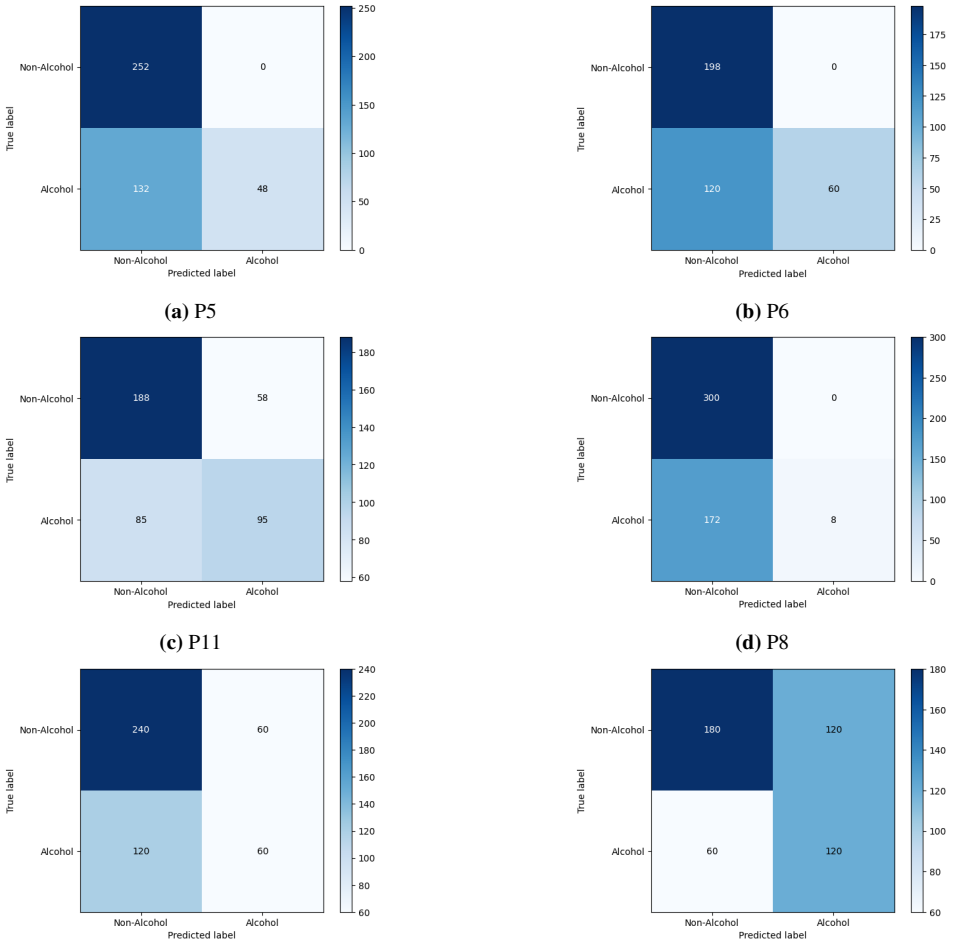
(c) P13

(d) P1

(e) P16

(f) P20

Figure F.1: Confusion matrices from the general model, for the best-performing participants based on accuracy.



(a) P5 **(b)** P6

(c) P11 **(d)** P8

(e) P7 **(f)** P3

Figure F.2: Confusion matrices from the general model for participants with an accuracy close to the model's average accuracy.

G Paper for IV LAWCN

EEG-Based Alcohol Detection System for Driver Monitoring

Molly Vassbotn[1,2*], Iselin J. Nordstrøm-Hauge[1,3*], Andres Soler[1,4], and Marta Molinas[1,5]

[1] Norwegian University of Science and Technology, 7491 Trondheim, Norway
[2] molly.vassbotn@hotmail.com
[3] iselinjnh@live.no
[4] andres.f.soler.guevara@ntnu.no
[5] marta.molinas@ntnu.no

Abstract. Today, alcohol drinking frequently accompanies socialising as a routine activity in various groups of society. 84.0% of individuals aged 18 and above in the United States have drunk alcohol at some point in their life [8]. Similarly, 81.7% of Norwegians in the age group 16 to 79 have drunk alcohol in 2021 [1]. Driving after the consumption of alcohol is a worldwide problem, causing a large number of deaths and injuries a year. This work proposes the first steps towards developing an electroencephalography (EEG)-based alcohol detector conceived with the idea to prevent people from driving under the influence of alcohol. This includes the design of an experimental protocol for EEG data collection, during which participants performed the Flanker task, and their blood alcohol concentration (BAC) was measured. The resulting data set consists of two sessions per participant, both while they are affected and not-affected by alcohol. Statistical analysis of the Flanker task indicated that the participants were affected by alcohol, and therefore, their EEG signals were expected to be affected as well. The collected EEG signals were used as input for three classifiers, all based on the EEGNet architecture. Two of these classifiers are individual models, with average accuracies of 100% and 90.7%. The last classifier is a general model with an accuracy of 62.9%. Although the performance of the presented classifiers should be improved for them to be used as a real-world alcohol detector, these results are encouraging.

Keywords: EEG · Alcohol · CNN · EEGNet · Classification · BAC · Flanker

1 Introduction

Electroencephalography (EEG) is a technique used for capturing brain signals by placing electrodes on the scalp. Today, EEG is a commonly used technique

* Equal contribution

for studying the brain. Many studies [4,7,11] have been investigating if it is possible to diagnose alcoholism using EEG data. Several of these studies have been successful, resulting in a high classification accuracy. However, only one study [2] has focused on using EEG data for the detection of alcohol in a healthy body.

Driving under the influence of alcohol is a worldwide problem. It is estimated to cause the death of at least 273 000 road users every year, although the actual number is believed to be higher [13]. The legal blood alcohol concentration (BAC) for driving varies for different countries, but, for most, the BAC limit is within the range of 0.2‰ to 0.8‰ [14].

In order to decrease the number of injuries and deaths, it is important to stop drunk drivers before the accident happens. Today, using breathalysers is the common way of detecting if a person is under the influence of alcohol. The breathalyser estimates the BAC value by using a single breath sample. It is the tool used by the police when they are suspecting that a person is driving under the influence of alcohol.

Although using breathalysers is a quick and inexpensive way of measuring the BAC, it has some disadvantages. Using breath samples is an indirect way of measuring the amount of alcohol in the blood, and incorrect measurements can occur due to factors such as residual alcohol or juice in the mouth, or temperature and humidity [10].

Creating an alcohol detector by using EEG signals could be of great interest. Using EEG signals instead of breath samples could potentially result in a system with higher precision and better accuracy than the breathalyser. Therefore, this paper proposes the first step towards the design of an alcohol detector system for onboard detection of the presence of alcohol in drivers.

2 Methods

2.1 Protocol for data collection

In order to collect alcohol- and non-alcohol affected EEG data, a protocol was designed. The protocol was designed for the collection of EEG data from 20 healthy participants, 10 males and 10 females. The participants were chosen based on a set of selection and exclusion criteria, including age (20-30 years old), health and no prior medical issues concerning alcohol or drugs. Each participant partook in two sessions, one where an alcoholic drink was served and another where a non-alcoholic drink was served. The alcoholic drink was vodka-based and consisted of 0.45 g/kg ethanol mixed with orange juice. In the non-alcoholic drink, the alcohol was substituted with a vodka-flavoured mixer.

Each session lasted approximately 66 minutes. During each session, EEG data were measured while the participant was resting, with their eyes open. The Flanker task [3] was performed two times during a session to investigate how alcohol affects the accuracy and the response time (RT) of the participants. The BAC of the participant was measured before and after the ingestion of alcohol by using a breathalyser. Figure 1 shows the data collection session in detail. A detailed presentation of the designed protocol and experiments are available in [9] and [12].

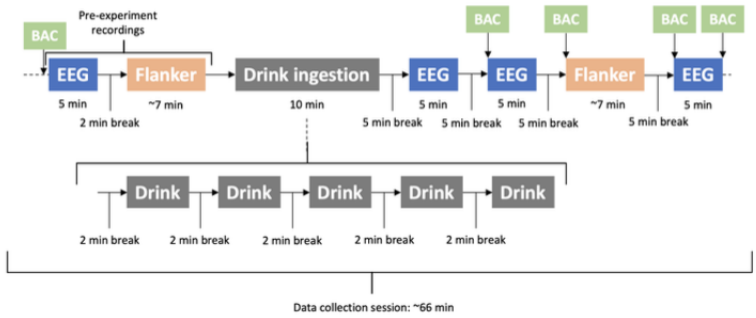


Fig. 1: A detailed overview of the data collection session [10].

2.2 Data preprocessing

Before the collected data set was used as input to the classifiers, some preprocessing was applied. First, the 5-minute EEG recordings were split into epochs of 5 seconds. After this, the data was split into a training and a test set. The data set was not subjected to any artifact removal or filtering procedures; only normalisation was performed.

2.3 Implemented models

Three classifiers were implemented to detect alcohol-affected EEG signals. All were implemented using the CNN architecture EEGNet [6]. This was chosen as it is made specifically for the classification and interpretation of EEG signals. EEGNet is known for performing well on different types of EEG signals, even when the available data set is very limited [6]. All models were optimised by using the Adam algorithm with a learning rate of 0.001. The used loss function was *binary cross entropy*. The hyperparameters of EEGNet were chosen to be their default values; $F_1 = 8$, $D = 2$, $F_2 = 16$. Two of the implemented classifiers were based on individual models while the third classifier was a general model. All are presented in the following sections.

Random individual model In the random individual model, all the 5-second epochs were split randomly into a training and a test set. Then, 5-fold cross-validation was used to train and validate the model to find the best hyperparameters. After this, the test set was used to get a final evaluation of the performance of the random individual model.

Individual model across sessions In this model, all 5-second epochs from the same 5-minute recordings were placed in the same set, instead of splitting epochs randomly into the training and test set. One 5-minute recording from the non-alcohol session and one 5-minute recording from the alcohol session were chosen randomly to be in the test set of each participant. The remaining data were used in 3-fold cross-validation. The test set was used for the final evaluation of the model.

The general model For the general model, EEGNet was trained using data from 19 of the 20 participants and then tested on the last, unseen participant. The hyperparameters of the model were chosen by using 3-fold cross-validation.

2.4 Metrics

The metrics used for evaluating the classifiers are accuracy, precision, recall, F1 score and specificity. These are further described in [10].

3 Results

3.1 BAC Evolution

The measured BAC values after ingestion of alcohol for each participant are presented in Table 1. Before the recordings, all participants had a BAC value of 0.000‰. In the non-alcoholic recording session, all participants had a BAC value of 0.000‰ throughout the session. Figure 2 shows the average BAC values for males, females and all participants at each BAC measuring point during the alcoholic recording session.

3.2 Behavioural Data From the Flanker Task

Figure 3 presents the average accuracy results from the Flanker task performed after drink ingestion. As seen in Figure 3a, the average accuracy for all participants decreases with a value of 0.01 from 0.987 to 0.977 when alcohol is present. The difference in accuracy is significant with a p-value of $p = 0.015836$. Figure 3b shows the average RT for all participants. After ingestion of alcohol, the RT decreased by 25 ms., from 675 to 650 ms. The difference in RT is significant with a p-value of $p = 0.000091$.

Table 1: The BACs of all participants at approximately 15, 25, 37, and 42 minutes after (m.a.) alcohol ingestion [10].

Participant	Gender	15 m.a.	25 m.a.	37 m.a.	42 m.a.
P01	Female	0.450‰	0.440‰	0.430‰	0.450‰
P02	Female	0.270‰	0.330‰	0.380‰	0.400‰
P03	Female	0.420‰	0.420‰	0.500‰	0.540‰
P04	Female	0.440‰	0.430‰	0.490‰	0.450‰
P05	Female	0.440‰	0.450‰	0.470‰	0.490‰
P06	Male	0.310‰	0.390‰	0.400‰	0.430‰
P07	Female	0.110‰	0.140‰	0.190‰	0.210‰
P08	Male	0.250‰	0.260‰	0.290‰	0.320‰
P09	Male	0.420‰	0.430‰	0.440‰	0.450‰
P10	Female	0.340‰	0.260‰	0.260‰	0.360‰
P11	Male	0.550‰	0.530‰	0.540‰	0.500‰
P12	Male	0.390‰	0.370‰	0.460‰	0.480‰
P13	Female	0.520‰	0.470‰	0.490‰	**
P14	Male	0.430‰	0.470‰	0.480‰	0.480‰
P15	Female	0.530‰	0.480‰	0.570‰	0.560‰
P16	Female	0.480‰	0.460‰	0.580‰	0.540‰
P17	Male	0.250‰	0.280‰	0.290‰	0.380‰
P18	Male	0.480‰	0.380‰	0.470‰	0.520‰
P19	Male	0.320‰	0.310‰	0.340‰	0.400‰
P20	Male	0.630‰	0.470‰	0.480‰	0.490‰

** No measurement due to technical issue with the breathalyser

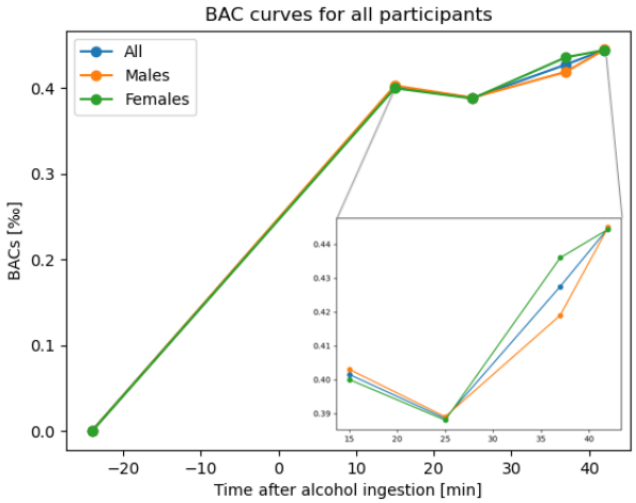
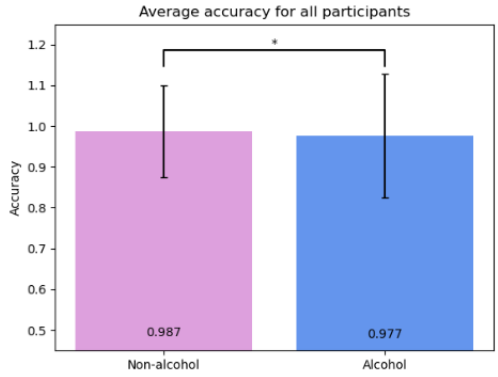
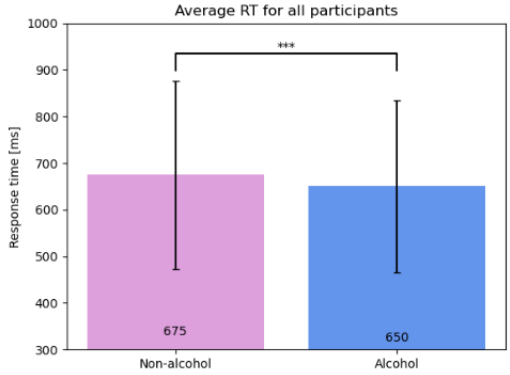


Fig. 2: Average BAC values for males, females and all participants during the alcoholic recording session. The inserted window shows an enlarged version of the most relevant area [10].



(a) Average accuracy. Significance:
* $p = 0.015836 < 0.05$.



(b) Average RT. Significance:
*** $p = 0.000091 < 0.0005$.

Fig. 3: Average accuracies and RTs for the Flanker task [10].

3.3 Individual models

Random individual model The performance of the random individual model can be seen in Table 2. As seen, the maximum score of 1 is achieved for all participants, across all metrics.

Table 2: The performance of the random individual model on the test set, when using epochs of length 5 seconds [10].

Participant	Accuracy	Precision	Recall	F1 score	Specificity
P01	1.000	1.000	1.000	1.000	1.000
P02	1.000	1.000	1.000	1.000	1.000
P03	1.000	1.000	1.000	1.000	1.000
P04	1.000	1.000	1.000	1.000	1.000
P05	1.000	1.000	1.000	1.000	1.000
P06	1.000	1.000	1.000	1.000	1.000
P07	1.000	1.000	1.000	1.000	1.000
P08	1.000	1.000	1.000	1.000	1.000
P09	1.000	1.000	1.000	1.000	1.000
P10	1.000	1.000	1.000	1.000	1.000
P11	1.000	1.000	1.000	1.000	1.000
P12	1.000	1.000	1.000	1.000	1.000
P13	1.000	1.000	1.000	1.000	1.000
P14	1.000	1.000	1.000	1.000	1.000
P15	1.000	1.000	1.000	1.000	1.000
P16	1.000	1.000	1.000	1.000	1.000
P17	1.000	1.000	1.000	1.000	1.000
P18	1.000	1.000	1.000	1.000	1.000
P19	1.000	1.000	1.000	1.000	1.000
P20	1.000	1.000	1.000	1.000	1.000

Individual model across sessions The results of the test set for the individual model across sessions are shown in Table 3. 13 of the participants scored perfectly across all metrics. The lowest accuracy was obtained for P14, with an accuracy of 50.8%. The best-performing metric is precision, with an average of 99.2%, closely followed by specificity with 99.1%.

3.4 General model

In Table 4, the results of the general model on the test set are listed. The participant column shows which participant was used in the test set. The average accuracy across all participants is 62.9%. The best-performing metric is specificity, with 81.5%. The best accuracies were achieved with P15 and P09 in the test set, with an accuracy of 94.1% and 90.2%, respectively. The lowest accuracies are obtained using P04, P19 and P02, resulting in accuracies of 25.0%, 29.4% and 29.5%, respectively.

Table 3: The performance of the individual model across sessions on the test set when the data is split into epochs of 5 seconds. Results marked in green are above the average value of that metric, and those marked in red are below average [10].

Participant	Accuracy	Precision	Recall	F1 score	Specificity
P01	1.000	1.000	1.000	1.000	1.000
P02	1.000	1.000	1.000	1.000	1.000
P03	1.000	1.000	1.000	1.000	1.000
P04	1.000	1.000	1.000	1.000	1.000
P06	1.000	1.000	1.000	1.000	1.000
P09	1.000	1.000	1.000	1.000	1.000
P10	1.000	1.000	1.000	1.000	1.000
P12	1.000	1.000	1.000	1.000	1.000
P13	1.000	1.000	1.000	1.000	1.000
P15	1.000	1.000	1.000	1.000	1.000
P17	1.000	1.000	1.000	1.000	1.000
P18	1.000	1.000	1.000	1.000	1.000
P20	1.000	1.000	1.000	1.000	1.000
P08	0.941	1.000	0.883	0.938	1.000
P11	0.908	0.845	1.000	0.916	0.816
P07	0.808	1.000	0.616	0.762	1.000
P16	0.808	1.000	0.616	0.762	1.000
P19	0.616	1.000	0.233	0.378	1.000
P05	0.555	1.000	0.466	0.635	1.000
P14	0.508	1.000	0.016	0.032	1.000
Average	**0.907**	**0.992**	**0.842**	**0.871**	**0.991**

Table 4: The performance of the general model on the test set. Results marked in green are above the average value of that metric, and those marked in red are below the average [10].

Participant	Accuracy	Precision	Recall	F1 score	Specificity
P15	0.941	1.000	0.856	0.922	1.000
P09	0.902	0.814	1.000	0.897	0.829
P13	0.875	0.750	1.000	0.857	0.800
P01	0.857	1.000	0.666	0.800	1.000
P16	0.760	1.000	0.361	0.531	1.000
P20	0.727	0.645	0.606	0.625	0.800
P05	0.694	1.000	0.267	0.431	1.000
P06	0.682	1.000	0.333	0.500	1.000
P10	0.682	0.000	0.000	0.000	1.000
P11	0.664	0.621	0.528	0.571	0.764
P08	0.642	1.000	0.044	0.085	1.000
P07	0.625	0.500	0.333	0.400	0.800
P03	0.625	0.500	0.666	0.571	0.600
P17	0.619	0.000	0.000	0.000	0.887
P12	0.550	0.000	0.000	0.000	0.880
P14	0.500	0.000	0.000	0.000	0.800
P18	0.394	0.000	0.000	0.000	0.630
P02	0.295	0.000	0.000	0.000	0.615
P19	0.294	0.000	0.000	0.000	0.470
P04	0.250	0.000	0.000	0.000	0.400
Average	**0.629**	**0.492**	**0.333**	**0.360**	**0.814**

4 Discussion

The aim of this study is to pave the way for the design of an EEG-based alcohol detection system for the onboard monitoring of drivers. In this section, the obtained results are discussed.

As Figure 2 shows, the average BAC value for all participants decreases from the first to the second measurement after drink ingestion. After this, the average BAC increases. Consequently, the peak BAC value was not reached at 25 minutes after ingestion. Even though the shape of the BAC curve can vary highly during the absorption phase, a decrease in the BAC normally indicates that the peak value has been reached and that the absorption phase is over. Therefore, this dip in value is most likely not due to the unpredictable nature of the absorption phase. A possible explanation can be that this dip is a part of a spike in the BAC value. These can be caused by the sudden opening and closing of the pyloric sphincter [5]. Using a breathalyser can also make these spikes more apparent. Therefore, the average dip in BAC value should not be emphasised. To avoid measuring a decrease like this, the BAC could be measured by using another instrument than a breathalyser, or it could be measured in intervals of more than 10-12 minutes.

Considering the increasing trend of the BAC curves (Figure 2), most participants did not reach the peak BAC value. The preferred outcome of the experiment would be to have the participants reach the BAC peak during the alcoholic recording session. This is to enable analyses of the behaviour of the participants while under peak BAC influence [10]. To increase the chances of the participants reaching the peak BAC, they could have been instructed to not eat beforehand, or they could have been served an undiluted alcoholic drink. Otherwise, the length of the each experiment session would have to be increased.

In the Flanker results, both the average accuracy (Figure 3a) and the average RT (Figure 3b) decrease from the non-alcoholic to the alcoholic Flanker task. As indicated by their p-values, both of these changes are significant. The decreases can be explained by the disinhibition caused by the alcohol. The participants answer faster due to impulsiveness, and therefore they might not be aware of the correct answer before they press a key. As all participants performed the Flanker task in the pre-experiment recording, the changes between the alcoholic and non-alcoholic Flanker tasks are believed to be caused by alcohol alone, and not nervousness.

The Flanker task was chosen as a part of this experiment to test the participants' ability to filter relevant information from irrelevant. This can be compared to how a driver needs to be aware of both the road they are driving on and their surroundings. As the results in 3 show, the consumption of even a small amount of alcohol seems to significantly affect a person's ability to make the right deci-

sion as fast as needed.

When comparing the individual models, it is clear that the random individual model performs better than the individual model across sessions. A reason for the worse performance of the latter model can be differences in the EEG signals across the recording sessions. These differences can be due to an increased impedance between the sessions, or due to the cap being placed a bit differently between the two sessions. Although the random individual model is performing better, the individual model across sessions provides a more realistic implementation of an alcohol detector. If an alcohol detector were to be calibrated for individuals before use, the EEG signals used to detect alcohol might differ slightly from the signals used to calibrate the device. This is comparable to how the data is split in the individual model across sessions. Therefore, this model is viewed as the best of these models. This is despite it performing slightly worse than the other model, but the realistic use of data outweighs the better performance.

The general model has large differences in its performance (Table 4). These results are not surprising, as there are differences in the EEG signals across participants [10]. This leads to difficulties when training and testing on different participants. The general model is struggling to correctly classify alcohol samples. Some of the participants it is struggling with have lower BAC values than the averages. This might lead to less clear alcohol features as the brain is less affected by alcohol, which can make the classification of alcohol-affected signals more difficult. Despite this, the general model does provide belief that an EEG-based alcohol detector can become a helpful tool in the future, as is is performing good on several participants (P15, P09, P13, P01). However, the performance must be significantly improved to use this approach as an alternative to a breathalyser.

The collected data set has some limitations. It was not possible to measure the scalp-electrode impedance while collecting the data. This means the impedance could be higher than desired, and this could have led to the data being more challenging to classify than it could have been with a lower impedance. During the collection of the data, construction work was performed outside the data collection room. The noise from this work could have affected the participants, and, consequently, the noise interference could have affected the classification results negatively.

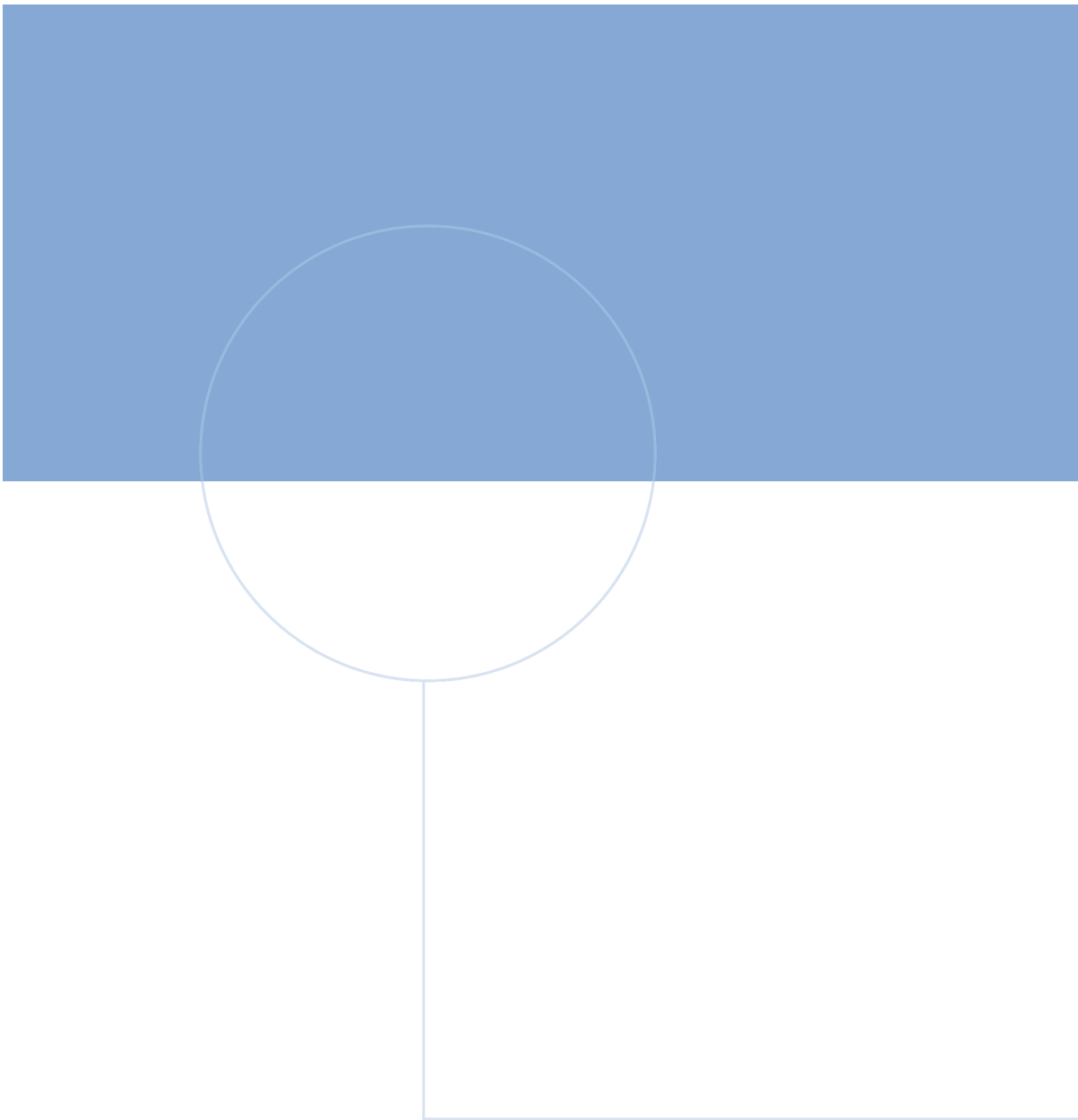
5 Conclusion

It is possible to differentiate alcohol-affected EEG signals from those that are unaffected. The Flanker results indicated that the participants were affected by alcohol, which indicates that the EEG signals might be affected as well. These results are supported by the performance of the classifiers, especially the

individual ones. The high accuracies indicate that EEGNet can extract features which characterise alcohol-affected signals. The performance of the general model is not as good, and it struggles to correctly classify alcohol-affected signals. There could be numerous reasons for this, and improving the performance should be further explored. Still, the models presented in this paper could be the first step towards creating an EEG-based alcohol detector.

References

1. Bye, E.: Alkoholbruk i den voksne befolkningen. Norwegian Institute of public Health, Webpublication **9** (2018)
2. Ek, Z., Akg, A., Bozkurt, M.R.: The classificaton of EEG signals recorded in drunk and non-drunk people. International Journal of Computer Applications **68**(10) (2013)
3. Eriksen, B.A., Eriksen, C.W.: Effects of noise letters upon the identification of a target letter in a nonsearch task. Perception & psychophysics **16**(1), 143–149 (1974)
4. Farsi, L., Siuly, S., Kabir, E., Wang, H.: Classification of alcoholic EEG signals using a deep learning method. IEEE Sensors Journal **21**(3), 3552–3560 (2020)
5. Jones, A.: Biochemical and physiological research on the disposition and fate of ethanol in the body. Medicolegal aspects of alcohol, 5th edition, Lawyers and Judges publishing company, Tucson pp. 47–128 (2008)
6. Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J.: EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. Journal of neural engineering **15**(5), 056013 (2018)
7. Mukhtar, H., Qaisar, S.M., Zaguia, A.: Deep convolutional neural network regularization for alcoholism detection using EEG signals. Sensors **21**(16), 5456 (2021)
8. National Institute on Alcohol Abuse and Alcoholism US: Alcohol use in the united states: Age groups and demographic characteristics. Available at https://www.niaaa.nih.gov/alcohols-effects-health/alcohol-topics/alcohol-facts-and-statistics/alcohol-use-united-states-age-groups-and-demographic-characteristics (2023), accessed 16.05.23
9. Nordstrøm-Hauge, I.J.: Design of protocol and collection of data for an EEG based alcohol detector (12 2022). https://doi.org/10.13140/RG.2.2.36378.11205
10. Nordstrøm-Hauge, I.J. and Vassbotn, M.: EEG-Based Alcohol Detection System with AI Techniques: Towards the Design of BCI Systems for Driver Monitoring. Master's thesis, Norwegian University of Science and Technology (June 2023)
11. Singhal, V., Mathew, J., Behera, R.K., et al.: Detection of alcoholism using EEG signals and a CNN-LSTM-ATTN network. Computers in biology and medicine **138**, 104940 (2021)
12. Vassbotn, M.: Design of protocol and collection of data for an EEG based alcohol detector (12 2022). https://doi.org/10.13140/RG.2.2.15013.37600
13. Vissers, L., Houwing, S., Wegman, F.: Alcohol-related road casualties in official crash statistics. ITF: International Transport Forum (2018)
14. World Health Organization: BAC limit data by country. https://apps.who.int/gho/data/node.main.A1002?lang=en (2020)



NTNU

Norwegian University of
Science and Technology