
**Kaunas University of Technology**

Faculty of Informatics

# EEG Signal Analysis of the Visual Cortex in Individuals with Alzheimer's Disease

Master's Final Degree Project

**Dovilė Komolovaitė**

Project author

**Prof. Dr. Rytis Maskeliūnas**

Supervisor

**Kaunas, 2022**


**Kaunas University of Technology**

Faculty of Informatics

# EEG Signal Analysis of the Visual Cortex in Individuals with Alzheimer's Disease

Master's Final Degree Project

Artificial Intelligence in Computer Science (6211BX007)

**Dovilė Komolovaitė**

Project author

**Prof. Dr. Rytis Maskeliūnas**

Supervisor

**Prof. Dr. Robertas Damaševičius**

Reviewer

**Kaunas, 2022**

**Kaunas University of Technology**

Faculty of Informatics

Dovilė Komolovaitė

# EEG Signal Analysis of the Visual Cortex in Individuals with Alzheimer's Disease

## Declaration of Academic Integrity

I confirm the following:

1. I have prepared the final degree project independently and honestly without any violations of the copyrights or other rights of others, following the provisions of the Law on Copyrights and Related Rights of the Republic of Lithuania, the Regulations on the Management and Transfer of Intellectual Property of Kaunas University of Technology (hereinafter – University) and the ethical requirements stipulated by the Code of Academic Ethics of the University;

2. All the data and research results provided in the final degree project are correct and obtained legally; none of the parts of this project are plagiarised from any printed or electronic sources; all the quotations and references provided in the text of the final degree project are indicated in the list of references;

3. I have not paid anyone any monetary funds for the final degree project or the parts thereof unless required by the law;

4. I understand that in the case of any discovery of the fact of dishonesty or violation of any rights of others, the academic penalties will be imposed on me under the procedure applied at the University; I will be expelled from the University and my final degree project can be submitted to the Office of the Ombudsperson for Academic Ethics and Procedures in the examination of a possible violation of academic ethics.

Dovilė Komolovaitė

*Confirmed electronically*

Komolovaitė, Dovilė. EEG Signal Analysis of the Visual Cortex in Individuals with Alzheimer’s Disease. Master's Final Degree Project / supervisor prof. dr. Rytis Maskeliūnas; Faculty of Informatics, Kaunas University of Technology.

Study field and area (study field group): Informatics (B01), Computer Science (B).

Keywords: EEG; classification; face inversion detection; Alzheimer's disease.

Kaunas, 2022. 65.

## Summary

Visual comprehension is essential activity in everyday interactions and occurs unconsciously. However, understanding the visual for individuals affected by Alzheimer's disease is often a challenging task. Because Alzheimer's disease is defined by a memory processing impairment, this study examines the visual cortex of an AD patient before memory processes begin. The study proposes a face inversion system that utilizes raw EEG data, performs preprocessing, transformation into GASF images, augmentation using the RGAN technique, and finally classification into two groups based on whether the visible face is upright or inverted. It has been demonstrated that transforming the EEG signal into images improves classification accuracy and reduces training time. Because the impacts of emotions and familiarity on this EEG signal data were found to be insignificant, study focused on the effects of colors. When comparing the performance of four different classifiers, including ResNet-50, Custom CNN, and EEGNet SSVEP; our proposed CNN, which was designed for smaller datasets utilizing the best regularization approaches, produced the greatest results in both color and grayscale analyses. Meanwhile, just a quarter of the data injections based on RGAN-generated signals demonstrated weak statistical significance. The study's findings suggest that extracting EEG signal features of facial processing in color images is more successful than in grayscale images. The control subject data achieved the average accuracy of 61.7% in detecting face inversion position processing color images and only 57% when processing grayscale images. Comparing the results between the oldest subject and the AD patient, the data from the Alzheimer's patient indicated very low feature extraction success. This could mean that the disease directly affects the human visual cortex, making it more difficult to recognize objects using memory resources.

Komolovaitė, Dovilė. Alzheimerio liga sergančių asmenų regos žievės EEG signalų analizė. Magistro baigiamasis projektas / vadovas prof. dr. Rytis Maskeliūnas; Kauno technologijos universitetas, Informatikos fakultetas.

Studijų kryptis ir sritis (studijų krypčių grupė): Informatika (B01), Informatikos mokslai (B).

Reikšminiai žodžiai: EEG; klasifikacija; veido inversijos nustatymas; Alzheimerio liga.

Kaunas, 2022. 65p.

## Santrauka

Vizualinis suvokimas yra kasdienių sąveikų procesas, vykstantis nesąmoningai. Tačiau suprasti atvaizdus Alzheimerio ligos paveiktiems asmenims yra dažnai sudėtinga užduotis. Kadangi Alzheimerio liga apibūdinama atminties apdorojimo sutrikimu, šiame tyrime analizuojama AD paciento regėjimo žievė prieš prasidedant atminties procesams. Darbe siūloma veido inversijos sistema, kuri naudoja neapdorotus EEG signalus, atlieka išankstinį apdorojimą, transformavimą į GASF vaizdus, padidinimą naudojant RGAN techniką ir galiausiai klasifikavimą į dvi klases pagal tai, ar matomas veidas yra vertikalus, ar apverstas. Įrodyta, kad transformavus EEG signalą į vaizdus, pagerėja klasifikavimo tikslumas ir sutrumpėja modelio apmokymo laikas. Kadangi buvo nustatyta, kad emocijų ir pažįstamumo įtaka šiems EEG signalo duomenims yra nereikšminga, tyrimas buvo analizuojamas pagal spalvų poveikį. Lyginant keturių skirtingų klasifikatorių, įskaitant ResNet-50, Custom CNN ir EEGNet SSVEP, našumą; mūsų siūlomas CNN modelis, kuris buvo sukurtas mažesniems duomenų rinkiniams, naudojant geriausius reguliavimo metodus, pateikė geriausius rezultatus tiek spalvotoms, tiek pilkoms nuotraukoms. Tuo tarpu RGAN generuojamais signalais duomenų padidinimas ketvirtadaliu parodė silpną statistinį reikšmingumą. Tyrimo išvados rodo, kad EEG signalo veido apdorojimo ypatybių ištraukimas spalvotuose vaizduose yra sėkmingesnis nei naudojant pilkų atspalvių vaizdus. Kontrolinio subjekto duomenys pasiekė vidutinį 61.7% tikslumą nustatant veido inversijos padėtį apdorojant spalvotus vaizdus ir tik 57% apdorojant pilkos spalvos vaizdus. Palyginus seniausio tiriamojo ir AD sergančio paciento rezultatus, Alzheimerio liga sergančio paciento duomenys rodo, kad esminių signalo savybių išgavimo sėkmė yra labai maža. Tai gali reikšti, kad liga tiesiogiai paveikia žmogaus regėjimo žievę, todėl naudojant atminties išteklius jam yra sunkiau atpažinti objektus.

## Table of Contents

## List of figures

## List of tables

## List of abbreviations and terms

| | |
|---|---|
| **2D** | Two-dimensional. |
| **AD** | Alzheimer's disease. |
| **Adam** | Non-stationary optimization method combining AdaGrad and RMSProp methods. The name comes from the adaptive moment estimation. |
| **AUC** | The area under the ROC curve. |
| **BCI** | Brain–computer interface. |
| **CNN** | Convolutional neural network. |
| **CUDA** | Compute unified device architecture. |
| **EEG** | Electroencephalography. |
| **ERP** | Event-related potential. |
| **FIR** | Finite impulse response. |
| **GAF** | Gramian angular fields. |
| **GAFD** | Gramian angular field difference. |
| **GAN** | Generative adversarial network. |
| **GASF** | Gramian angular summation fields. |
| **Git** | Global information tracker. It is an open-source distributed version control system. |
| **GPU** | Graphics processing unit. |
| **LSTM** | Long short-term memory. |
| **ResNet** | Residual neural network. |
| **RGAN** | Recurrent generative adversarial network. |
| **RNN** | Recurrent neural network. |
| **ROC** | Receiver operating characteristics. |
| **STFT** | Short-time Fourier transform. |
| **SVM** | Support-vector machine. |
| **WGAN** | Wasserstein generative adversarial network. |

## Introduction

The ability to assign meaning to visible objects is an essential part of everyday life. The brain constantly performs visual recognition of the environment, although we do not always consciously understand the result. To monitor brain activity information, the most popular tools are computed tomography, magnetic resonance imaging, but electroencephalography (EEG) systems are currently the most available. Therefore, this work uses EEG data that records the voltage at each electrode placed on the scalp to measure electrical potentials.

Accurate prediction of visual stimuli can help design better human-computer interactions for both healthy and sick individuals. Knowing what type of object, word, number, or other visual stimuli a person is concentrating on might help provide more targeted tool commands or help with better communication. For instance, paralyzed patients and those with memory loss may benefit from the ability to predict visual cues. This can assist paralyzed people connect with the outside world and express needs, and conversely, it can help patients with Alzheimer’s communicate with themselves. As in the latter case, trained models based on brain signals may be able to give additional information about the visible face and the state of the relationship. Similar applications could be developed for integrated computer games or smart house devices.

Unlike recent publications that try to determine whether a person is suffering from memory loss diseases. This study aims to discover the possibilities of recognizing facial stimuli from brain signals, thus paving the way for the development of brain-computer interfaces for patients with memory problems. Faces have been proven to be processed in the brain differently from other stimulus types, making them a distinct niche. Given that memory impairment is frequently linked to issues with facial recognition, there may be possibilities to classify an image category during the visual processing region of the brain before the brain understands how to label it. If it is true, this might be a valuable insight into the effects of memory loss on the brain.

The study uses electroencephalography data, which helps to analyze the brain and its behavior in relation to the signal frequency, as well as allowing for a more accurate interpretation of brain diseases such as epilepsy or memory loss. The participants in the data collected are the elderly and one patient with Alzheimer's disease with a facial recognition deficit. During the experiments, photos were shown in both upright and inverse positions. The first experiment we focus on employs color face photos with an emotion factor, whereas the second uses grayscale images with a familiarity component. The face inversion classification provides additional information on whether picture color, emotion, or familiarity influence the visual processing of the face. The study leverages EEG signal channels from the occipital lobe of the brain to create a facial inversion recognition system from a single-trial epoch.

## Project novelty and relevance

Although the face inversion classification task has been performed in the past, in this work, inversion recognition based on color influence is performed. The data used in the research comes from older people's signals, which are sensitive, noisy, and unstable, which is unusual in non-sickness related tasks. Because Alzheimer's disease is more common in elderly people, this group is chosen as a control group. Consequently, data from an Alzheimer's patient with a face recognition deficit is used to test the hypothesis that if memory processing in the brain is impaired but not the part of the brain responsible for visual comprehension, then brain-computer interfaces can be used to help memory loss patients identify visible stimuli. This hypothesis is being explored by seeing if deep learning

models can extract essential features from Alzheimer's patient data in the same way that they do from healthy people. Such a condition has never been studied previously, to our knowledge.

## Aim and Objectives

The aim of the study is to determine if visual stimuli could be predicted from EEG signals as precisely in Alzheimer's patient as they can in healthy people, in order to better understand how similarly these two groups process visual information.

**Objectives:**

1. Predict the scope of attention of the control subject and the patient with Alzheimer's disease from single-trial EEG data from color images of visual stimuli.
2. Predict the scope of attention of the control subject and the patient with Alzheimer's disease from single-trial EEG data from grayscale images of visual stimuli.
3. Investigate how augmenting synthetic data created by generative adversarial networks to a training dataset may improve the classifier's performance.

## Structure of the Report

This paper focuses on face inversion processing investigation for a healthy subjects and an Alzheimer's disease patient using EEG signals. The structure is organized as follows. The first section discusses similar work in visual stimulus observation, a literature analysis, and technical details of the study's approach. The second section focuses on project planning, visioning, and architectural construction. The final section highlights and examines the results. Finally, the final section summarizes the work and makes recommendations for future improvements.

## 1. State of the Art

The human brain is one of the most complex organs that interacts with billions of neurons [1]. There are several techniques to map brain signals, but EEG technique is non-invasive, quick to use, and provides high temporal resolution at milliseconds time precision [2]. Not to mention lower costs than other devices, EEG has been shown to be effective in extracting brain signal features, even when the subject is not yet aware of the stimuli [1].

Furthermore, the ability to detect abnormal rhythms by monitoring brain potential can assist disabled individuals in properly diagnosing illnesses [3]. EEG has been proven to be a suitable tool for distinguishing between dementia severities [2], traumatic brain injuries, and post-traumatic stress disorder [4]. Individuals with brain problems tend to lose long-term attention throughout a task, making cognitive and perceptual processes harder to complete [4]. Based on the classified EEG frequency bands, subjects affected by Alzheimer's have higher activity in the theta and delta frequency bands and lower activity in the alpha and beta frequency bands [2]. Meanwhile, the normal signal frequency range is between 1 Hz and 100 Hz, bearing in mind that the frequency of 100 Hz is very rare, and the amplitude varies between 10μV and 100μV [1].

An important feature of EEG signal is that it is highly sensitive to individual differences [5], therefore any brain-computer interface (BCI) must be user-specific. The activity itself also depends on age, mental state, and many other factors [1]. Consequently, only adjusted BCIs can be used for people with disabilities to give commands using a brain signal to a computer [1].

Poor spatial resolution and low signal-to-noise ratio cause many problems for EEG data. Due to the ongoing signals of various brain activities, the signal has various artifacts. Another important challenge is the feature extraction of random time-varying EEG signals [1]. The other problem is how to classify EEG as accurately as possible using only one stimulus presentation [1], as it is known that multiple presentations can drastically improve performance [5], but this is not suitable for use in real-time applications. However, with the development of signal processing techniques and innovative solutions for machine learning and deep learning, it is now possible to better understand and predict brain impulses than ever before.

### 1.1. Visual Observation using EEG

Visual observation, otherwise known as attention, is used to analyze whether different visual stimuli cause different brain responses. For many years, scientists have been trying to distinguish what a subject observes in front of using only brain signals. Such a task is usually successful in predicting patterns of two very different classes of displayed images. The following is a brief overview of the articles that try to anticipate the scope of attention.

In a study that analyzes attentional focus, it was discovered that single-trial EEG activity could be used to predict stimulus type or even performance errors (whether the observer was correct or not). These conclusions were made using a machine learning algorithm - a **linear pattern classification**. In the first experiment, thirteen participants received stimuli consisting of an odd or even number, and in the second experiment, a word, or a random set of letters. The authors used standard linear discriminant analysis for pattern classification, where the input data were all 64 electrodes. For the dimensionality reduction, the responses of each electrode were divided into non-overlapping 20 ms time bins and averaged, starting with stimulus onset. Until finally, the results were evaluated using a

10-fold cross-validation scheme. The ability to distinguish between two classes of stimuli and the correctness, even when individuals are distracted, suggests that pattern classification algorithms combined with continuous measurement of EEG response may be a promising tool for monitoring cognitive states [6].

The study, which included eight individuals, used a **linear support vector classifier**, which reliably decoded the perception of upright and inverted faces against a gray background. The EEG signal data were used at each time point (~ 1ms resolution) using approximately 62 electrodes. The results were validated by a 10-fold cross-validation procedure. The aim of the study was to identify the time bins in which the model successfully determined the EEG correlate. The results showed that the accuracy of facial inversion (56% on average) peaked over the 125–250ms latency after stimulus onset, which coincides with the ERP component N170. In future research, the authors recommend further increasing the accuracy of prediction by using the average signal over time or by applying Bayesian statistics [7].

Another article used a deep neural network model to predict human response times to visual stimuli. The models were fed with periodogram features, and the highest accuracy was achieved using the **convolutional neural network (CNN)** - around 94% for binary class classification and 78% for 3-class classification. This research was unique because it used spectrotemporal features from different EEG channels rather than temporal ones. And finally, after evaluating the importance of each channel, it was found that at least 7 channels are needed to perform the visual response task of attention with almost maximum accuracy. Out of all 30 EEG channels, the dominant ones were distinguished to be the left central (C3, CP3 and CPZ), parietal (P3, PZ and P4) and occipital (OZ) lobes. Although this study is not directly related to visual stimulus detection, but only to reaction time prediction after visual stimuli, its success in achieving high accuracy and faster model performance that would be suitable for real-time use reveals deep neural network's potential in working with EEG signals [8].

Meanwhile, in the other two articles, in which the face class was included as a visual stimulus, the **support vector machine (SVM)** classifier was used and an accuracy of 77% [4] and ~80% [9] was achieved. The first article used composite images of overlapping face and outside pictures, and the second used face versus non-face data.

**Table 1.1.** Summary of papers classifying EEG visual stimulus

| Article | Methods | Goal | Accuracy |
|---|---|---|---|
| „Decoding Information Processing When Attention Fails: An Electrophysiological Approach“ [6] | Linear Pattern Classification | predict stimulus type:<br>• odd/even number<br>• word/letters | 90% and 72% respectively |
| „Pattern classification of EEG signals reveals perceptual and attentional states“ [7] | Linear support Vector | • upright/inverted faces | 56% |
| „Deep Neural Network for Visual Stimulus-Based Reaction Time Estimation Using the Periodogram of Single-Trial EEG“ [8] | Convolutional Neural Network | • estimate visual stimulus reaction time | 94% (binary) and 78% (3-class) |
| „Decoding Attentional State to Faces and Scenes Using EEG Brainwaves“ [4] | Support Vector Machine | • scene/face images | 77% |
| „EEG Correlates of Categorical and Graded Face Perception“ [9] | Support Vector Machine | • natural scenes/genuine faces | ~80% |

A summary of the reviewed articles analyzing brain responses to visual stimuli is provided in **Table 1.1.** The classification of upright and inverted faces was achieved with only 56% accuracy, while there are situations where visual stimuli can be predicted with 94%. Although there are many factors that can lead to incomparable results: the inability to replicate studies identically, the different data quality, the different number of channels, and so on. Nevertheless, the analyzed papers provide a global picture of the methods used and the accuracies achieved.

## 1.2. EEG Channels

EEG signal data is collected from a device called an electroencephalogram, in which electrodes are placed to perform electrical activity measurements from the scalp of the brain [3]. However, it is important to understand that this data is only a crude representation of brain activity. Electrodes on the scalp measure the activity of large groups of neurons, and small signals are attenuated and scattered due to the physical barrier between the brain and the electrodes [5]. These electrodes are arranged according to the specified standards and marked by letters (F - Frontal, T - Temporal, C - Central, P - Parietal, O - Occipital, z - midline region). Here, odd numbers indicate the left hemisphere and even numbers the right hemisphere [3]. Because the brain is divided into two hemispheres, lower case letter $r$ or $l$ can be added to the component name to indicate from which side of the brain it was measured.

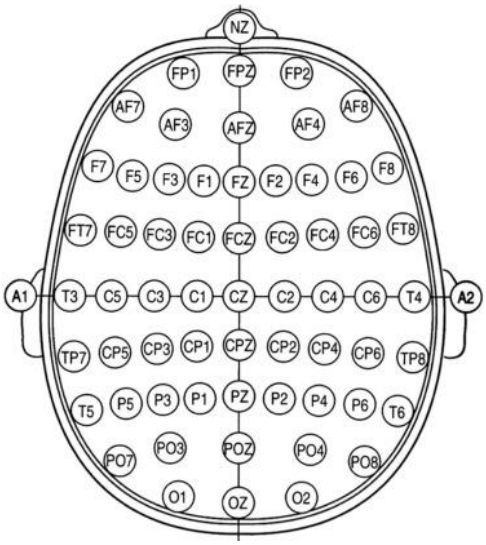



**Fig. 1.1.** 10-10 electrode system for EEG recording [10]

Different electrode positions describe different brain functions:

- "Cortex around Cz, C3 and C4 locations deals with sensory and motor functions" [3].
- "Pz, P3 and P4 are related to cognitive processing" [3].
- "Oz, O1 and O2 deals with visual processing stimuli" [3].
- "Fz is placed near intentional and motivational centers" [3].
- "F8 and F7 are located close to emotional and verbal expressions" [3].
- "F3 and F4 are located at motor planning activities" [3].
- "FPz, FP1 and FP2 deal with attention and judgment impulses" [3].

As a result, it is possible to make judgments about distinct parts of the brain by analyzing brain activity. The frontal lobe, for example, can have an impact on emotions and memory. The occipital lobe, on the other hand, is involved in perception and visual information processing [3].

### 1.3. Face Inversion Effect in EEG Signal

Face inversion effect occurs when stimuli are inverted 180$^{o}$. This effect tends to increase the perception error rate and response time of inverted faces compared to upright faces. For example, in one study with normal vision volunteers, the processing of inverted faces in ERP components such as N170 and N250r has a significant delay of about 30ms [11]. The amplitude of N170 facial component is consistently higher and delayed longer, there is evidence that it is sensitive to facial identity and facial emotions as well [9]. Another study with the same data found that P1, N170, and N250 are related to face-like processing, including the inversion effect [12]. Meanwhile, there are sources where the delay of N170 is thought to be small, about 10ms, but consistent, and facial inversion is also observed with a delay of 11ms from the N200 [13]. There have been a lot of debates whether the latency in inverted faces is caused by memory or perceptual encoding. However, because inversion in ERPs is obtained on a very early potential peaking around 170ms after stimulus, there is a growing consensus that face inversion effect occurs primarily through perceptual level encoding rather than long-term memory encoding [13][11]. Processing tends to be delayed because when individuals see upright faces, they perform holistic processing, and when they see inverted faces, they perform feature processing [12].

### 1.4. Emotional and Familiar Faces

Visual processing in EEG is influenced by emotions and familiarity. It was found that facial expressions encoded in EEG from 120ms to 750ms after the visual stimulus onset, with the highest peak at 270ms [14]. The largest variances in different emotions were seen in the alpha and beta waves, according to an investigation that detected the emotional stages in EEG signals while listening to music [15]. Whereas facial familiarity increases signal amplitude at an early stage due to the knowledge and memories of the individuals involved [16]. The sensitivity of familiar faces is determined at a frequency of approximately 250ms [14].

### 1.5. Color Influence on Visual Stimuli

Color enhances environmental perception. While it may appear to be a minor feature, color presence has a significant influence, therefore it is worth considering while dealing with EEG signals. For example, in an article classifying landscape images such as forest, desert, and water, it was found that color images were more accurately predicted than gray images [17]. Using the SVM classifier, performance varies with an accuracy of 5% [17]. This task was investigated using data from 20 healthy volunteers aged 25 to 55 years [17]. Another article, with an average age of 31 years old participants, supported the same idea that better classification results could be achieved using color images rather than grayscale, as this allows more useful information to be extracted [18]. Here, it was demonstrated that the average accuracy of color images was 3.4% greater for the SVM classifier [18]. Furthermore, research on face perception discovered that color information improves the classification of human faces, with colored faces again being classified more accurately [19]. This leads to the conclusion that the impact of colors must also be considered.

### 1.6. Different Frequency Brain Waves

„One way to analyze EEG data is to look at the spectral power of the signal in a set of frequency bands, which have been observed to correspond with certain types of neural activity“ [5]. EEG waveforms consist of different frequencies that determine a normal or abnormal rhythm. The usual

classification according to the frequency of the signal is divided into alpha, beta, theta, delta, and gamma waves [3]. They can be derived from the frequency bands listed in the **Table 1.2.**

**Table 1.2.** Brain wave frequency bands [3]

| Frequency Band | Frequency Range (Hz) |
|---|---|
| Delta | 0 - 4 |
| Theta | 4 - 8 |
| Alpha | 8 - 13 |
| Beta | 13 - 30 |
| Gamma | 30 - 100 |

„Brain wave frequency differs corresponds to different behavior and mental states of the brain“ [3].

- **Delta** wave activity is found at all stages of sleep and reflects gray matter in the brain [3].
- **Theta** waves reflect subconscious activity and can be observed in deep relaxation and meditation [3].
- **Alpha** waves are observed during the relaxed awakening phase and reflect white matter in the brain [3]. These waves have a higher amplitude when the participant is sleepy or bored [20].
- **Beta** waves are related to behavior and are observed in a conscious state [3].
- **Gamma** waves reflect processes of perception and consciousness [3].

Performing visual tasks, theta and alpha waves were found to be the most active ones. Their fluctuations play an important role in sensory tasks associated with constant attention and mental fatigue [8].

## 1.7. Event-Related Potential

The event-related potential (ERP) is the most widely used technique to study cognitive processing in response to a stimulus. Since the EEG signal activity is noisy, the experiment is repeated several times for each individual and then the average of the EEG signals is taken to derive the ERP values, resulting in a *grand average ERP* [6]. „These grand average responses can be compared across different stimulus classes to make statistical statements about the signals“ [5]. By averaging and aligning multiple signals with the onset of the stimulus, noise is reduced and only the systematic electrical potential associated with the processing of that stimulus remains [6]. Grand average ERP statistical analysis is useful and helps to better understand different brain responses, however, to adapt to real-time computer interfaces, a single-trial signal must be sufficient to predict visual stimuli [5].

The most important feature of ERP is that it has temporal resolution that measures brain activity in milliseconds, where states of perception and attention appear. In G. F. Woodman’s article, he summarized it as follows: "ERP components are defined by their polarity (positive or negative going voltage), timing, scalp distribution, and sensitivity to task manipulations" [20]. He argues, based on many sources and experiences, that all the different ERP components are different in amplitude when presented with different stimuli and varies by individuals [20]. In general, „ERP components such as N100, N170 and N200 which appears 100, 170 and 200ms after the onset of a visual stimulus correspond respectively to a selective attention, the processing of color, shape, and rotation (e.g., processing of human faces) and a degree of attention“ [21]. Meanwhile, in a more general case, ERP is divided into early (100–220ms) and late (350–550ms) features. Early features correspond to visual

processing, and late features - to post-sensory processing, where the brain consciously decides what was presented as a stimulus [5].

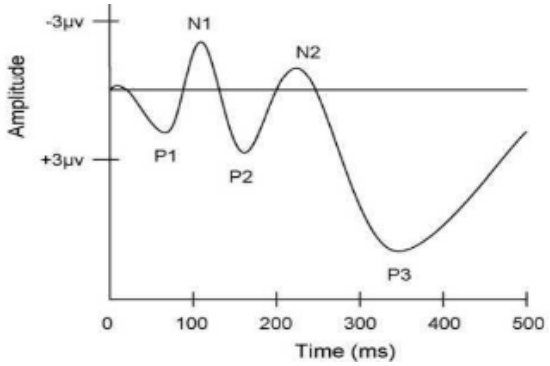



**Fig. 1.2** Event-related potentials main components [21]

In the **Fig. 1.2** the ERP waveforms voltages are plotted with negative going up to make them easier to interpret [21]. Here, N means "negative" and P means "positive" polarity, and the number defines the delay in milliseconds.

Here are the most influential components in terms of peak time after face stimulus onset:

- **The N170** component is particularly important for visual discrimination and perceptual processing, especially when distinguishing faces [20]. Moreover, N170 response is sensitive not only to human faces but schematic faces as well, this component is higher in the left hemisphere for featural processing (eyes, nose, and mouth), and in the right - for holistic processing [12].
- **The N250** is particularly sensitive to face identity. This component relates to personal detection processing in the right hemisphere, where the amplitude increases by observing familiar objects such as friends or family [12].
- **The P1 or P100** „reflects the processing of low-level physical properties, including contrast, luminance, spatial frequency, and color" [12]. It has also been observed that the amplitude of unpleasant stimuli P1 is higher than that of pleasant stimuli, so there is a correlation between P1 and facial expression [21].

Because the post-stimulus signals up to 250ms contains the most descriptive information for face processing, 250ms EEG windows can be used as input for the predictor [22].

## 1.8. EEG Signal Processing Stages

Traditional EEG signal processing stages consist of 4 main steps (see **Fig. 1.3**):

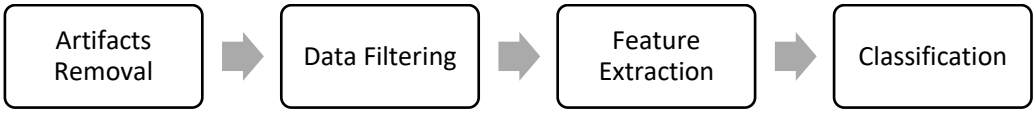



**Fig. 1.3.** Traditional EEG signal processing stages [3]

Artifact removal and data filtering is a preprocessing step, and feature extraction and classification defines post-processing [3]. However, the structure of this study will be similar but different, the time series are encoded as images and the deep neural network learns to find the optimal features itself. The following sections describe the steps of the implemented design.

## 1.9. Preprocessing

The raw EEG signal data record electrical activity (voltage fluctuations) from multiple neurons around the electrode, therefore the signal itself is reduced due to the resonant signals and contains noise [3]. Consequently, the preprocessing stage is very important and usually consists of following steps (see **Fig. 1.4**): channels selection from relevant areas of interest, data filtering, data trimming into epochs, and artifact removal.

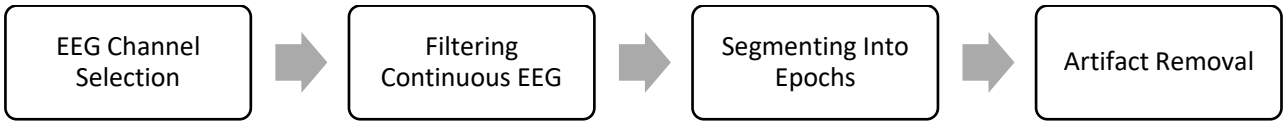



**Fig. 1.4.** EEG signal preprocessing steps

Some noise in the continuous raw signal can be filtered out. To eliminate low-frequency noise caused by breathing, it is common to use high-pass-frequency filters with a cut-off frequency of 0.5Hz. In contrast, low-pass-frequency filters with a 40-70 Hz cut-off are used to eliminate high-frequency noise. Also, spatial filtering can be performed using the Common Average Reference technique, using a window length of 4s with a 3.75s overlap [1]. However, most ERP components are in the same frequency band as the noise, and it is not possible to remove the noise without distorting the signal of interest [20]. Therefore, usually studies select lower frequency limits (~0.1–30Hz) [5][7][9] or (~0.1–40Hz) [23][4] and use a band-pass hamming windowed finite impulse response (FIR) filter [23] for inherited stability and computational efficiency [5]. A high-frequency filter of 1 Hz was selected in a study examining the deficits of Alzheimer's facial recognition [24]. Pre-filtering at very low frequencies (less than 0.5 Hz) may be inefficient, while filtering high-frequency data between 1 and 2 Hz has been demonstrated to improve signal-to-noise ratio and single-trial classification accuracy [25][26].

Prior to working with artifacts, the continuous signal must be segmented into epochs of the desired length. In one article, a segment begins 100ms before the stimulus onset with a length of 600ms [9]. Elsewhere, 1-s epochs are used with 250ms before the stimulus to 750ms after stimulus onset [7]. It is very important to choose the right window size, as incorrect settings can corrupt time and amplitude measurements, especially when removing artifacts [20].

Identifying and removing artifacts is a complex step as they can be caused by a variety of factors such as head movement, blinking (±100μV), device and head connection problems. These artifacts will create signals of abnormal frequency and shape [3]. It is common practice to simply discard trials containing artifacts and perform further analysis of the "clean" data [6]. In this way, ambiguous data are avoided, as eye movements radically change the input of the visual system. As an example of artifact removal, in one article that analyzed face perception using EEG data, segments were excluded for blinking activities (threshold +/− 70μV) or eye movements (threshold +/− 50μV). The individual channels were then marked as bad if the amplitude difference between the maximum and minimum values across the segment exceeded 80μV. If more than 10% of the channels in the segment were marked as bad, the entire segment was not included, otherwise the bad channels were replaced using spherical spline interpolation [9]. In another article, meanwhile, the trials with an amplitude greater than 75μV were eliminated to exclude facial movement artifacts [4].

All preprocessing thresholds are approximated and selected for a specific task. Experiments are usually conducted in a relatively calm atmosphere so that the brain signals are not affected by background noises and the subject can maintain a stable signal. Each person, however, is unique, and therefore each signal strength in the brain is distinct. One of the most common issues with EEG is that the characteristics are difficult to generalize across all participants [27]. All EEG signal acquisition is basically done in the time domain. For a method to be valid for a BCI, it must use EEG data from a single-trial and be able to perform calculations quickly. Thus, after data processing follows a further step – encoding time series as images.

### 1.10. Encoding Time Series as Images

Feature extraction techniques are used to obtain the most informative properties that reflect brain activity. Signal features can be obtained by combining them from both domains: temporal and frequency [4] or used separately. The most popular way to extract features is to use the Fourier transform for the divided signal into four frequency bands (delta, theta, alpha, and beta) [3]. However, this approach is noise sensitive [3]. Not to mention, the accuracy and time efficiency of the classifier are highly dependent on this choice of features. High processing time can affect the interface so that it is not suitable for real-time solutions.

In time series modeling, extracting temporal correlations from time series data remains a major difficulty [28]. Therefore, one study proposed a novel method to encode time series into images called Gramian Angular Summation (GASF), Difference Fields (GAFD) and Markov Transition Fields [29]. This approach was inspired by deep learning success in computer vision and created by Zhiguang Wang and Tim Oates in 2015 [29]. Since medical images have recently been classified with very high accuracy using CNN architectures, scientists believe the possibility to explore signals in the form of images [30]. This transformation to images proved to be affected: "The recognition accuracy of motor imageries of movement with the right and left hand and the state of rest was 97% for the studied EEG signals [31]". In another article, the epilepsy detection from EEG signals was classified using GASF transformation and an F1 score of 0.90 was obtained [30]. And using a new method based on visualization of GASF images of Hjorth parameters, the study achieved 86% accuracy in classifying an arrhythmia case [32]. After time series encoding to images, the pre-trained or customized convolutional neural network architectures can be trained to automatically extract features. And, once trained, these networks quickly calculate outputs, making them more suitable for brain-computer interfaces [33].

#### 1.10.1. Gramian Angular Fields

The Gramian Angular Fields (GAF) method converts the signal data to a polar coordinate system to get a 2D matrix [31]. Firstly, signals are treated as time series, then a two-step procedure is performed to normalize into the range of [0,1] and map signals into polar coordinates. Finally, the temporal correlation over different time intervals is determined using the angular field between each observation point [33]. A detailed explanation with formulas is provided below.

Having a time series $X = x_1, x_2, \dots, x_T$ with the number of real-value observations $T$, rescale values to the interval [0, 1] by [29]:

$$\hat{x}_i = \frac{x_i - \min(X)}{\max(X) - \min(X)} \quad (1)$$

Rescaled data to $[0, 1]$ will have angular bounds in $\left[0, \frac{\pi}{2}\right]$.

Then, the procedure for mapping the normalized values to the polar coordinates are obtained by the following formula [33]:

$$\begin{aligned} \varphi_i &= arccos(\hat{x}_i), \qquad \hat{x}_i \in [-1,1] \\ r_i &= \frac{t_i}{T}, \qquad\qquad\quad t_i \in N \end{aligned} \tag{2}$$

Here $t_i$ – the time instant, $T$ – the vector length, $r_i$ – the radius, $\varphi_i$ – the angle in the Cartesian coordinate system. Due to the arccos bijective characteristic of the values in the range [0,1], it associates the data of each time series with only one possible polar coordinate [33].

To form a Gramian (or so-called Gram) matrix, the angle operation is considered [33]:

$$x + y = cos(\varphi_x + \varphi_y) \tag{3}$$

Then, a Gram matrix is created where each element is equal to the cosine of the sum of angles [33]:

$$GAF = \begin{bmatrix} cos(\varphi_1 + \varphi_1) & \dots & cos(\varphi_1 + \varphi_T) \\ \dots & \dots & \dots \\ cos(\varphi_T + \varphi_1) & \dots & cos(\varphi_T + \varphi_T) \end{bmatrix} \tag{4}$$

And **Gramian Angular Field Summation** is build using this formula [33]:

$$GAFS = cos(\varphi_i + \varphi_j) = X' \cdot X - \left(\sqrt{I - X^2}\right)' \cdot \sqrt{I - X^2} \tag{5}$$

Here $X$ – the time series row vector and $I$ – the unity vector $[1,1, \dots, 1]$. Given the sine of the angle difference, the **Gramian Angular Field Difference** matrix can be defined and calculated analogously [33].

$$GAFD = sin(\varphi_i + \varphi_j) = \left(\sqrt{I - X^2}\right)' \cdot X - X' \cdot \sqrt{I - X^2} \tag{6}$$

After the transformation, GAFS and GAFD matrices are converted to images with a single-color channel (see **Fig. 1.5**) [33].

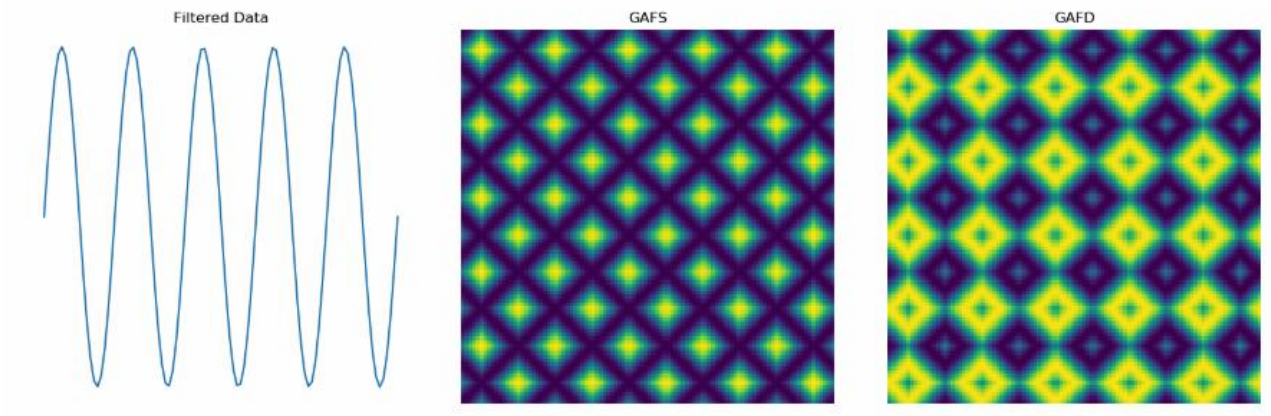



**Fig. 1.5.** An example of GAFS and GAFD building by taking a sin-wave [33]

In summary, the advantage of these matrices is that the diagonal values are rescaled initial time series values [33]. Then, the temporal relations are considered, because as time increases, the position moves from top left to bottom right [29]. The main drawback is that the size of the data has changed from a single vector to a 2D matrix based on the length of the time series [33].

### 1.10.2. Signal Encoding to Images Using Other Methods

Traditionally, signal data is converted to an image using a Short Time Fourier Transform (STFT). In this way, the analysis of the time-frequency signal can be considered as the analysis of non-stationary signals with time-varying frequency content. Analyzing the signal by frequency usually improves the performance of classification systems [34]. However, most methods require some domain knowledge [35], then parameters need to be optimized as well, such as that STFT has window type and window length parameters that need to be set and can significantly distort the time and frequency spectrum if not set properly [36]. Such dependencies, as well as the changing signal-to-noise ratio, have attracted researchers to use neural networks for the ability to learn to distinguish features based on data and the problem [37]. Numerous studies with different approaches have been performed to obtain a better classification accuracy and a shorter training time, but often they cannot be compared due to differences in data, and the inability to replicate experiments identically. Information on the studies performed where images of EEG signals were used is presented in the **Table 1.3**.

**Table 1.3.** Summary of papers that used signal encoding into images

| Study | Experiment | Goal | Feature extraction | Classifier | Accuracy |
|---|---|---|---|---|---|
| [38] | Participants who watched music videos and classified them into 4 emotional states | Emotion recognition | Time frequency domain features | Optimized SVM | 93.86 % |
| [39] | Participants who watched music videos and classified them into 4 emotional states | Emotion recognition | The spatio-temporal representations | Three-dimensional convolutional neural networks | 99.73% |
| [36] | Patients who underwent ENT surgery | Prediction of depth of anesthesia | Modified Short-time Fourier transform | CNN (VGG) | 93.50% |
| [35] | Child scalp EEG data from 23 cases with 163 epileptic seizures | Classification of four epilepsy conditions | Power spectrum density energy diagrams | DCNNs based on transfer learning | 90% |
| [30] | Normal and epileptic participants | Epilepsy detection | Gramian Angular Summation Field | The Custom CNN | 90% |
| [37] | Participants who performed imagery tasks | Identifying left and right hands motion imagery | Continuous Wavelet transform | Simplified convolutional neural network | 83.2% |
| [31] | Participants watched a black screen and imagined the movement of the hand for three seconds | Motion imagery classification | GAF conversion for each signal from 64 channels | CNN | 97% |

## 1.11. Deep Learning for EEG classification

Deep Learning models learn hierarchical representations of input data through sequential nonlinear transformations. This could improve EEG processing to make it more general: steps such as

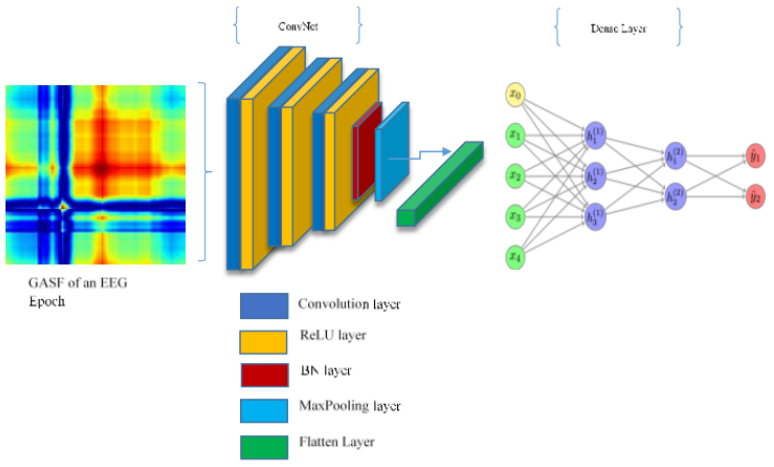
preprocessing, feature extraction, and classification would be trained and learned automatically. However, there has been minimal research on the architectures of deep learning models in the context of time series analysis tasks [28]. One study reviewing 154 articles found that the most used classes of deep neural networks for EEG are convolutional neural networks (40% of all papers) and recurrent neural networks (13%) [40]. As a result, transforming time series to images is beneficial because there is more scientific support for two-dimensional data.

Since EGG signal features are encrypted into images, several studies have already been conducted that show that CNNs work well with these new images. One such study shows that CNN does not overfit, and another important finding was that adding layers or increasing the complexity of the model does not appear to have helped achieve higher accuracy [33]. In another research where normal and epileptic cases were categorized using transformed images from GASF, the EEG classification task was performed using several CNN architectures, such as AlexNet, VGG16, VGG19, feature-based Deep-ANN, and custom CNN. In their findings, all approaches generated less than 0.8 F1-scores, except for the custom CNN, which gave 0.9 F1-score and surpassed all methods tested [30].

There are also possibilities to use recurrent neural networks (RNNs), it is recommended in some articles [33], but there are no studies that apply RNN to EEG data using Gramian Angular Field coded images on this day.

### 1.11.1. Custom CNN

Here, the custom CNN architecture provided by the article "Implementation of Deep Neural Networks by Classifying EEG Signals Using a Gram Angular Summation Field for the Diagnosis of Epilepsy" [30] is described. This high-precision custom CNN consists of three convolutional layers with a ReLU activation function, followed by a Batch Normalization and a Max Pooling layer. Then, using Dense layers with a Sigmoid activation function, high-level features are obtained. Until the final classification layer consists of the SoftMax activation function for binary classification [30].




| Layer Name Kernel Size, # Filters | Input Size | Output Size | Stride | Activation Function |
|---|---|---|---|---|
| Input Image | [224,224, 3] | [224,224, 3] | - | - |
| ConvNet-1 [3x3], 32 | [224,224, 3] | [222, 222,32] | 1 | ReLU |
| ConvNet-2 [3x3], 64 | [222, 222,32] | [110, 110,64] | 2 | ReLU |
| ConvNet-3 [3x3], 64 | [110, 110,64] | [54, 54,64] | 2 | ReLU |
| Batch Norm | [54, 54,64] | [54, 54,64] | - | - |
| Max Pooling [2x2] | [54, 54,64] | [27, 27,64] | 1 | - |
| Flatten | [27, 27,64] | [46656,1] | - | - |
| Dense-1 | [46656,1] | [1024,1] | - | Sigmoid |
| Dense-2 | [1024,1] | [512,1] | - | Sigmoid |
| Dense-3 | [512,1] | [2,1] | - | SoftMax |
| Classifying Layer | [Normal, Focal] | | | |

**Fig. 1.6.** Deep learning blocks and configurations of the custom CNN architecture [30]

More informative parameter configurations and visual representation of blocks according to the analyzed study are shown in the **Fig. 1.6**.

### 1.11.2. Residual Neural Network

The performance of a neural network is heavily influenced by its depth. When the number of layers is too big, however, degradation occurs [41]. To overcome this problem, a ResNet residual learning framework was created [42]. The layers are reformed as learning residual functions with reference to the layer inputs; such networks are easier to optimize and can benefit from increasing depth [42].

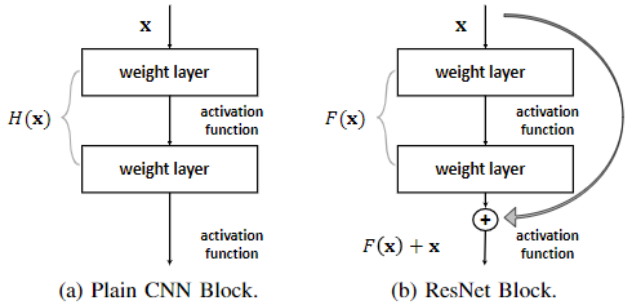



**Fig. 1.7.** Example of a simple CNN block and a ResNet block [41]

In the basic CNN block, layers are immediately fitted to the activation function $H(x)$. On the other hand, a ResNet block has a distinct residual learning objective $F(x) = H(x) - x$, which implies that it learns the residual of $x$ in $H(x)$ [41].

With ReLU and batch normalization blocks, the ResNet model is implemented by skipping connections on two or three layers. The key advantages of this network are that it is not computationally costly and that it does not suffer from performance degradation due to deep layers. Resnet has been tested using 18-layer and 50-layer networks for biomedical imaging, with ResNet-50 proving to be the most trustworthy [43].

## 1.12. EEG Data Augmentation

Deep learning methods require a large amount of training data to obtain sufficient accuracy. For this reason, the use of augmentation models is recommended if it is difficult or impossible to collect more data [44]. This challenge constantly occurs in EEG signal experiments because data collection is time-consuming and demanding on human subjects [45]. Signal data could be augmented using interpolation or mathematical modelling, but it was proven to have drawbacks such as information loss and suffering from artifacts [46]. Therefore, with the growing popularity of neural networks, Generative adversarial networks (GANs) were introduced as a suitable way to learn the statistical characteristics and generate artificial EEG signals [44][33][47]. Several modifications have been made to better adapt this framework for EEG signals. The study "Biosignal Data Augmentation: Based on Generative Adversarial Networks" showed that data augmentation based on RNN using hidden layers of long-term short-term memory (LSTM) is effective for time-series data [44]. Meanwhile, other studies have used the Wasserstein GANs (WGAN) [47] and the recurrent GANs (RGAN) models [33]. However, it is not known which modification is most suitable in general, but all papers suggest a similar conclusion that GAN significantly improved the performance of classification models using small data sets. More information on specific cases provided in **Table 1.4**.

**Table 1.4.** Summary of papers on EEG data augmentation

| Study | Goal | Augmentation method | Classifier | Improved accuracy |
|---|---|---|---|---|
| [48] | Classify Steady State Visual Evoked Potential (SSVEP) | Improved Wasserstein GAN (WGAN) | SSVEP Convolutional Unit (SCU) CNN | From 69% to 71% |
| | | Deep Convolutional GAN (DCGAN) | | From 69% to 72% |
| | | Variational Auto-Encoder (VAE) | | From 69% to 67% |
| [45] | Classify EEG Motor movement/imagery events | The recurrent generative adversarial network (RGAN) | Deep feed-forward neural network | From 54.3% to 93.4% |
| | | | SVM | From 59.7% to 72.5% |
| | | | Random forest tree (RFT) | From 67.1% to 79.7% |
| [46] | Reconstruction from low-sampling-sensitivity EEG (LSS-EEG) signals to high-sampling-sensitivity EEG (HSS-EEG) signals to improve performance of motor related datasets | WGAN and a temporal-spatial-frequency (TSF-MSE) loss function | Filter bank common spatial patterns (FBCSP) and SVM | From 65.1% to 68.5% |
| | | | CNN model FBCSPNet | From 68.8% to 70.3% |

The augmentation of EEG data can be done in the image space or directly in the signal space. Augmenting signal data consumes less computational resources, but this type of model is less researched than image augmentation. Moreover, EEG augmentation results are difficult to interpret even for professionals, they are noisier and more complex [49]. As can be seen from **Table 1.4**, RGAN performance far surpasses other methods. Adding 50% of the synthetically generated data increased the accuracy of neural network classification by 39.1% [45]. The article also compared Autoencoder and Variational Autoencoder augmentation models and found that RGAN also outperforms them by 34.8% and 19.9%, respectively [45]. This is due to the efficiency of this method and the ability to model repetitive signal patterns over time [45]. Nevertheless, almost all the augmentation models presented improved accuracy, despite being affected by a non-stationary EEG signal. Adding augmented data could help overcome the limitation of small EEG data size in various BCI applications.

### 1.12.1. RGAN Framework

Generative adversarial networks were introduced in 2014 [50] and since then received considerable attention in image generation, but limited work has been exploited to generate time series data [49]. The main characteristic of adversarial networks is that they consist of two agents that are trained simultaneously. They are called a discriminator (D) and a generator (G). The generator network randomly generates data with the same dimensions as the training data and learns to capture the real data distribution [50]. And the discriminator is trained to discriminate between real and generated data [50]. By competing, G tries to maximize the likelihood of D making a mistake [50]. As a result, the generator is increasingly generating more accurate data, but not the same data as training [44].

The two main advantages of GANs models are the ability to generate data with gradient values, and the representation of degenerate distributions [50]. In order to generate data for more than one distribution, the Conditional GAN model was presented [51]. Both the generator and the discriminator

can be conditioned by additional information $y$ such as class labels [51]. The objective function, where the discriminator tries to minimize the mean square error and the generator tries to minimize the log $(1 - D(G(z|y)))$, is described as follows [51]:

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[\log D(x|y)] + E_{z\sim p_z(z)}\big[\log(1 - D\big(G(z|y)\big))\big]; \quad (7)$$

here $p_{data}$ and $p_z$ are the distributions of real and generated data, respectively, while $x$ is an actual data sample and $z$ is a random noise.

The recurrent generative adversarial network (RGAN) has shown promising results in dealing with time-series, as recurrent neural network capture signal time dependencies [45]. In one study, the Recurrent Conditional GAN model was proposed to generate realistic medical multi-dimensional time series based on certain conditional inputs, such as class labels [49]. The RGAN corresponds to the architecture of a regular GAN, the only difference being that LSTM is used in the generator and discriminator [49]. Here, the discriminator tries to minimize the average negative cross-entropy between predictions and labels, and the generator minimizes the average negative cross-entropy between the discriminator predictions [49]. The RNN generator uses a different random seed for each epoch along with the additional input and classifies whether it is fake or real data [49].

### 1.13. Evaluation Metrics for Classification

Choosing proper evaluation metrics assists in the development of the optimal classifier. Metrics are used to assess the trained classifier's generalization ability and the overall quality. The evaluation metrics are used in the training, validation, and testing phases. In training and validation stages, the chosen loss function is used to optimize and tune the classifier. And in the testing stage, the chosen metrics are used to evaluate the classifier with the unseen data. The following sections discuss cross-validation technique and different binary classification metrics.

#### 1.13.1. Five Times Repeated 2-fold Cross-Validation

To compare different classifiers, it is standard to calculate their average performance over cross-validation iterations [52]. In $k$-fold cross-validation, the training set is randomly partitioned into $k$ smaller sets, then a model is trained using $k - 1$ subsamples and validated on the remaining part of the data [52]. This approach is repeated until all $k$ subsets have been used as validation sets. Cross-validation can be repeated $n$ times with different $k$-fold subsets to reduce the variance of the performance metric [53]. For classification tasks with limited amounts of data, $5 \times 2$ cross-validation is recommended [54]. This technique was proposed by T. G. Dietterich as a good technique to estimate the generalization error, and the variance of that error [55]. Although it is well known that when $k$ is small, less training data is used to learn the task, it is preferable to perform more iterative tests to assess the model's generalization ability [54]. To compare the performance of two models trained for 5 times and validated using 2-fold cross-validation technique, the author also proposed a powerful $5 \times 2$cv statistic test to detect algorithm differences [55]. The null hypothesis that models $A$ and $B$ have similar performance is measured by the $5 \times 2$cv $t$-test statistic as follows:

$$t = \frac{p_1^{(1)}}{\sqrt{(1/5)\sum_{i=1}^{5} s_i^2}} \quad (8)$$

Where the variables are calculated according to the following formulas:

$$p^{(1)} = p_A^{(1)} - p_B^{(1)};\ p^{(2)} = p_A^{(2)} - p_B^{(2)};\ \bar{p} = \frac{p^{(1)} + p^{(2)}}{2}\ ;\ s^2 = \left(p^{(1)} - \bar{p}\right)^2 + \left(p^{(2)} - \bar{p}\right)^2; \tag{9}$$

here, the performance of model $A$ for the first fold is $p_A^{(1)}$, and for the second fold is $p_A^{(2)}$. Similarly, this applies to model $B$, only with $B$ indices. Then, the differences in the performance measurements of the two models are $p^{(1)}$ and $p^{(2)}$. Meanwhile, $\bar{p}$ is the average and $s^2$ is the variance of the differences. The null hypothesis is rejected when the $p - value$ is less than $\alpha = 0.05$, then the models considered to be very different.

### 1.13.2. Confusion Matrix

The confusion matrix describes the classification efficiency of the model on a test data set. There are two aspects in this table: the actual value (the actual class of the object) and the predicted value (the value assigned by the classifier). When there are only two classes in a confusion matrix, these classes are called positive and negative. In this context, these four values in the table are also called: true positives (TP), false positives (FP), true negatives (TN), false negatives (FN) [56].

**Table 1.5.** Confusion matrix for positive and negative classes

| | | **Predicted Class** | |
|---|---|---|---|
| | | **Positive** | **Negative** |
| **Actual Class** | **Positive** | TP | FN |
| | **Negative** | FP | TN |

**Table 1.5** shows the expected classes on the x-axis and the actual classes on the y-axis. With these values in the matrix, some additional metrics can be obtained:

- **Accuracy**. It measures the ratio of correctly predicted observation to total observations. Accuracy is the most appropriate metric when data sets are balanced. It is calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{10}$$

- **Precision**. It measures the ratio of correctly predicted positive observations to total predicted positive observations. It is calculated as follows: [56]

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

- **Recall/Sensitivity**. It measures the ratio of correctly predicted positives to all observations in the actual class. It is calculated as follows: [56]

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

### 1.13.3. Receiver Operating Characteristic

The receiver operating characteristics (ROC) curve is a graphical-based metric that is considered to be superior to accuracy [57]. It is represented in a two-dimensional probability graph in which the TPR (sensitivity) represents the y-axis and FPR (1-specificity) is the x-axis [58]. The ROC curve is made by choosing different threshold values and afterwards recalculating the specificity and sensitivity [59]. These values are then shown on a graph, allowing for a more detailed analysis of the data [57]. The 45° diagonal line represent a random classifier, the higher the curve, the more perfect the classifier is (see **Fig. 1.8**). Many systems, including diagnostic systems, medical decision-making systems, and machine learning systems, have been evaluated using the ROC curve [58].

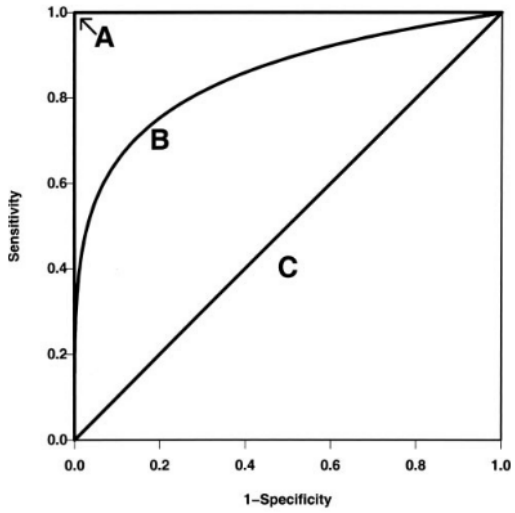



**Fig. 1.8.** ROC curves, where A is the perfect classifier, B - the normal, C - the random [59]

The area under the ROC curve (AUC) was developed as an overall summarizing metric of the ROC curve [59]. It represents a classifier's total ranking performance [57]. When the AUC equals 0.5, the ROC curve represents random chance, and 1.0 represents perfect accuracy [59]. The AUC value for a two-class problem can be determined as follows: [57]

$$AUC = \frac{S_p - n_p(n_n + 1)/2}{n_p n_n} \tag{13}$$

Here $S_p$ is the sum of all ranked positive samples, and $n_p$ and $n_n$ are the numbers of positive and negative samples, respectively [57].

## 1.14. Conclusions

The analyzed literature suggests many ways to predict visual stimuli based on brain signals. However, traditional methods suggest taking several test data from a single image to obtain an average signal value and thus eliminate noise in the data. Such an approach produces sufficiently good results, but it is not a suitable way to use for applications in real time. Data should be analyzed using only single trial EEG activity, signal processing should not take too long for results to be achieved quickly.

The visual stimulus studied in this research consists of upright and inverted faces. The inversion effect has been shown to affect response time and delay of ERP components. This is explained by the fact that the brain performs feature processing for inverted faces before understanding the image. EEG

data also have the peculiarity that the data are extremely sensitive to noise, with a higher noise ratio when participants are older or unable to maintain attention. Given that EEG data varies from individual to individual, this can lead to some problems in analyzing data from Alzheimer's disease patients. Therefore, preprocessing steps should be performed with caution.

To avoid mistakes due to lack of domain knowledge and bias in selecting the most influential features, it was proposed to analyze brain signals encoded into images. In fact, this approach ensures that by applying the data to the CNN architecture, the application speed will be more suited to the brain-computer interfaces. However, the use of deep learning architectures requires a larger data set for the method to optimize and extract relevant features. Unfortunately, EEG data often suffers from a lack of data because medical experiments take a long time to collect and then identical or similar experiments cannot be performed, so the data cannot be naturally supplemented from other resources. Therefore, to obtain sufficient accuracy, the RGAN augmentation method will be used directly from the signal domain to allow the algorithm to learn to generate synthetic data that should help improve accuracy and convergence rate. Then with data from healthy elderly individuals who observed visual stimuli and one patient with Alzheimer’s disease, it should be possible to predict the visual stimuli using electrode data from the areas of cognitive processes. The test dataset should contain data from one individual that was not previously seen in the training set to verify the subject independent results. The results of the control group and the Alzheimer's patient can then be compared in terms of visual processes' influence in the presence of different visual stimuli.

## 2. EEG Materials and Methods

This section describes all the design process. The "Research Data" section provides the data used in the study, the experiments performed to collect brain signals, and the results of the primary research. The following sections give an overview of the project as well as the design of potential systems and experiments. Then, the architectures of deep learning networks with all adjustments are described, such as will be used in the study for the augmentation and classification tasks. The section “Baseline Results using Raw EEG signals” discusses results from prior research using the same data for the classification of four classes of visual stimuli. Finally, the final chapters define the functional and non-functional requirements, as well as the quality criteria and development tools that were employed.

### 2.1. Research Data

Data for EEG signal analysis were taken from an open-source web interface "Figshare", which is used to manage and disseminate academic research data [60]. The data provided were used to study ERP neural correlates in an Alzheimer's disease (AD) patient [24]. Data were collected from healthy and AD patient by observing visual stimuli in three different experiments, details of which are described in the following sections.

#### 2.1.1. Subjects

The experiments were conducted between 2017 and 2018. All participants are females between the ages of 63 and 70. One of them is a 67-year-old Alzheimer's patient (initials MCG) with a facial recognition deficiency. She completed a separate test in which she failed to distinguish any of the famous faces, or the emotions expressed in the faces [24]. A total of eight healthy female volunteers without neurological abnormalities were also examined and used as a control group [24]. The findings may be influenced by numerous physical attributes or brain activity artifacts because the trials were conducted on older women who often have health issues at that age.

#### 2.1.2. Experimental Design and Visual Stimuli

In all experiments, seated participants looked at the monitor at 57 cm distance. On the monitor, visual stimuli were refreshed at a frequency of 85 Hz at a pixel resolution of 1280 × 1024. The experiments were constructed in three parts, where participants identified stimuli by clicking buttons 1000 ms after the stimulus onset. The electroencephalogram signal was continuously captured at a frequency of 1000 Hz with 64 electrodes placed according to the 10-10 international standard, then downsampled to 250 Hz. Each control participant performed the experiments on a single day and the AD patient on different days. The three experiments are detailed below [24]:

1. **Object recognition**: they were asked to distinguish the visible object from three categories: face, house, scrambled image [24]. See examples in **Fig. 2.1**.

**Fig. 2.1.** Example of experimental data (a face, a house, and a scrambled image)

**Face emotion detection**: they were asked to report a face inversion (upright or inverted) in the presence of different emotional visual stimuli (neutral or fearful expression). A total of 72 unique colorful images of male and female faces were presented [24]. See examples in **Fig. 2.2**.

**Fig. 2.2.** Example of experimental data (neutral and fearful emotions)

2. **Familiar and unfamiliar face recognition**: they were asked to identify a face inversion (upright or inverted) in the presence of different familiarity visual stimuli (famous or unknown). A total of 36 unique black-and-white male and female face photographs were used in this part. In a random order, each image was repeated four times [24]. See examples in **Fig. 2.3**.

**Fig. 2.3.** Example of experimental data (famous and non-famous people)

### 2.1.3. Experimental Results of the Original Study

The goal of the experiments described in the article was to see whether indicators of facial processing might be detected in Alzheimer's patient. The patient recognized the upright position substantially faster and more accurately in both second and third experiments. However, unlike the control group, MCG showed no inversion effect at the ERP level, supporting the holistic hypothesis of face perception whereby vertical face stimuli are processed as a whole whereas inverted stimuli are processed using feature-based processing. In addition, the N170 component, responsible for face processing, was not present in the EEG signal of an AD patient. Unfortunately, this visual processing was unaffected by emotional content as well. The negative component N400, on the other hand, shows the memory trace's recruitment [24].

## 2.2. Baseline Results using Raw EEG signals

To establish a baseline, a side study was conducted. Raw EEG signal data and well-performing deep learning algorithms including EEGNet, EEGNet SSVEP, and DeepConvNet are used in the study. L1 and L2 regularization were added to the best performing EEGNet SSVEP, which designed to classify the visual evoked potential. FIR bandwidths ranging from 4 to 40Hz were used to filter the data from this experiment. A time window of 200ms before and 800ms after stimulus onset was chosen. At the 150V peak-to-peak threshold, epochs were eliminated and normalized between -1 and 1. In order the deep-learning model to learn the most significant features on its own, 35 EEG channels from the back of the head were chosen to preserve a larger number of channels. Additionally, Variational Autoencoder was used to train the model to generate synthetic data to pre-train the EEGNet

architectures. The test dataset contains all data from one subject (the oldest) while the other test dataset contains signals from an Alzheimer's patient. For training and validation, the remaining data were combined and divided by 75/25 ratio in order to develop a subject-independent model [61].

A total of two experimental studies were performed. In the first, visual stimuli are divided into 4 classes of emotion and face inversion: neutral/upright, neutral/uspsidedown, fearful/upright, fearful/upsidedown. And in the second, divided into 4 familiarity and face inversion classes: famous/upright, famous/upsidedown, unfamous/upright, unfamous/upsidedown. This helps to distinguish the influence of different types of images presented and to investigate the influence of lower-level facial characteristics on brain signals [61].

The models were allowed to train for 500 epochs until the validation loss started declining in the last 20 epochs. This strategy helps to prevent overfitting. The training procedure was started with a batch size of 64 and an Adam learning optimizer with a rate of 0.001. The conclusions were made using the pre-trained model "EEGNet SSVEP with regularization" which results were evaluated using 5 times 2-fold cross validation technique. Finding reveal that identifying the emotion and face inversion effect resulted in the maximum accuracy of 50.2% for validation dataset. The accuracy of the oldest person test dataset was 32.75%, and Alzheimer's disease was 24.41%. In the second classification scenario, the same model classified familiarity and face inversion types with an accuracy of 43.25% in the validation dataset, 30.23% in the control group, and 27.72% in the Alzheimer's patient group. Both investigations reveal that the model attempts to learn the face inversion effect, but it does not demonstrate any patterns of emotion or familiarity learning [61].

The Alzheimer's patient was unable to appropriately label emotions and familiarity categories during the initial experiment [24], hence deep learning models were also unsuccessful. However, because the classification results showed some patterns in attempting to categorize the face inversion effect, and because a patient with Alzheimer's disease had difficulty differentiating face position, this task is examined further.

### 2.3. Project Scope

The project steps identified based on the literature used. To extract meaningful information about visual stimulus, single-trial event-related responses from the electrodes that are responsible for visual processing will be taken. The preprocessing steps will be performed as described in the **section 1.9**. ERP response will be encoded into an image using GAF method. This approach will help shorten classification time and allow CNN to learn features automatically. Therefore, decisions about the scope of attention from the EEG signal will be made quickly and will help overcome the biggest challenge - to make it available for real-time systems. Eventually, the accuracy of the test sets will be compared without and with RGAN augmented data to see if this could improve model performance. The final conclusions will be drawn on whether it is possible to predict the scope of attention of a patient with Alzheimer's disease from the occipital lobe EEG channels responsible for visual processing. The goal here is to predict the face inversion effect and it will be tested under two conditions. The first is in an environment with color images and an emotion factor. And the second is in an environment with gray images and the familiarity factor. The essential project steps for creating a visual stimulus prediction solution are briefly presented in **Fig. 2.4** and **Fig. 2.5**.


**Fig. 2.4.** Scope of the project

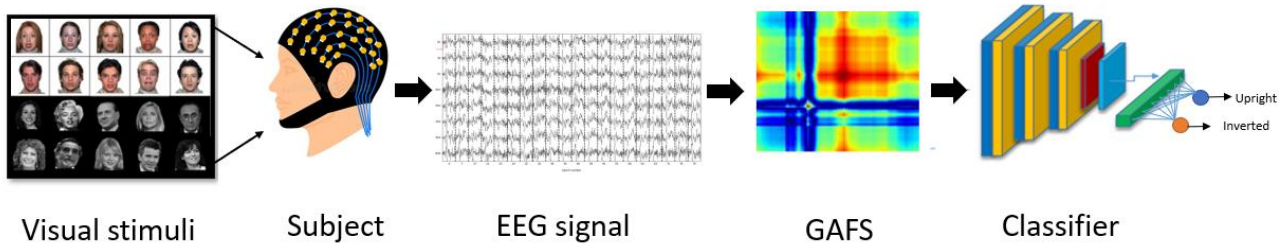



**Fig. 2.5.** The overall concept of the project

## 2.4. System Design

The user of the system can be any person with the required data of the electroencephalography signal. They run the model on the selected data segment and receive a response about the stimulus that is visible to the subject. Or the user can turn off the system in cases where the response time is too long. Use case diagram (see **Fig. 2.6**) describes the system from the user point of view. Meanwhile, the system will initiate the start of the process, perform pre-processing steps on the given signal data, encode epochs into images, and provide prediction of visual stimuli using a pretrained model.

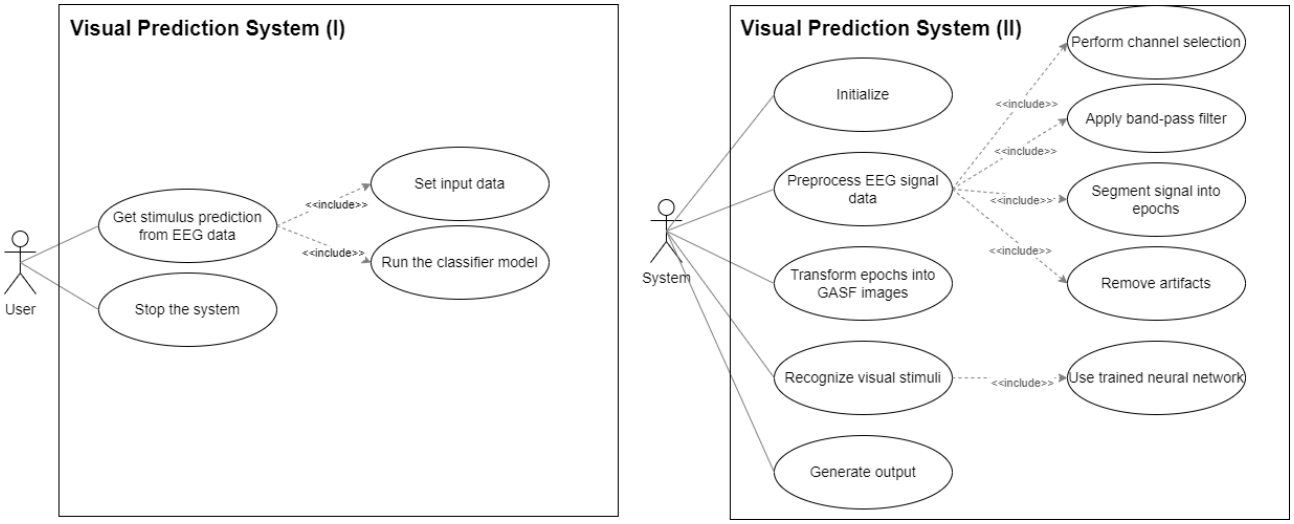



**Fig. 2.6.** Use case diagram

Activity diagram (see **Fig. 2.7**) describes sequences of actions in three swim lanes. Activities are grouped by who performs the action, whether it is a user, an interface, or an algorithm. Because the algorithm uses pre-trained neural network, its operation is faster, and the process is not complicated. Much of the complexity is training that model. In this case, the interface is the Python console. No exclusive user interface is being developed, given that it is not the system that is being built, but the potential performance of the model for future systems is being demonstrated.

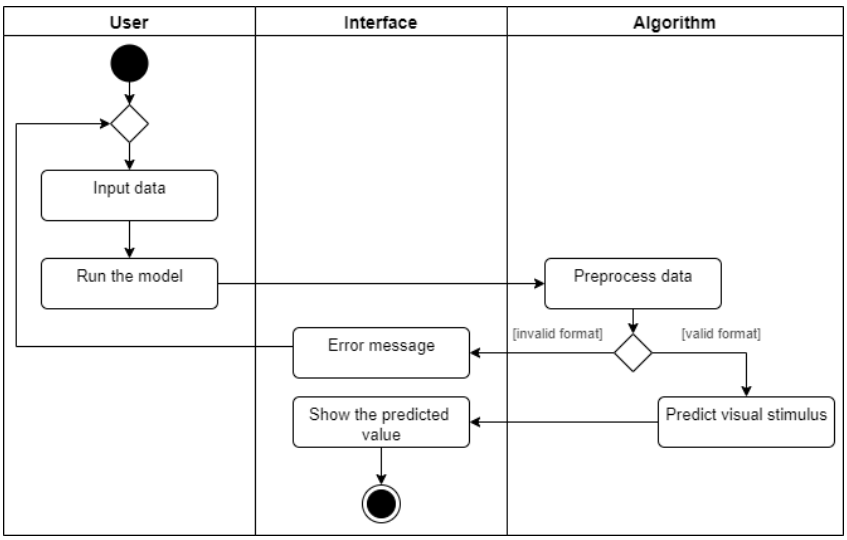



**Fig. 2.7.** System activity diagram

To predict a visual stimulus, the model must first be trained. Several experiments will be performed (as shown in **Fig. 2.8**) to select the most accurate model process.

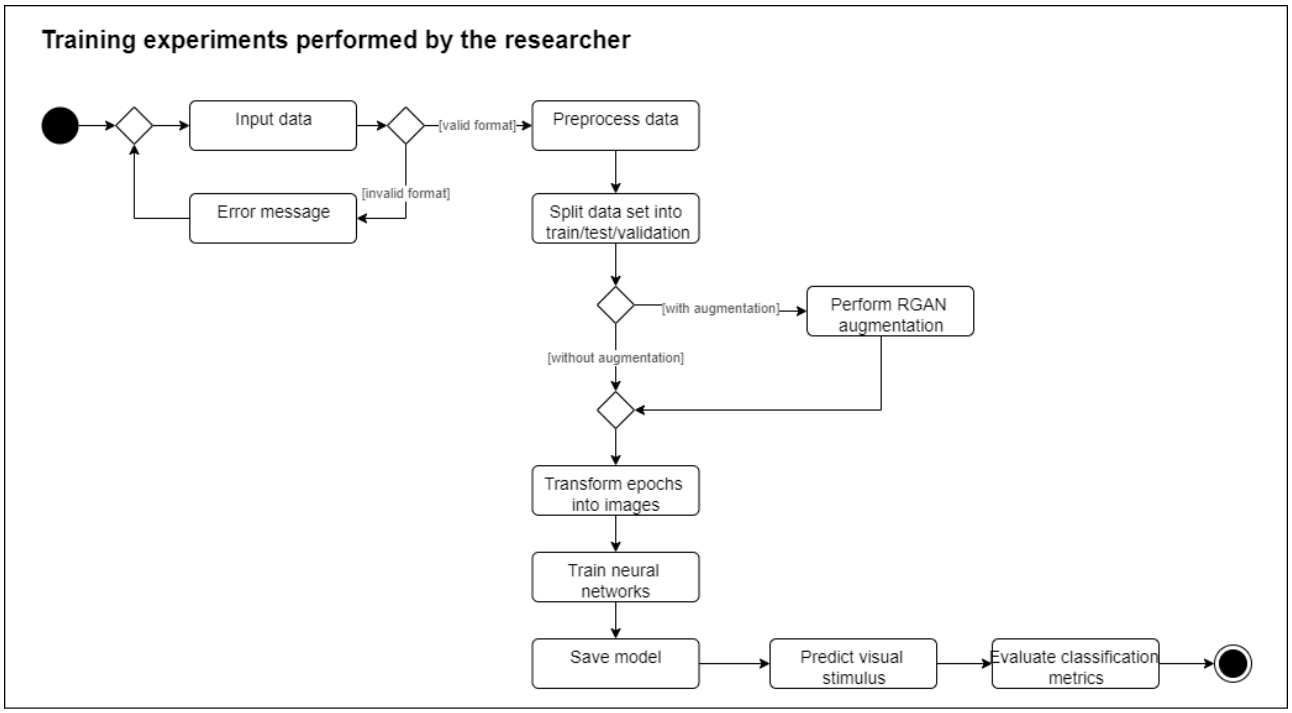



**Fig. 2.8.** Training experiments diagram

The goal of the research is to find the best classifier for identifying facial inversion stimuli. And then to make conclusions about the similarities between an Alzheimer's patient's brain region of visual processing and healthy persons of a comparable age. Because there were two sorts of photographs: grayscale face images and color face images, the same experimental procedure was used on both data sets, and findings were derived correspondingly. Initially, the data are pre-processed to obtain epochs corresponding to the given stimulus from the raw continuous EEG signals. The data set is then divided into train, test, and validation sets. The test dataset will consist of data from two subjects: the oldest subject and a patient with Alzheimer's disease. Such a set of test data guarantees user-independent

testing because the training data did not see examples of signals from these participants. Finally, the following process is performed with and without data augmentation performed by the RGAN model:

1) All selected EEG signal channels are converted to GASF images.
2) Selected and created neural networks are trained to find the best generalizing classifier.
3) The trained model is saved.
4) The visual stimuli of the test data are predicted by the trained model and the classification metrics are evaluated.

## 2.5. Network Structures

The system consists of two main neural network structures: one belongs to the data augmentation method RGAN, and the other to CNN for the classification of the visual stimulus class. Experimenting with different parameters and architectural layers is often a challenging task since it is compute-intensive and takes a significant amount of time. For these reasons, already assessed architectures and RGAN for multi-channel signal augmentation (**section 1.12.1**), and for training the classifier 4 different architectures has been tested: Custom CNN (**section 1.11.1**), ResNet (**section 1.11.2**), EEGNet SSVEP and a new proposed network. This section discusses all network structures used in the research work.

### 2.5.1. RGAN Framework

Vanilla GAN augmentation is not suitable for generating EEG data as it has features such as training instability and model collapsing. For these reasons, the RGAN framework is used. Looking at the bigger picture, the discriminator is trained to distinguish the real from the fake data (while the generator parameters are kept constant). And the generator is then trained to generate fake data that the discriminator can no longer distinguish as fake (while the discriminator parameters are constant). This process continues for multiple epochs while networks compete. Finally, the generator model is saved so that it can be reused to generate realistic data. The diagram of recurrent generative adversarial network suitable for EEG signals is shown in **Fig. 2.9**.

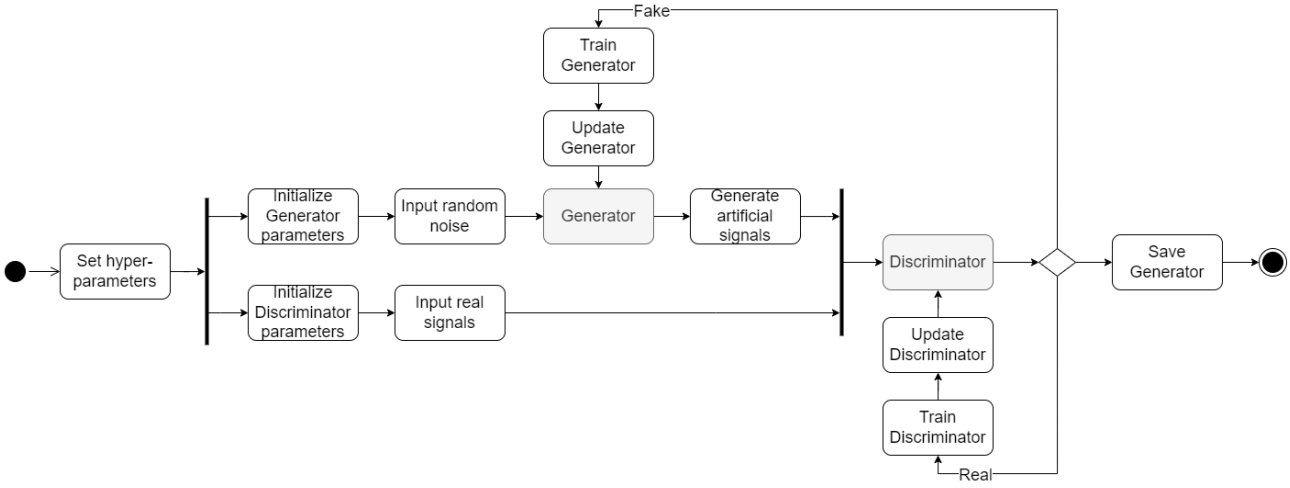



**Fig. 2.9**. RGAN model flowchart for EEG signals

A hyper-parameter is a parameter that directly affects the learning process. Its values are set once and do not update itself. Hyperparameter values can be an algorithm selection or parameters that affect speed and quality. The choice of these values is extremely important and have a significant impact

on the results [62]. The predefined hyperparameter values for EEG signal augmentation shown in the **Table 2.1**.

**Table 2.1.** RGAN hyperparameters

| Hyper-parameter | Description | Value |
|---|---|---|
| **Learning rate** | It is the step size to the minimum loss function. | 0.001 |
| **Batch size** | The number of samples to be propagated in the network. The larger this size, the more memory is required. | 32 |
| **Number of epochs** | The number of neural network cycles describing how many times the weights will be updated. | 1000 |
| **Loss function** | The loss function calculates the distance between the generated data distribution and the actual data distribution. Sigmoid cross entropy with logits is suitable for binary classification. | sigmoid cross entropy with logits |
| **Generator optimizer** | The optimization algorithm chosen is Adam (adaptive learning rate method) with momentum constants $\beta_1$ and $\beta_2$. | Adam |
| **Discriminator optimizer** | The optimization algorithm chosen is the gradient descent. | Gradient descent |
| **Random latent space dimension** | It is an embedding space, representing features in multi-dimensional space. | 16 |
| **Generator rounds** | The number of rounds of generator training. | 1 |
| **Discriminator rounds** | The number of rounds of discriminator training. | 3 |

Both the generator and the discriminator have separate network structures. The study [51] proposed this multi-channel network architecture for the generator and discriminator (see **Table 2.2** and **Table 2.3**).

**Table 2.2.** The Generator architecture

| Type | Activation | Options |
|---|---|---|
| Input | | $(z, \text{condition})$ |
| LSTM | tanh | $\text{num\_units} = 100$<br>$\text{state\_is\_tuple} = \text{True}$ |
| RNN | | $\text{sequence\_length} = [\text{time points}] * \text{batch size}$ |
| Dense | tanh | $\text{units} = 8$ |
| Output | | $(\text{batch size}, \text{time points}, \text{channels})$ |

**Table 2.3.** The Discriminator architecture

| Type | Activation | Options |
|---|---|---|
| Input | | $(\text{batch size}, \text{time points}, \text{channels})$ |
| LSTM | tanh | $\text{num\_units} = 100$<br>$\text{state\_is\_tuple} = \text{True}$ |
| RNN | | |
| Dense | sigmoid | $\text{units} = 1$ |

At each step, the generator takes a random seed and conditional embedding that encodes the visual stimulus classes. The input is passed to the LSTM layer with the hyperbolic tangent activation

function, followed by the RNN layer, followed by a fully connected layer reproducing the multi-channel EEG signal output. The discriminator architecture is very similar, except that the last fully connected layer is used with a sigmoid activation function that determines whether the signal is real or fake.

### 2.5.2. EEGNet SSVEP Architecture

The EEGNet SSVEP is a convolutional neural network for steady-state visual evoked potentials [63]. While observing visual stimuli, these potentials are recorded from the occipital lobe of the brain. This network was shown to be the top performer in a study using the identical EEG signal data [61]. As a result, rather of focusing on the time-series domain, the application of EEGNet SSVEP for imaged EEG data will be investigated. The network design is shown in **Table 2.4**, which uses temporal convolution in conjunction with depth wise spatial convolution and separable convolution. This way, the number of parameters is reduced. After all convolutional layers, a final dense layer with the SoftMax activation function is used for classification. Here in the **Table 2.4**, $C$ - number of channels, $T$ - number of time points, $F_1 = F_2 = 96$ - numbers of temporal and pointwise filters, $D = 1$ - number of spatial filters, $F_s$ - sampling rate, $N$ - number of classes.

**Table 2.4.** EEGNet SSVEP architecture [63]

| Layer | Type | Filters | Size | Pad | Activation | Options |
|---|---|---|---|---|---|---|
| | Input | | | | | input=(T, T, C) |
| **1** | Conv2D | $\boldsymbol{F_1}$ | (1, T/2) | same | none | |
| | Batch Normalization | | | | | |
| **2** | DepthwiseConv2D | | (C, 1) | valid | ELU | bias=False<br>depth multiplier=D<br>depth wise constraint=max norm (1.) |
| | Batch Normalization | | | | | |
| | AveragePooling2D | | (1, 4) | valid | | |
| | Dropout | | | | | rate=0.5 |
| **3** | SeparableConv2D | $\boldsymbol{F_2}$ | (1, 16) | same | ELU | bias=False |
| | Batch Normalization | | | | | |
| | AveragePooling2D | | (1, 8) | valid | | |
| | Dropout | | | | | rate=0.5 |
| **4** | Flatten | | | | | |
| **Classifier** | Dense | $\boldsymbol{N}$ | | | SoftMax | kernel constraint=max norm (0.25) |

### 2.5.3. Custom CNN Architecture

The custom CNN architecture, which was used specifically for Gramian Angular Summation Field transformed images, is used as a reliable classifier. This network has been described previously in **section 1.11.1**. Because the size of the input image is twice as small, some of the preset settings have been changed accordingly (see **Table 2.5**).

**Table 2.5.** Adjusted Custom CNN architecture, here T- time points, C- channels

| Layer | Type | Filters | Size | Strides | Activation | Options |
|---|---|---|---|---|---|---|
| | Input | | | | | input=(T, T, C) |
| 1 | Conv2D | 16 | (3, 3) | 1 | ReLU | |
| 2 | Conv2D | 32 | (3, 3) | 2 | ReLU | |
| 3 | Conv2D | 32 | (3, 3) | 2 | ReLU | |
| | MaxPooling2D | | (2, 2) | 1 | | |
| | Dropout | | | | | rate=0.25 |
| 4 | Flatten | | | | | |
| | Dense | 1024 | | | Sigmoid | |
| | Dense | 512 | | | Sigmoid | |
| **Classifier** | Dense | 2 | | | SoftMax | |

The main differences from the original network are that the size of the convolutional layer filter is reduced by 2 times. And the batch normalization layer was removed because the resulting training was unstable.

### 2.5.4. Proposed CNN Architecture

Knowing that EEG signal data are difficult to obtain in large quantities, the neural network created must be adapted to smaller data sets. Therefore, a network architecture that is suited for smaller sets and uses main generalization techniques are proposed in **Table 2.6**.

**Table 2.6.** Proposed CNN architecture. Here C – number of channels, T – number of time points.

| Layer | Type | Filters | Size | Activation | Options |
|---|---|---|---|---|---|
| | Input | | | | (T, T, C) |
| | GaussianNoise | | | | stddev= 0.1 |
| **1** | Conv2D | 2 | (5, 5) | ReLU | padding= “same” |
| **2** | Conv2D | 4 | (5, 5) | ReLU | padding= “same” |
| **3** | Conv2D | 8 | (5, 5) | ReLU | padding= “same” |
| **4** | Conv2D | 16 | (5, 5) | ReLU | padding= “same” |
| **5** | Conv2D | 32 | (5, 5) | ReLU | padding= “same” |
| | Flatten | | | | |
| **6** | Dense | 128 | | ReLU | kernel_regularizer= “l2” |
| | Dropout | | | | rate=0.25 |
| **7** | Dense | 128 | | ReLU | kernel_regularizer= “l2” |
| | Dropout | | | | rate=0.25 |
| **Classifier** | Dense | 2 | | Softmax | |

Using a smaller data size makes it more difficult for the neural network to learn the most relevant features, which might lead to overfitting. As a result, at the start of the architecture, Gaussian noise is injected to the input. Models trained with noise have often been proven to perform better on test

data and are less prone to overfit [64]. Including noise in the training phase allows CNN to learn more robust features that are unaffected by noise [65]. The selected noise distribution has a standard deviation of 0.1. Five two-dimensional convolutional layers with an increasing number of filter sizes and a $5 \times 5$ kernel size follow the input. All layers except the last dense layer employ the rectified linear unit's activation function. A L2 regularization penalty is applied to dense layers. The regularization strategy also contributes to the generalization and improvement of the model's performance [66]. Following dense layers, a $25\%$ dropout is employed to randomly remove the units and their connections from the neural network, preventing the units from over co-adapting [67]. Afterwards, a final dense layer with a SoftMax activation function classifies the input into two classes: 0 and 1 for vertical and inverted classes, respectively.

### 2.6. Functional and Non-functional Requirements

The program has the following functional requirements:

- The system processes (.set) and (.fdt) raw EEG data.
- The parameters of the trained model must be saved.
- A constant file must be created that can be easily adjusted to modify the experiments.
- The results of all experiments are collected and saved.
- Each part of the process must have a separate pipeline.
- A trained model provides an answer from a single epoch after visual stimulation of what the subject sees.

The program has the following non-functional requirements:

- The system response time is 10 second or less.
- The system is able to determine the visual stimulus more precise than the random classifier.
- Model training (retraining) time is limited to 12 hours.

### 2.7. Quality Criteria

Experimental EEG signal data containing visual stimulus events are used to train the neural network. To assess the quality of the model, it should meet the following criteria:

- The model should not be overfitted.
- A confusion matrix should be used to assess the performance of the model.
- The process should be automated in cases when the experiment with different parameters needs to be repeated.

It is important to emphasize that the proposed solution would be biased towards women and people of a certain age. While age bias should not be a major concern because Alzheimer's disease affects people aged 65 and up, it should be noted. However, the model may incorporate gender differences, which might be a concern with a different data set in the future.

### 2.8. Solution Development Tools

Software components are selected based on their growing popularity and personal experience. The requirements mainly consist of the fact that the programming language must be able to work with signal data and must be well adapted to deep learning. Hence, the following tools were selected:

- *Python programming language*. It includes libraries such as TensorFlow which supports the CUDA GPU and withstand high performance computing. This allows experimentations with GPU acceleration.
- *Pycharm.* It is a development environment that helps ensure code quality, supports debugging, and has Git integration.
- *Git.* This is a distributed version control system that is important to protect against the risk of losing code and maintaining the ability to revert to previous versions.
- *Google Colaboratory.* This is a Google product that enables the use of computing resources, including the GPU, in a cloud environment.

Python version 3.8 is used. To keep this version stable, *pipenv* library is used to create a virtual environment with required packages. This will avoid the risk that a newly released version of the package will corrupt the developed program. Python has many packages that can be used for different types of data. *NumPy* works with n-dimensional arrays while the backend code is well-optimized with C language. *Pandas* works with structured data and provides tabular results that are easy to read and use. Electroencephalography signals can also be analyzed using the *MNE* package, which provides state-of-the-art algorithms [68]. *Scipy* could be used for linear algebra and signal processing. And *Matplotlib* for visualizations.

## 3. Experimental Evaluation of EEG Signal Classification and Results

In this work, EEG signal data with classes of face inversion visual stimuli are analyzed. Initially, a pre-processing process is used, which includes data review, EEG channel selection, noise filtering, epoch-cutting, and artifact removal. The extracted epochs are then converted to images using the GAF technique. The raw EEG epochs are augmented using the RGAN method, and then the artificial signals are also transformed into the image space. Input images are supplied to the network input in the form of 8-channel images. Four different classifiers are trained, and performance is assessed using 5 time repeated 2-fold cross-validations. Each process is performed for both experiments in an attempt to classify the face inversion while seeing the color images (experiment 2) and the grayscale images (experiment 3).

### 3.1. Data Preprocessing

The section provides steps for pre-processing the data. The data is reviewed first. The accuracy of the Alzheimer's patient's response to determine the position of the visual stimulus is then analyzed. After reading the data, a channel selection from the occipital lobe is performed with data filtering to remove noise. The term “epoching” is introduced, which defines the process of cutting the epochs associated with an event from the raw continuous signal. The removal of artifacts is done in a way that removes epochs affected by physical movements. In the final step, event-related possibilities are explored in order to discover some visible patterns.

#### 3.1.1. Data Review

The second and third experiments were chosen for face features characteristic analysis after examining the experimental data. The second experiment included several emotion types, such as fear and neutral, as well as an inverted face effect, with all photos being in color. The third trial had photos in black and white with familiarity categories such as famous and unfamous. A patient with Alzheimer's disease has the initials “MCG”, the remaining participants are all in the same age range from 63 to 70. The information is labelled as "correct" or "incorrect" because the AD patient did not always reply accurately by identifying the face inversion, whether the face was inverted or upright. **Fig. 3.1** and **Fig. 3.2**. shows how different stimuli were arranged on the time axis during the experiment, the legend also shows the labels of each sample, and how many experimental results were correctly and incorrectly predicted by the patient. Here boundary events indicate discontinuity in the data when continuous datasets are concatenated.

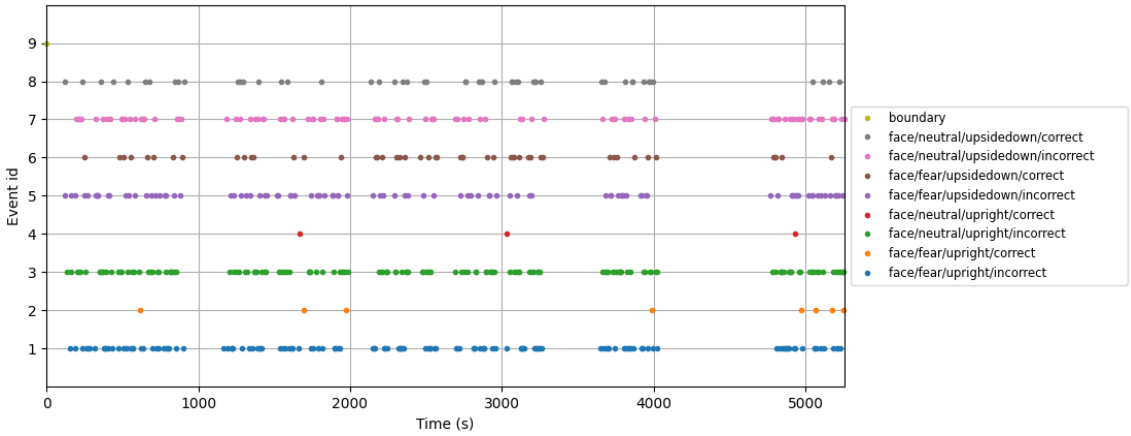



**Fig. 3.1.** Distribution of observed events in a patient with AD in the second experiment

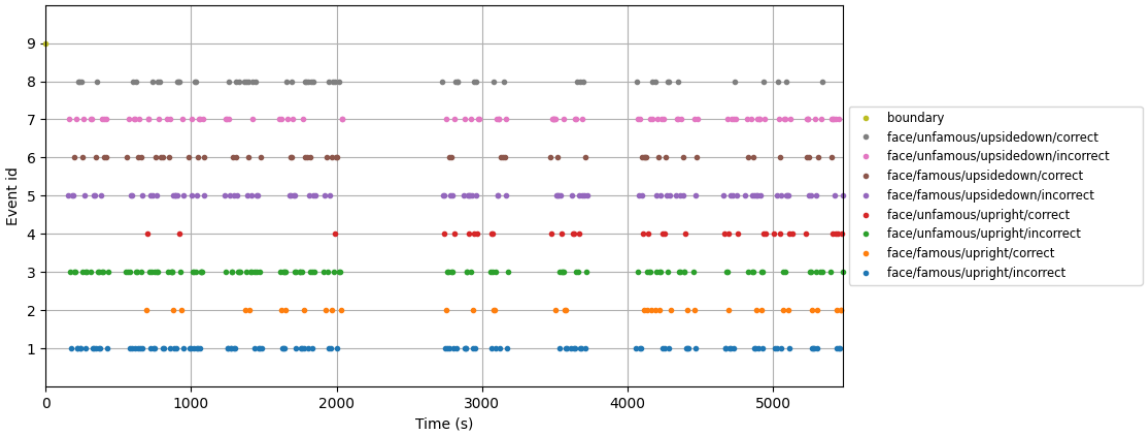
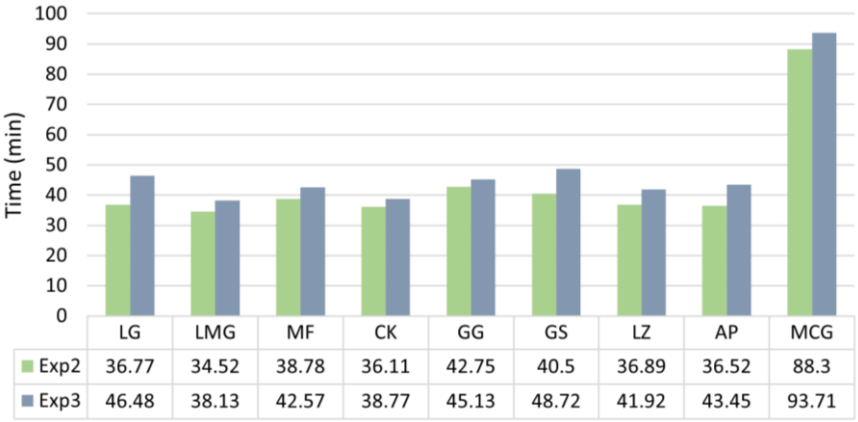



**Fig. 3.2.** Distribution of observed events in a patient with AD in the third experiment

Looking at **Fig. 3.3**, the experiments with the control subjects lasted on average about 38 minutes in the second experiment and about 43 minutes in the third. The experiment with grayscale images (experiment 3) reveals a tendency to take longer, which is a common aspect of grayscale image observation since the response time is on average longer. Meanwhile, experiments with a patient with Alzheimer's took more than twice as long to keep the patient from getting tired or distracted. Each AD patient experiment also took place on different days.




| | LG | LMG | MF | CK | GG | GS | LZ | AP | MCG |
|---|---|---|---|---|---|---|---|---|---|
| Exp2 | 36.77 | 34.52 | 38.78 | 36.11 | 42.75 | 40.5 | 36.89 | 36.52 | 88.3 |
| Exp3 | 46.48 | 38.13 | 42.57 | 38.77 | 45.13 | 48.72 | 41.92 | 43.45 | 93.71 |

**Fig. 3.3.** Time allocation for individuals during each experiment

### 3.1.2. Accuracy of Visual Stimulus Perception in an Alzheimer's Patient

The correctness of the responses of Alzheimer's patients in predicting visual stimuli in both experiments is shown in **Fig. 3.4**. All the upside-down classes received the most misclassifications. The upright photos were guessed inaccurately in 51% of the unfamous/upside-down classes and 39% of the famous/upside-down classes. Given that the images were black and white, darker hair color may blend into the background, and therefore make it difficult for an Alzheimer's patient to detect the face inversion. Under this scenario, the AD patient must perform feature processing rather than a holistic face processing. The inaccurate classification of inverted emotion images, on the other hand, was not far behind. Fear/upside-down stimulus is anticipated wrongly 34% of the time, whereas neutral/upside-down is predicted incorrectly 35% of the time. These results indicate that MCG has some difficulty in detecting an inverted face. Overall, most misclassifications were made when processing gray images with 31% incorrect responses. Meanwhile, there were 19% incorrect responses when processing color images.

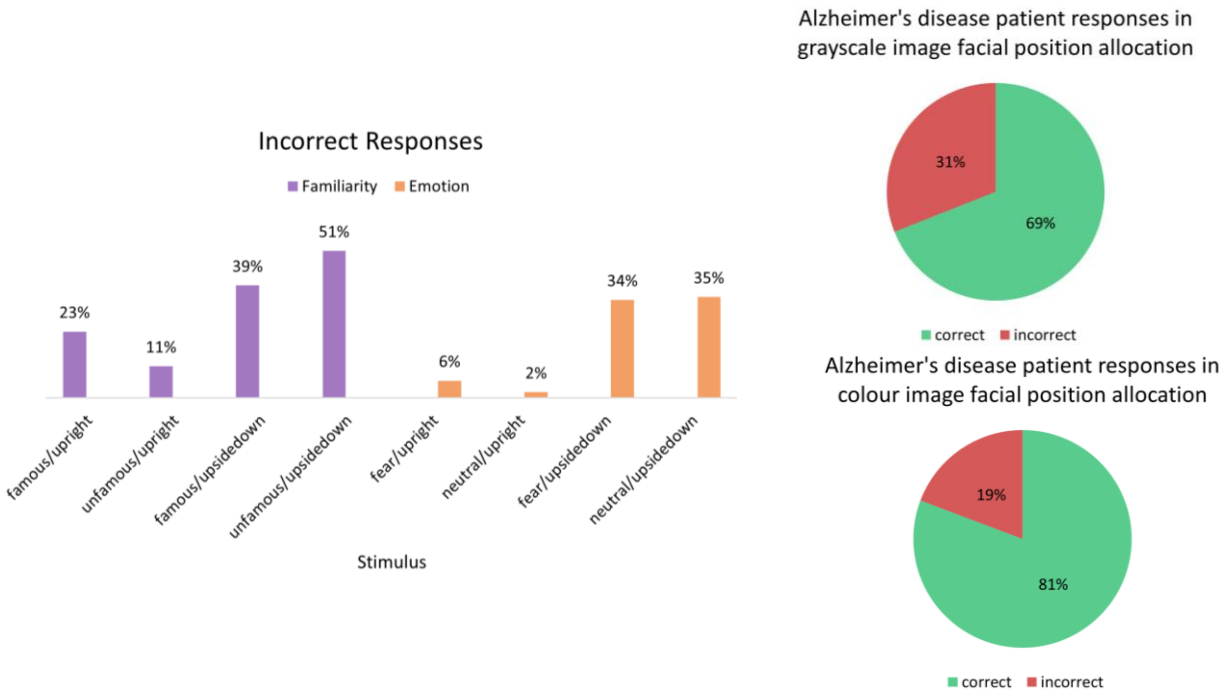



**Fig. 3.4.** Alzheimer's patient responses in both experiments

### 3.1.3. Channel Selection and Data Filtering

Only the most influential EEG channels for visual processing are selected. These channels are from the occipital lobe with indicators O and PO. Selected sensors are marked in black, and channels not used in this experiment are marked in red (see **Fig. 3.5**). Here, the positions of the EEG sensors are in three-dimensional space.

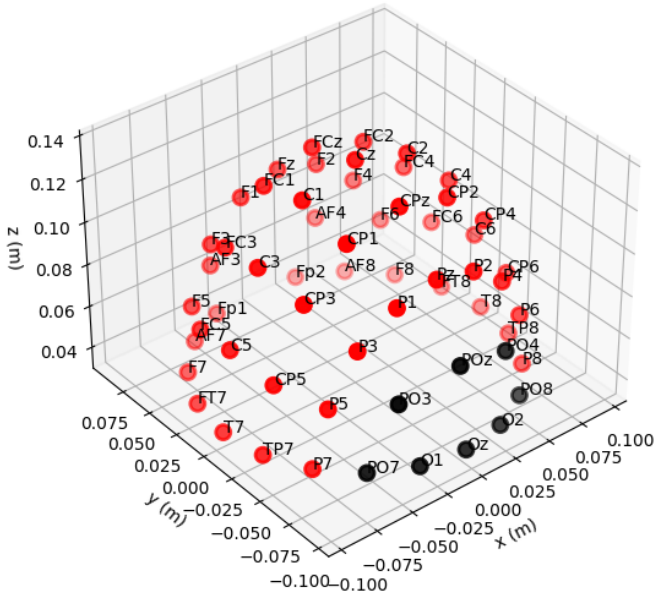



**Fig. 3.5.** EEG sensor positions

Filtering process is performed on continuous signal data rather than on each epoch to minimize filtering artifacts in signal boundaries. The most used zero-phase FIR band-pass filter with a high-pass of 1Hz and a low-pass of 30Hz is used to purify signal response and eliminate low and high frequency noise. The hamming window method is used in the FIR design, with a 6 dB cutoff

frequency and a filter order of 111. Breathing or movement artifacts can be removed with the right filtering procedure. The chosen frequency range includes delta, theta, alpha, and beta waves because theta and alpha waves have been proven to be most active in visual activities, meanwhile alpha and beta waves have been demonstrated to identify differences in different emotions. Additionally, a 1 Hz high-frequency filter was chosen since it has been found to be the most effective at removing noise and boosting classification accuracy. **Fig. 3.6** presents raw experimental data from a single control subject for the selected channels, with the onset of stimulus events marked at the top of the diagram.

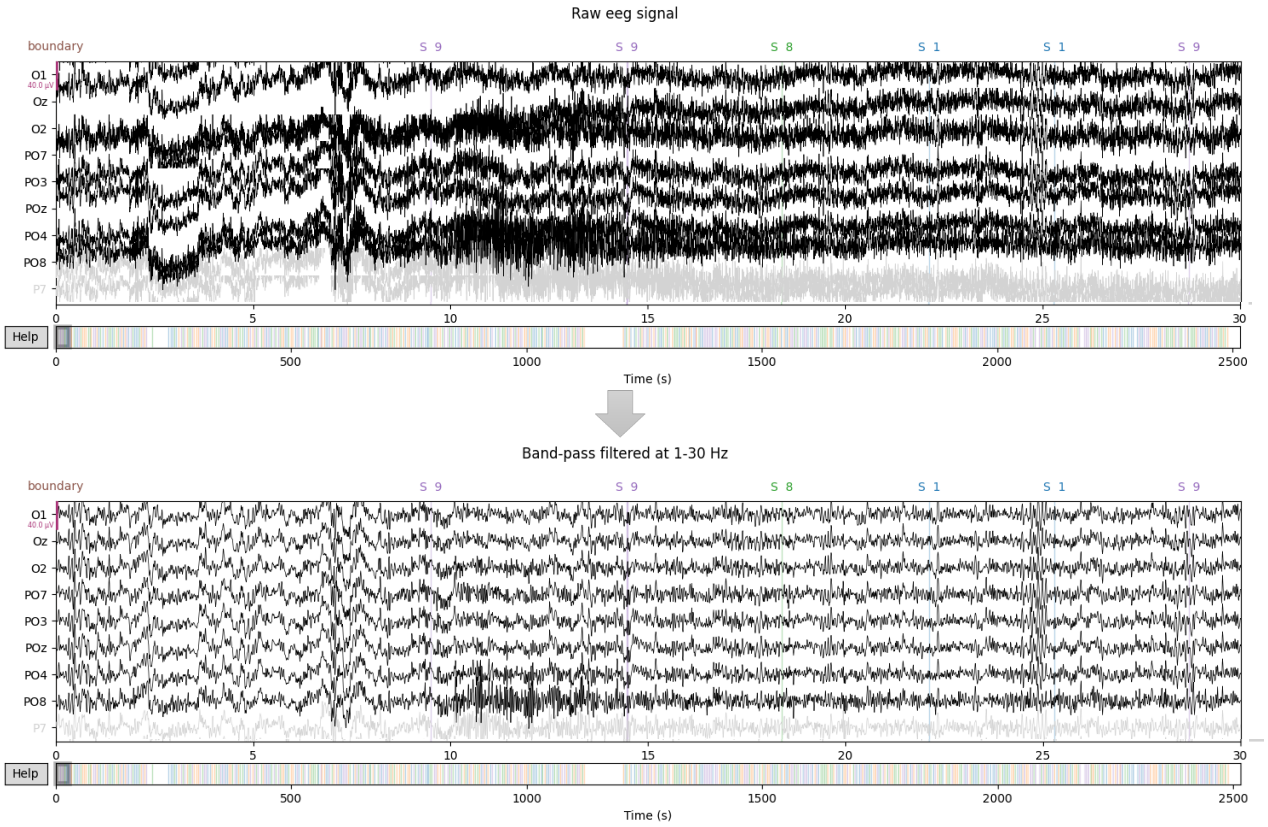



**Fig. 3.6.** Example of raw and filtered at 1-30Hz EEG signals

### 3.1.4. Epochs Segmentation and Artifact Removal

EEG “epoching” is the process of extracting the desired time windows around a specific event from a continuous EEG signal. Here, epochs are extracted 0.2 seconds before the visual stimulus and 0.8 seconds after it. Thus, one epoch consists of 1 second of signal with 250 time points. The baseline correction is done to each epoch and channel independently, right before the stimulus occurs, using a range of -0.2 to 0 seconds. The average baseline signal is calculated and subtracted from the whole epoch in this correction. This is done to remove temporal drift from a signal and increase the visibility of patterns in the EVP. **Fig. 3.7** shows an example of extracted epochs from experiment 2, when the visual stimuli are upside-down. The numbered epochs corresponding to the given stimulus are displayed on the x-axis, while the names of the 8 EEG channels are presented on the y-axis.

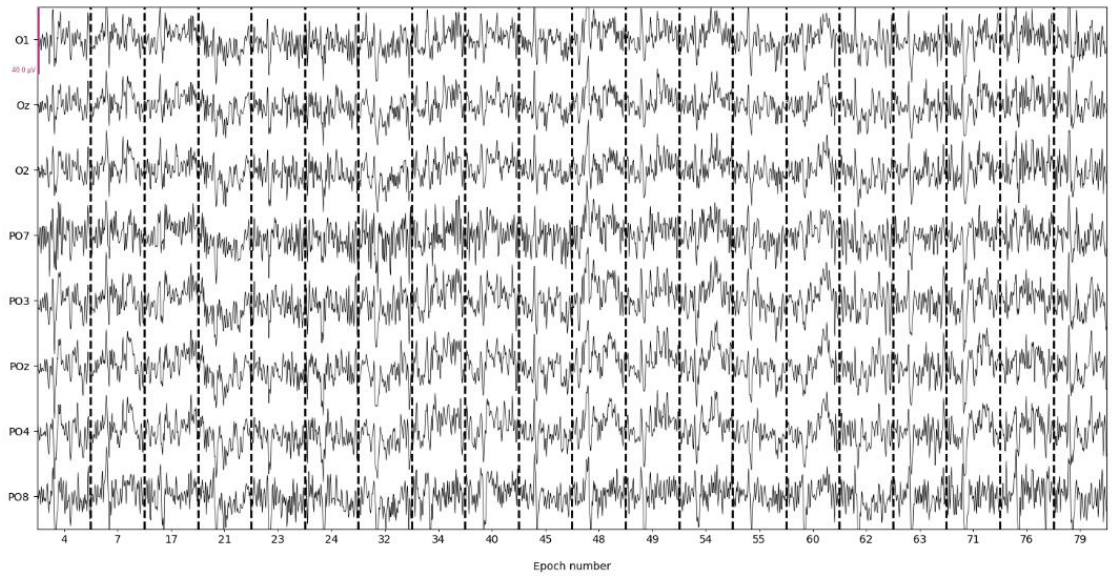



**Fig. 3.7.** Epochs of upside-down face stimuli (experiment 2, LG subject)

Having the epochs obtained, the next step is to determine if they contain artifacts of abnormal shape. This appears in the signal most often during blinking, eye, or face movement. If the artifacts are in the epoch, the whole epoch must be discarded. The easiest way to identify artifacts is to calculate peak-to-peak values. The limit chosen in the experiment is $100\mu V$, which means that the segment values can fluctuate $\pm 50\mu V$ and will not be discarded.

Looking at **Fig. 3.8**, the MCG patient had the most artifact-affected segments in both experiments. This seems logical, given that it is more difficult for an Alzheimer's patient to concentrate and maintain attention. The second highest place in terms of the most rejected artifacts is CK participant, she is the oldest of all subjects. In general, the number of dropped epochs per participant is less than 10%, which is a small part of the data sample, so no action is required to preserve or transform the compromised data. Thus, the data is considered pre-processed and ready to use.

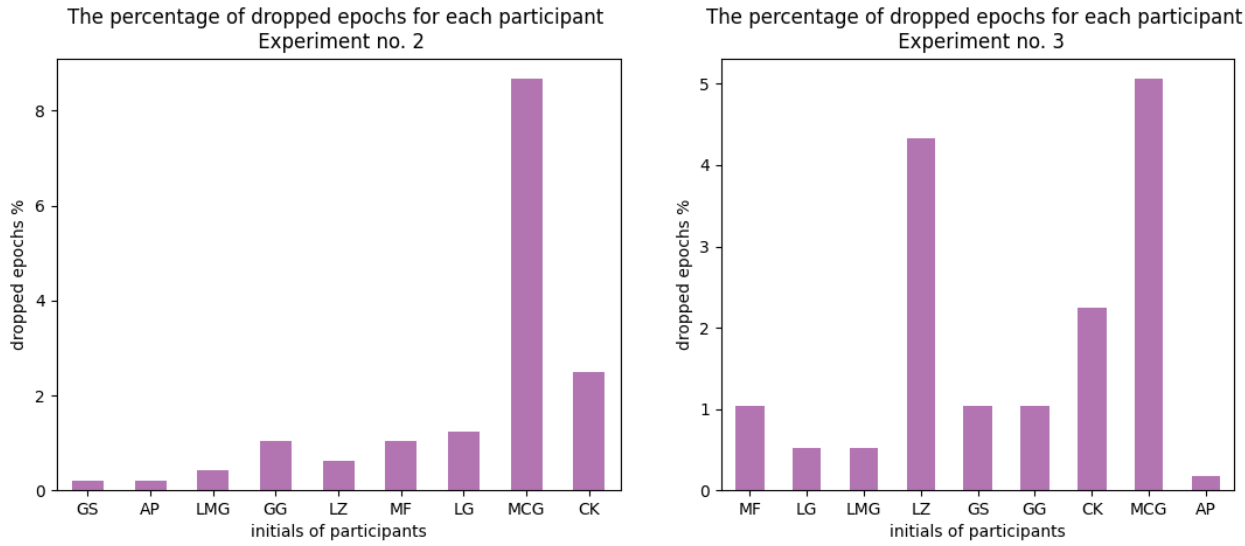
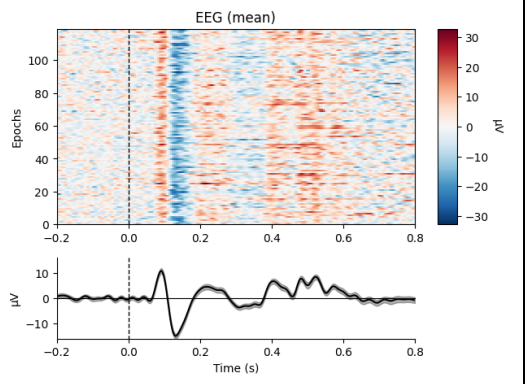
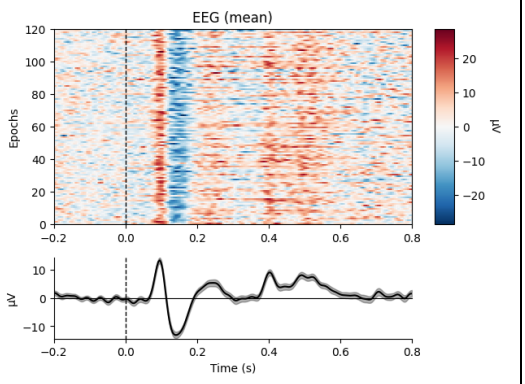
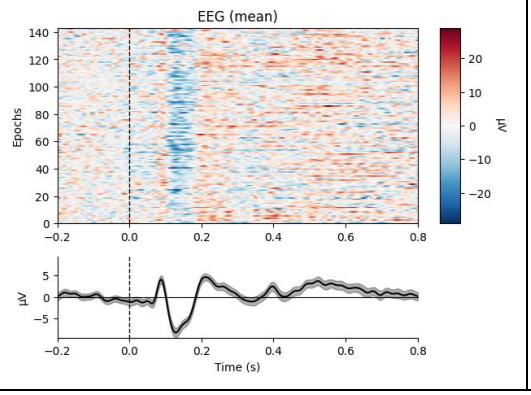
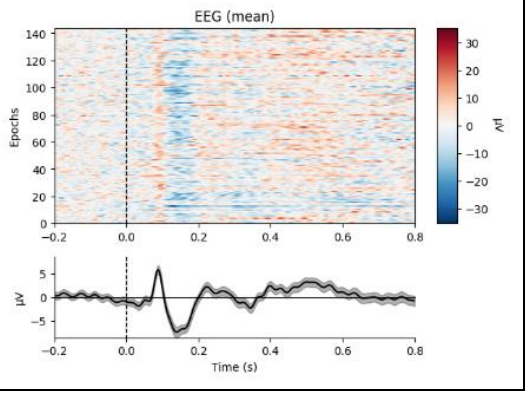



**Fig. 3.8.** Percentages of dropped epochs for each participant during experiments

### 3.1.5. Event-Related Potentials

Plotting event-related potential helps to investigate whether signal-specific epochs associated with specific stimuli have a certain pattern. This reduces noise and helps to better understand the various brain reactions. After rejecting bad epochs, all O and PO channels for each stimulus were averaged over epochs to obtain event-related potentials. In both **Table 3.1** and **Table 3.2**, the black line shows the mean ERP during the epochs, and the gray area shows the noise around the stimuli. A heat map showing the amplitude of the epochs is also shown, colored by the voltage strength of the electrodes. This allows to examine how frequently a comparable paternal appears in the presence of the same visual stimulus.

**Table 3.1.** Epochs and event-related field of control subject (LG) for experiment no 2

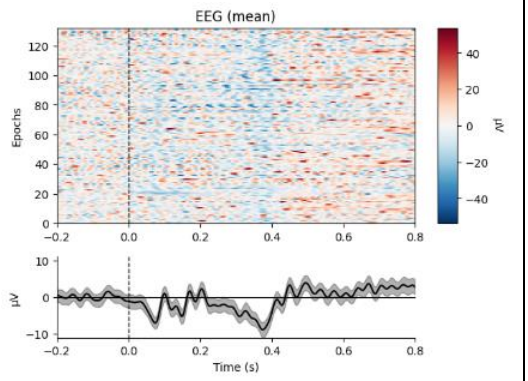
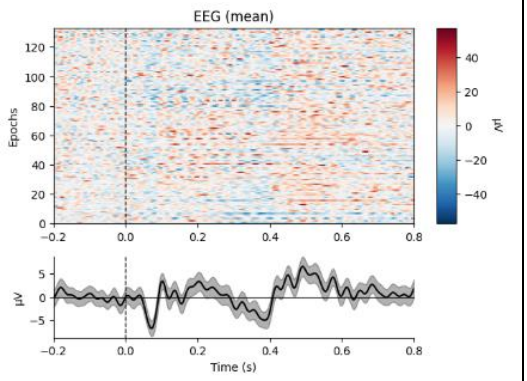
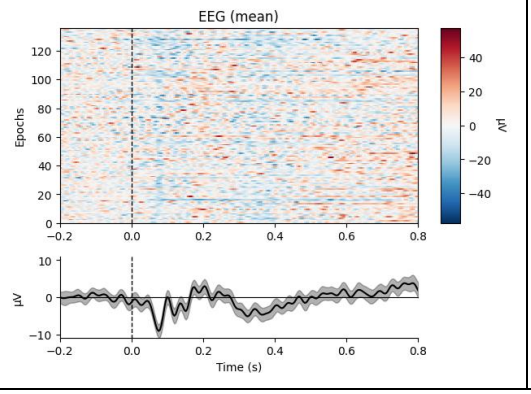
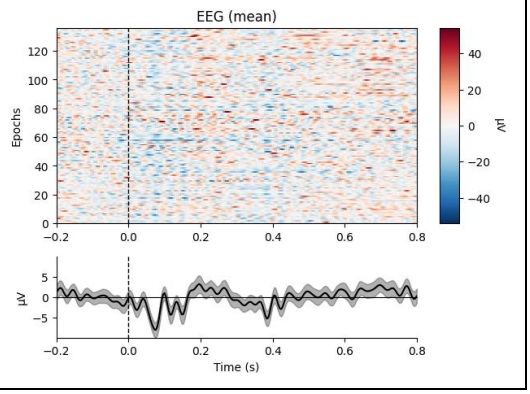
| | Exp. | Upright | Upside down |
|---|---|---|---|
| Control Subject (LG) | 2nd |  |  |
| | 3rd |  |  |



It can be seen that the control subject (LG) has a constant similar pattern in EEG signals by observing visual stimuli. As an indication of visible faces and face inversion, a negative peak increase of around 170ms after stimulus onset was seen. This N170 seemed greater and clearer in the second experiment than in the third experiment, in which gray pictures were observed. This might indicate that colorful faces are recorded more firmly in the EEG signal than black and white ones.

And as expected, there is more noise in the AD patient data and no clear paternal (see **Table 3.2**). Although, as demonstrated in a previous article [24], data from patients with Alzheimer's disease show a negative component N400. In the second experiment, this component is also more pronounced than in the third. When visually clear patterns cannot be observed, neural networks can demonstrate their potential by learning the properties of complex signals.

**Table 3.2.** Epochs and event-related field of AD patient (MCG) for experiment no 2

| | Exp. | Upright | Upside down |
|---|---|---|---|
| AD Patient (MCG) | 2nd |  |  |
| | 3rd |  |  |



## 3.2. Signal Transformation to Images

Preprocessed epochs in the form of time series are transformed into GASF images. This is done by conducting a polar encoding followed by a Gram Matrix to have spatial representation and retain the temporal dependency. At first, the values of all eight channels are normalized between 0 and 1. Because the epochs are truncated at one-second intervals, we have 250 time points, resulting in a final output image of $250 \times 250$ dimensions, with values ranging from $-1$ to 1. Examples of signals transformed into a Gramian angular summation field are shown in **Fig. 3.9**.

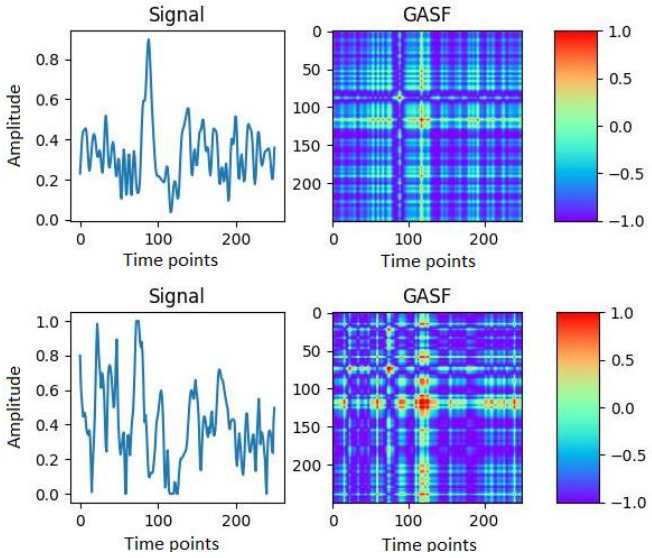



**Fig. 3.9.** Examples of signal transformation into GASF images

### 3.3. EEG Signal Data Augmentation

Since medical data often lacks large amounts of data, synthetic data augmentation can be very beneficial. Generative Adversarial Networks generates artificial data that looks real. Adding data can help solve problems of generalization and over-fitting. It should be emphasized that only the training dataset is augmented. In the GAN model, the job of the generator is to generate synthetic samples similar to the distribution of a training data set, and the task of the discriminator is to train oneself to better detect real ones from false samples. Both networks will compete until no one can improve any further. Because training for data augmentation can be very time demanding, augmentation is performed in preprocessed signal epochs rather than transformed images. To produce multi-channel EEG data, a conditional RGAN with stated architecture and hyperparameters was employed (see **section 1.12.1**). **Fig. 3.10** shows the average negative cross-entropy loss over training epochs. It is observed that the losses of both the discriminator and the generator converge with a small movement in the end as there are still attempts to improve. With less than 2000 samples in each class, the RGAN training durations for both studies were roughly 3 hours and 40 minutes. See **Fig. 3.11** for the real and generated synthetic signals.

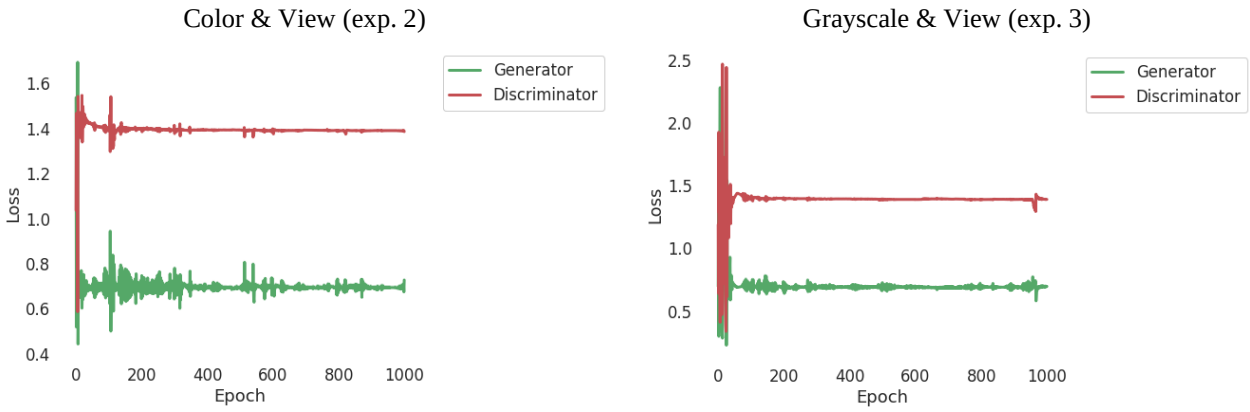



**Fig. 3.10.** RGAN performance for face inversion stimuli in both experiments

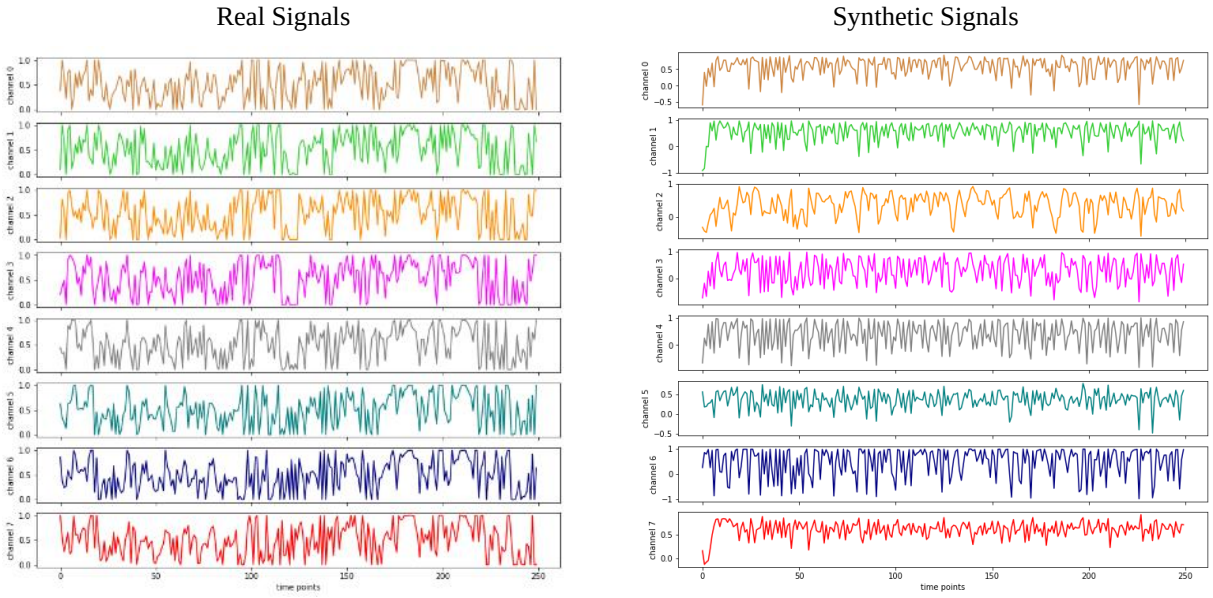



**Fig. 3.11.** Example of real and synthetic EEG signals for all 8 channels

In **Fig. 3.11**, the time points on the x-axis are denoted from 0 to 250 because there is an epoch of 1 second measured at 250 Hz. The y-axis contains different channels with normalized amplitude values. The convergence of the generator and discriminator networks, on the other hand, does not imply that the generated samples are similar to the actual data samples. It is critical to double-check the generated samples. This procedure, however, cannot be accomplished manually since the EEG signals are not intuitive.

Generator performance must be evaluated to determine when the generator produced the most realistic samples. This is accomplished using the novel idea proposed in an article "train on real, test on synthetic" [49]. The classifier model is trained on actual data and tested on artificial data that was generated after randomly selected epochs. To evaluate the model performance on different validation splits and gain a comprehensive view of the model noise, the evaluation was conducted using 5 times repeated 2-fold cross-validation, as suggested for smaller sets. For RGAN evaluation the classifier model EEGNet SSVEP was selected, as it was the best performing model in a previously conducted study [61]. The hyperparameters employed in this modeling method are listed in **Table 3.3**.

**Table 3.3.** EEGNet SSVEP classifier hyperparameters for RGAN testing

| Hyper-parameter | Value |
|---|---|
| Validation ratio | 0.25 |
| Batch size | 32 |
| Number of epochs | 1000 |
| Early stopping epochs | 20 |
| Learning rate | 0.0001 |
| Optimizer | Adam |

The validation data trained up to 90% on average for both studies, as shown by box plots (see **Fig. 3.12**) and averaged AUC metrics for randomly selected training points for the model reported in **Table 3.4**.

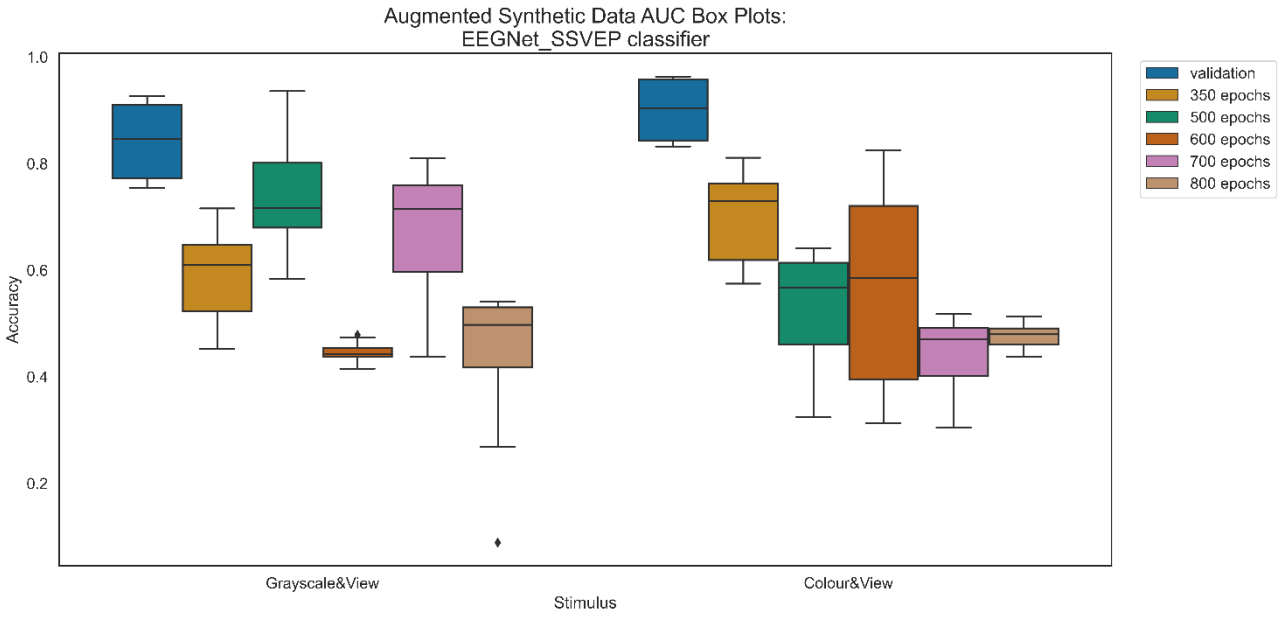



**Fig. 3.12.** Augmented synthetic data AUC box plots

**Table 3.4.** Averaged AUC metrics tested using synthetic data

| Stimulus | Run time (sec) | AUC - validation | AUC - 350 epochs | AUC - 500 epochs | AUC - 600 epochs | AUC - 700 epochs | AUC - 800 epochs |
|---|---|---|---|---|---|---|---|
| Color & View | 193.412 | 0.901 | 0.7 | 0.524 | 0.577 | 0.437 | 0.477 |
| Grayscale & View | 180.227 | 0.842 | 0.592 | 0.738 | 0.446 | 0.672 | 0.435 |

For color images the generator model scored the greatest AUC 70% after 350 epochs. And after 500 epochs, the generator model produced the most realistically generated data for grayscale pictures, with an AUC score of 73.8% on average. Even though the data came from mentally healthy people, the signals had a lot of noise and artifacts, which led to the data being rejected in previous steps. As a result, it is not surprising that synthetic data produces up to 74% accuracy. Consequently, the artificial data generated by these selected generator models will augment 25% of the training data.

### 3.4. Visual Stimuli Classification Results

At the time when raw EEG signals and artificial EEG signals are converted to images using the GAF technique, classification of visual stimuli can be performed. Because there are 250 time points in each epoch, the data generated is also $250 \times 250$ dimensional. However, larger images result in a lot of training parameters and require a large amount of GPU memory, so images are downsampled by applying a "mean" function to local blocks by factor 2. This results in images of the size $125 \times 125$. The input of size $(125,125,8)$ is then passed to different classifiers with the same hyperparameters given in **Table 3.5**. To allow objective interpretation of the models, the training is repeated 5 times and verified using the 2-fold cross-validation technique, which considerably assesses generalization ability, especially for small data. This assures that the classification results are unaffected by the training and test data selections. Meanwhile, the findings are compared to data from Alzheimer's patient and one control group participant who is the oldest. This testing separation ensures that the subject-independent results are evaluated, and that the model is never exposed to any signal from these individuals. Testing on healthy individuals reveals the performance for control group participants, whereas testing on Alzheimer's patient reveals the prediction outcomes for a particular patient, allowing researchers to compare healthy and not-so-healthy visual processing in the brain.

Google Colab is used to train classification models since it executes Python code on Google's servers and connects with Google Drive to retrieve data and save results. The Python version is 3.7, while the TensorFlow version 2.8.0 used with GPU support. Collaboratory utilizes NVIDIA-SMI 460.32 with CUDA version 11.2. As defined in **Table 3.5**, the models are trained with a maximum of 1000 epochs and an initial learning rate of 0.0001 using the adaptive moment estimation algorithm called Adam. This approach is one of the most effective, requiring little memory. To avoid overfitting, the model training is terminated if the validation loss stops decreasing over the past 20 epochs. Because 2-fold validation is performed, the validation ratio is 0.5, which means the model is trained on half of the data and verified on the other half. And to recalculate the gradients, batches of 32 samples are employed to preserve memory space.

**Table 3.5.** Visual stimuli classifier training hyperparameters

| Hyper-parameter | Description | Value |
|---|---|---|
| Validation ratio | The ratio of training data that should be kept aside for validation. | 0.5 |
| Batch size | The number of samples to be propagated in the network. The larger this size, the more memory is required. | 32 |
| Number of epochs | The number of neural network cycles describing how many times the weights will be updated. | 1000 |
| Early stopping epochs | Stops training when a loss function has stopped improving for the defined number of epochs. | 20 |
| Learning rate | It is the step size to the minimum loss function. | 0.0001 |
| Optimizer | The optimization algorithm. | Adam $beta_1 = 0.9$, $beta_2 = 0.999$ |

### 3.4.1. Classification of Facial Inversion in Color Images

Imbalanced training data when the number of samples for each class is unequal can lead to falsely excellent results. As a result, before training any model, it is critical to examine the data distribution (see **Fig. 3.13**). The figure on the left shows the number of epochs in real data for both upright and inverted stimuli. The figure on the right shows the distribution of real data combined with 25% of the augmented data, colored differently for training and two test data sets. We can observe that both stimulus classes have almost the same sample size, indicating that this data set is balanced as a minor variation makes minimal impact.

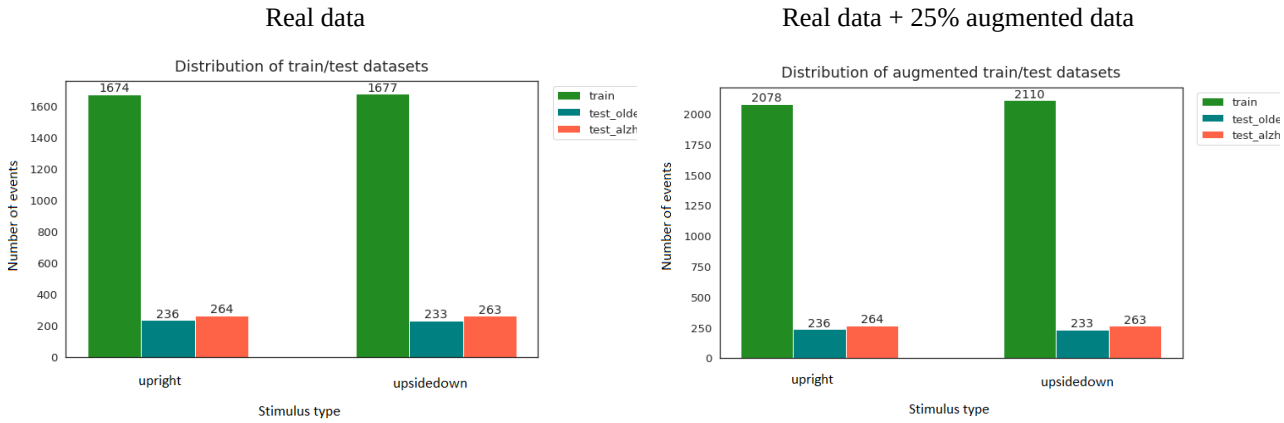



**Fig. 3.13.** Distribution of training and testing data with real and augmented data in color images

When both classes are equally balanced, accuracy is the best metric for measuring performance. However, since the amount of data in each class is slightly varies, we will measure two key metrics in order to obtain robust estimates: accuracy and AUC (area under the curve). The higher the score for both selected metrics, the better the classifier. AUC assesses performance across all classification thresholds, whereas accuracy provides an overall assessment of the classifier's correctness.

The findings of 5 times repeated 2-fold cross validation are summarized in the boxplot diagrams (see **Fig. 3.14**). Boxplots are used to compare distributions between different classifiers to determine how

data values are skewed, with whiskers values showing the minimum and maximum values observed, and a black line in the interquartile range indicating a typical median value.

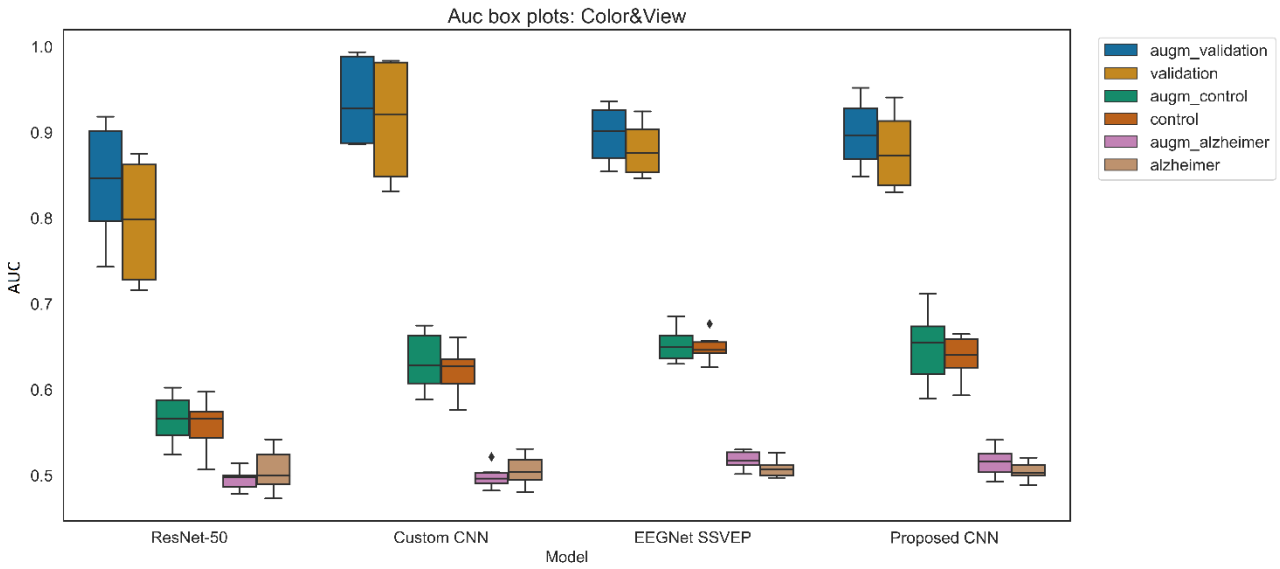



**Fig. 3.14.** Boxplots for estimating the AUC of facial inversion classes in color images

Comparing the results with raw and supplemented data, it can be observed that the augmented data helped to achieve better results. Since 10 different evaluations were generated using 2 cross-validation for 5 iterations, the value of the $t-test$ can be calculated as described in **section 1.13.1.** The calculated results for the validation AUC metrics to test the null hypothesis that the models being compared are similar are presented in **Table 3.6**. In this table, the $t-statistic$ values and the two-tailed $p-values$ are calculated, and if $p-value$ is less than 0.05, then the null hypothesis is rejected, meaning that the models are very different. The performed augmentation was statistically significant only for the ResNet-50 model, meanwhile the result for Custom CNN and Proposed CNN was weakly significant.

**Table 3.6.** 5x2 cross-validation statistics comparison for networks trained to recognize face inversion in color images on raw data and on augmented data.

| Network name | t | p-value | Is significant |
|---|---|---|---|
| **Proposed CNN** | 1.666 | 0.157 | False |
| **Custom CNN** | 1.993 | 0.103 | False |
| **EEGNet SSVEP** | 1.350 | 0.235 | False |
| **ResNet-50** | 3.967 | 0.011 | True |

Looking at the average evaluation metrics (see **Table 3.7**) for all architectures tested, it can be seen that the injected synthetic data slightly improved the estimated metrics. On average, training accuracy increased from 2.4% to 4.6%, according to validation data. However, the accuracy growth was smaller for testing control subject data, with an averaged accuracy decrease of 0.2% for EEGNet SSVEP and an increase of 1.3%, 1.1%, 1.1% for Custom CNN, Proposed CNN and ResNet-50, accordingly.

**Table 3.7.** The average evaluation metrics for all tested architectures with 5 times repeated 2-fold cross validation of facial inversion classes in color images. The results are sorted by the control subject's accuracy.

| Model | Run time (sec) | Validation accuracy | Control accuracy | Alzheimer accuracy | Validation AUC | Control AUC | Alzheimer AUC |
|---|---|---|---|---|---|---|---|
| Proposed CNN augmented | 150.93 | 0.853 | 0.617 | 0.51 | 0.9 | 0.648 | 0.515 |
| EEGNet SSVEP | 223.046 | 0.804 | 0.606 | 0.496 | 0.88 | 0.649 | 0.508 |
| Proposed CNN | 128.713 | 0.829 | 0.606 | 0.509 | 0.879 | 0.638 | 0.505 |
| EEGNet SSVEP augmented | 206.866 | 0.832 | 0.604 | 0.512 | 0.899 | 0.652 | 0.517 |
| Custom CNN augmented | 66.808 | 0.9 | 0.594 | 0.502 | 0.937 | 0.633 | 0.498 |
| Custom CNN | 54.12 | 0.876 | 0.581 | 0.501 | 0.914 | 0.622 | 0.505 |
| ResNet-50 augmented | 138.252 | 0.805 | 0.554 | 0.494 | 0.844 | 0.566 | 0.495 |
| ResNet-50 | 144.709 | 0.759 | 0.543 | 0.514 | 0.796 | 0.56 | 0.506 |

In **Table 3.7**, the values are sorted according to the highest accuracy for the data of the control subject. This shows that the best model for this particular task was our Proposed CNN, when the data were supplemented with artificial samples. Our classifier for the control subject was 1.1% more accurate than the second-best method EEGNet SSVEP in classifying the visible facial position.

The following are the results of the best-performing model. This was the Proposed CNN network's third training phase as it achieved the highest accuracy for the control subject - 67.59% (AUC = 71.19%) and the accuracy of the visual identification for an Alzheimer's patient - 53.32% (AUC = 53.53%). **Fig. 3.15** shows the training history of the proposed CNN model for the 50 epochs before it was stopped by the callback to avoid overfitting. The training and validation accuracies are shown on the left, while the training and validation binary cross entropy losses are shown on the right. The training and validation loss curves appear to behave similarly as one increases and the other increases as well.


**Fig. 3.15.** The proposed CNN model's training history in color images

At the end of the training, the model achieves up to 100% training accuracy and 82.23% validation accuracy. This means that the model can predict unseen data with 82.23% accuracy, but it has potentially begun to adapt to noise to achieve greater training accuracy. Hyperparameter optimization may be performed in the future to increase training performance.

The receiver operating characteristic curves for a control group subject and Alzheimer's patient data are shown in the **Fig. 3.16**. The 1-specificity (false positive rate) is represented on the x-axis, while the sensitivity (true positive rate) is plotted on the y-axis, using various thresholds. Images indicate that the model is moderate (AUC = 0.75) for the control subject, however the model is only slightly above the random for Alzheimer's visual classification (AUC = 0.55).

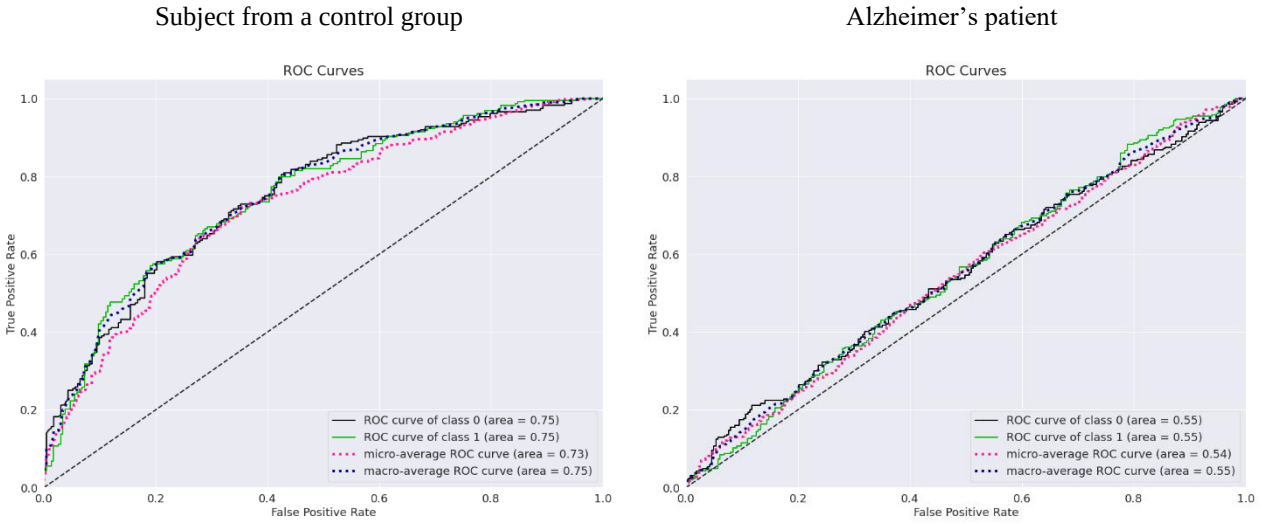



**Fig. 3.16.** The ROC curves for a control group subject and Alzheimer's patient using the Proposed CNN model in classifying facial inversion in color images

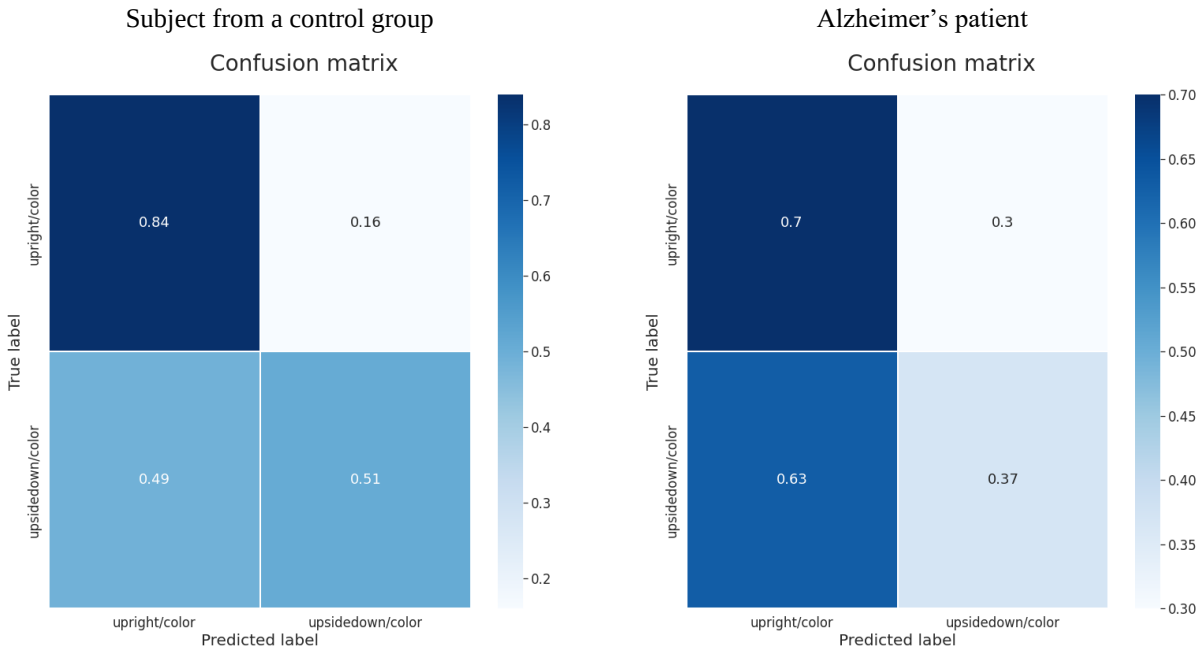



**Fig. 3.17.** Confusion matrices for a control group subject and Alzheimer's patient using the Proposed CNN model in classifying facial inversion in color images

The horizontally normalized confusion 2 × 2 matrices (see **Fig. 3.7**) provide an overview of correctly classified and misclassified data using the Proposed CNN model. Here, on the x-axis are predicted classes, while on the y-axis are ground truth classes. The aspiration to be seen in the figures is the greater true positive and true negative numbers in the diagonal, indicating that the model correctly predicted an upright or inverted visual stimulus. For a control group subject, the upright facial position is classified with less falsely predicted upside-down classes. The same pattern is present in Alzheimer's upright stimuli prediction.

### 3.4.2. Classification of Facial Inversion in Grayscale Images

First, looking at the distribution of visual stimuli in the training and testing datasets (see **Fig. 3.18**), we make sure that the datasets are evenly distributed among the classes listed on the x-axis. As the samples differ only slightly, the data set is considered to be balanced. Therefore, the same accuracy and AUC metrics continue to be used.

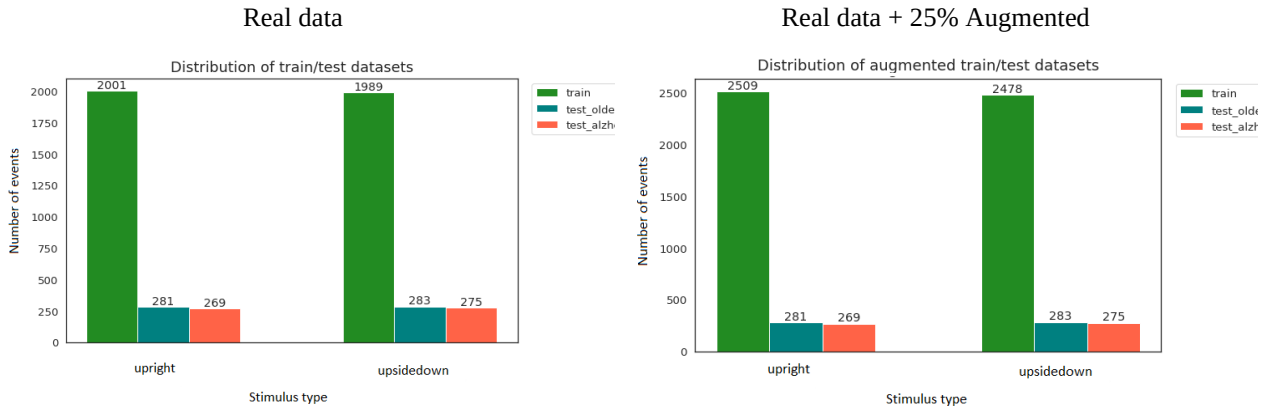



**Fig. 3.18**. Distribution of training and testing data with real and augmented data in gray images

The results of 5 training iterations using the 2-fold cross-validation technique are shown in the box plots (see **Fig. 3.19**) for the validation set, control's subject set, and Alzheimer's set. Because there are multiple randomized events throughout the training process, such as random training set selection or a dropout layer, the same model will train slightly differently each time.

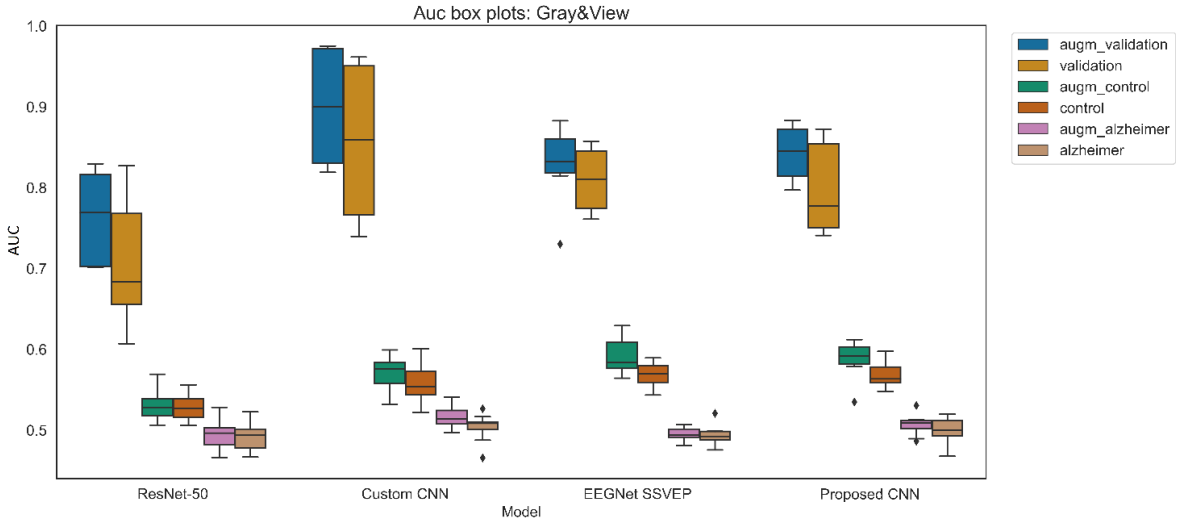



**Fig. 3.19.** Boxplots for estimating the AUC of facial inversion classes in grayscale images

As observed in the box plots (see **Fig. 3.19**), classifiers with supplemented data achieved higher validation AUC values than those trained with raw data alone. To determine whether this artificial signal generation process using the RGAN network was statistically significant, t-test statistics were calculated. **Table 3.6** shows the statistics obtained to assess the differences in validation AUC performance between the two models. The 5% alpha value was used to evaluate whether the models differ significantly. However, since all p-values are larger, the null hypothesis is not rejected meaning that the classifications used for supplemented and raw data are similar. On the other hand, the proposed CNN and custom CNN models have generated a p-value of less than 15%, suggesting that the augmentation process is slightly beneficial.

**Table 3.8.** 5x2 cross-validation statistics comparison for networks trained to recognize face inversion in grayscale images on raw data and on augmented data.

| Network name | t | p-value | Is significant |
|---|---|---|---|
| **Proposed CNN** | 2.172 | 0.082 | False |
| **Custom CNN** | 1.873 | 0.120 | False |
| **EEGNet SSVEP** | 1.482 | 0.198 | False |
| **ResNet-50** | 0.963 | 0.380 | False |

**Table 3.9** shows the average metrics for each trained classifier. We can estimate that the accuracy of custom CNN, EEGNet SSVEP, Proposed CNN, and Resnet-50 validation increased by 4.1%, 2.1%, 4.5%, and 5%, respectively. Meanwhile, the accuracy of the control subject's data increased only from 0.4% to 1.5%. Nonetheless, the augmentation slightly enhances the model's perception of the stimuli.

**Table 3.9.** The average evaluation metrics for all tested architectures with 5 times repeated 2-fold cross validation of facial inversion classes in gray images. The results are sorted by the control subject's accuracy.

| Model | Run time (sec) | Validation accuracy | Control accuracy | Alzheimer accuracy | Validation AUC | Control AUC | Alzheimer AUC |
|---|---|---|---|---|---|---|---|
| Proposed CNN augmented | 139.543 | 0.793 | 0.57 | 0.503 | 0.844 | 0.589 | 0.507 |
| EEGNet SSVEP augmented | 225.91 | 0.759 | 0.559 | 0.493 | 0.833 | 0.592 | 0.495 |
| Proposed CNN | 113.453 | 0.748 | 0.555 | 0.498 | 0.798 | 0.569 | 0.499 |
| Custom CNN augmented | 67.494 | 0.854 | 0.552 | 0.51 | 0.901 | 0.572 | 0.516 |
| EEGNet SSVEP | 250.49 | 0.738 | 0.547 | 0.493 | 0.81 | 0.57 | 0.494 |
| Custom CNN | 53.971 | 0.813 | 0.545 | 0.501 | 0.857 | 0.558 | 0.504 |
| ResNet-50 augmented | 158.834 | 0.719 | 0.52 | 0.492 | 0.763 | 0.53 | 0.495 |
| ResNet-50 | 165.107 | 0.669 | 0.516 | 0.498 | 0.706 | 0.527 | 0.493 |

The best performing model with the highest average accuracy score for control group subject is our Proposed CNN model with augmented data. The model achieved 1.1% higher accuracy than the EEGNet SSVEP classifier. Therefore, the results of the Proposed CNN model are presented, where

the maximum accuracy of the control group is 58.33% (AUC = 61.22%) and the accuracy score of the Alzheimer's patient is 52.21% (AUC = 50.86%).

**Fig. 3.20** shows the accuracy and loss curves during the Proposed CNN training, when the maximum training epoch was set to 1000. However, after 30 epochs of training, the model was stopped since the validation loss stopped decreasing, as the model goal was to minimize binary cross-entropy loss. The validation data had a maximum accuracy of 75.54%, which corresponded to a training accuracy of 98.75%. The wide gap between training and validation results indicates that there is some overfitting. Because grayscale images showed a less distinct pattern in ERP, training data may be unrepresentative, as it has fewer established patterns and contains more noise, making it difficult for the classifier to learn the task.

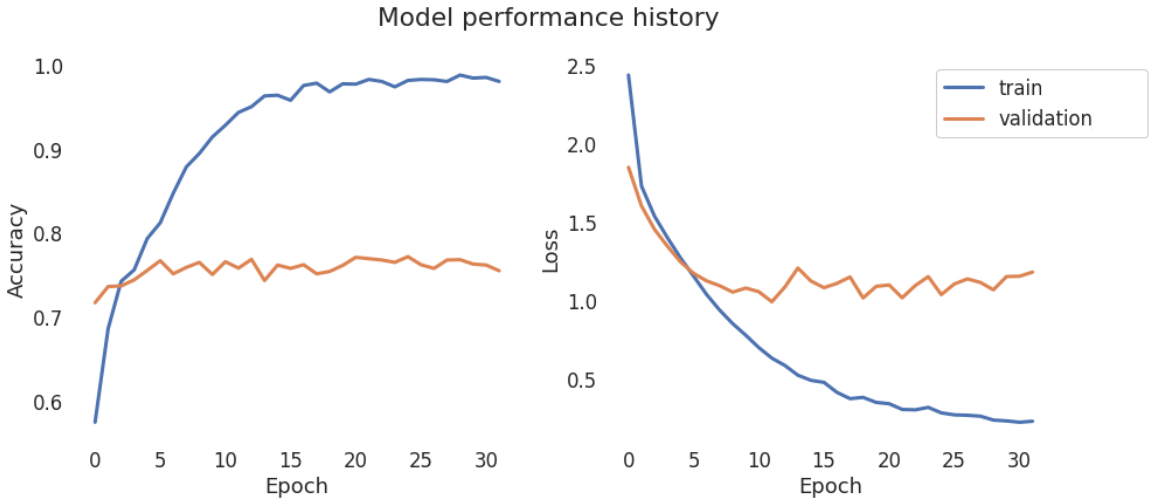



**Fig. 3.20.** The proposed CNN model's training history in grayscale images

The receiver operating characteristic curves (ROC curves) (see **Fig. 3.21**) are one of the most essential evaluation measures for evaluating the classification model's performance. The ROC of the control group is on the left, and the ROC of the AD patient is on the right. Presented plots indicate that the model is better than random (AUC = 0.61) for the control subject, however the model is completely random for Alzheimer's visual classification (AUC = 0.51).

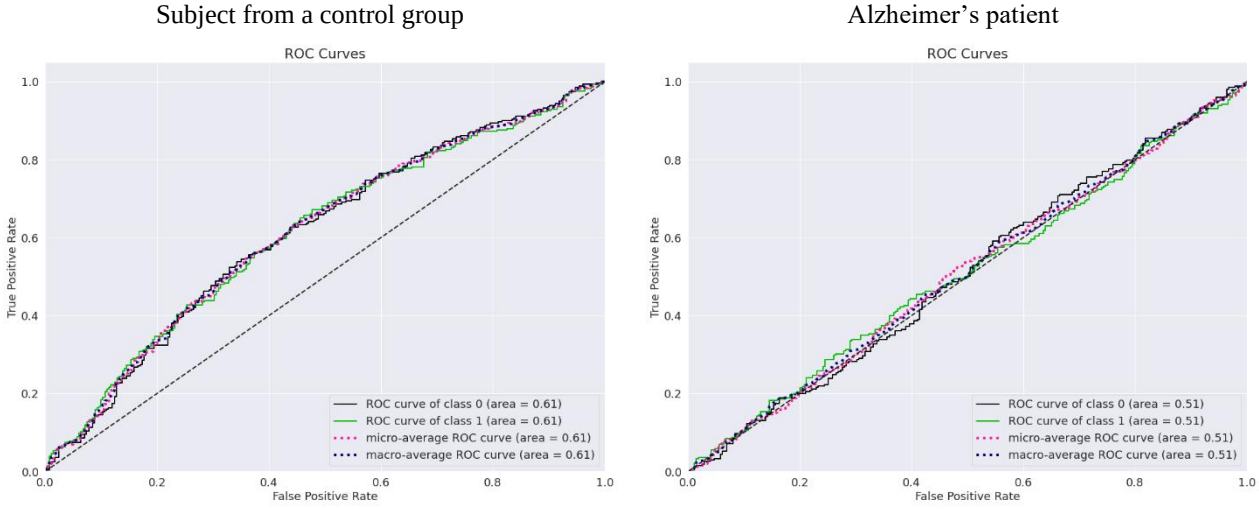



**Fig. 3.21.** The ROC curves for a control group subject and Alzheimer's patient using the Proposed CNN model in classifying facial inversion in grayscale images

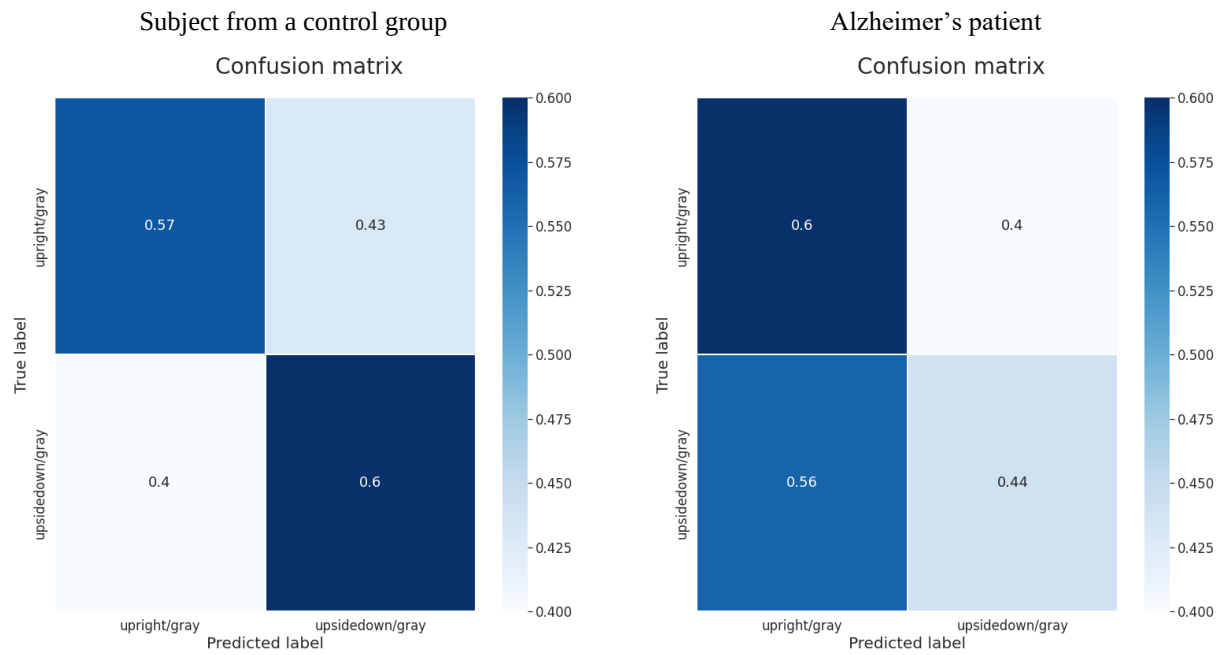



**Fig. 3.22.** Confusion matrices for a control group subject and Alzheimer's patient using the Proposed CNN model in classifying facial inversion in grayscale images

To get an overall picture of the classifier's performance, **Fig. 3.22** shows normalized confusion matrices. Matrices presents comparison for real and predicted response values to describe how often visual stimuli were correctly identified. The left graph depicts the performance of the control subject, while the right graph depicts the performance of the AD patient. For a control group subject, the upright and upside-down facial positions are classified equally correctly. While in Alzheimer's data, there are more falsely predicted upright stimuli.

## 3.5. Overview of Visual Prediction System Requirements

At the beginning of the research, certain requirements were raised, which are taken into account in this section. All the described functional requirements have been implemented in the code. For non-functional requirements, a single-epoch response time test was performed. The average processing time for epoch processing is 1.75 seconds, which includes converting the EEG signal into GASF, reducing the image size, loading a pre-trained model, and predicting the label. The visual stimulus prediction system was able to classify the class with greater accuracy than the random classifier, while color images had higher accuracy than gray images. In addition, the model's one training procedure lasted less than 3 minutes, which is a satisfying amount of time. However, due to the generalization gap, the learning curves showed a minor overfitting. This could be enhanced by tuning hyperparameters or otherwise pre-processing the data. In conclusion, all project requirements were met.

## Conclusions

In this study, the binary classification of face inversion visual stimuli in a control group and a patient with Alzheimer's disease was investigated. Electrodes from the occipital lobe were used to determine the visual stimuli using EEG signal data from 8 healthy women aged 63–70 years. It has been hypothesized that if a patient with Alzheimer's disease has memory loss rather than visual processing impairments, the system should provide classes similarly to healthy people. However, the findings were considerably different. In the second experiment, a 20% difference in AUC between healthy and AD patient was observed using color images. In the third trial, employing gray images, a 10% difference was detected. Noise in all brain regions is known to affect signals for Alzheimer's sufferers. The developed system is unable to predict the visual stimulus with greater accuracy than the individual, which is 81% and 69%, respectively for color and gray images. This could indicate that not only memory, but also visual processes are impaired in this person's brain.

The developed system consisted of pre-processing of the EEG signal, augmentation using RGAN, transformation of the brain channels responsible for image processing into a 2D domain using the GAF technique, and classification. Unlike other studies, this one combined training data from seven people and tested with data from an oldest person and an Alzheimer's patient separately. In this way, the training data did not see any signals from any of the test subjects, indicating that the user-independent system is being tested that does not require calibration with new user data. Meanwhile, it has been common practice in prior studies to conduct training individually for each participant, resulting in a unique model for each person. Our approach has the advantage of attempting to extract the most common features of visual processing from signals that are experienced by everyone.

Four different classifiers were investigated in this study: the classical image network ResNet-50, Custom CNN as it performed best on EEG signals transformed into GASF images, and EEGNet SSVEP as it was proposed as a model for steady-state visual evoked potentials. Our Proposed CNN architecture for smaller data sets, on the other hand, produced the best accuracy and AUC results, with a 75% AUC in the control group and a 55% AUC in the AD patient while processing color images. Meanwhile, grayscale images achieved 61% AUC in healthy subjects and 51% in AD patients. The proposed new architecture consisted of five convolutional layers with the most common best practices for generalization, such as the use of dropout, l2 regulation, and the addition of Gaussian noise. However, due to the limited and noisy training data, a minor overfitting was detected during model training. The RGAN data augmentation method has been applied directly to the signal space in an attempt to tackle this problem. Although the results showed that the 25% of the data augmentation procedure was not statistically significant, it did have small importance, and the models did not behave identically.

The findings should be made considering traditional limitations in medical data, that limited data sets were used in the study, having only 9 subjects, all of whom were female and elderly. In summary, promising results in healthy patients have been seen, indicating that the visual stimulus of facial inversion has the potential to be further researched and developed. This was the first study to our knowledge that used deep learning technology to compare the visual cortex of a healthy person and an Alzheimer's patient to determine whether a brain area was damaged. The following are the findings of this study:

1. By classifying the inversion of faces in visible color images with an emotion component, an average accuracy of 61.7% was achieved for the control subject. Compared to previous work [7] to detect upright and inverted faces, this proposed method achieves 5.7% better results. However, the classification using data from Alzheimer's patients remains almost random. Because Alzheimer's disease patients' brain signals are noisier than healthy people's, the Alzheimer's-specific noise reduction might be investigated.
2. Predicting the scope of attention when seeing grayscale facial images with a familiarity component reached only 57% when tested with data from a healthy elderly subject. Predictions for Alzheimer's disease patients, on the other hand, remain random. This might be explained by the fact that in this experiment, the AD patient made the most mistakes, accurately estimating just 69% of the stimuli. Grayscale pictures with a black background may blend in with darker color hair, and hair position may be a critical identifier for detecting a face's position. These results suggest that the AD patient did not process the visual information in the same way as healthy individuals, and the visual cortex may be affected by the disease.
3. Training the classifier with the augmented data using RGAN model increased the training validation overall performance from 2.1% to 5%. The proposed CNN network processes both color and gray images with weak statistical significance when compared to the performance with 25% augmented data and the performance with the original data. This suggests that the outcomes of the two models were not identical. In the future, the augmentation process could be tested using various percentages of augmented data and tuning hyperparameters.

Finally, several recommendations are suggested for the future. Collecting more data with a larger group of patients is always desirable, but at the same time it is very difficult and time consuming. As a result, smarter preprocessing might be done specifically for AD patient data to improve the facial inversion system. Preprocessing was done automatically in this study using several techniques, however manual removal of artifacts could also improve performance. Additionally, more sophisticated feature extraction techniques could be tested, for example, using a new method based on GASF images of Hjorth parameters [32]. Finally, optimizing the architecture and hyperparameters would contribute to enhance the accuracy of model predictions.

## List of References


[1] S. Vaid, P. Singh, and C. Kaur, "EEG Signal Analysis for BCI Interface: A Review," in *2015 Fifth International Conference on Advanced Computing & Communication Technologies*, 2015, pp. 143–147. doi: 10.1109/ACCT.2015.72.

[2] G. Fiscon *et al.*, "Combining EEG signal processing with supervised methods for Alzheimer's patients classification," *BMC Medical Informatics and Decision Making*, vol. 18, no. 1, 2018, doi: 10.1186/s12911-018-0613-y.

[3] J. S. Kumar and P. Bhuvaneswari, "Analysis of Electroencephalography (EEG) Signals and Its Categorization–A Study," *Procedia Engineering*, vol. 38, pp. 2525–2536, 2012, doi: 10.1016/j.proeng.2012.06.298.

[4] R. Abiri, S. Borhani, Y. Jiang, and X. Zhao, "Decoding Attentional State to Faces and Scenes Using EEG Brainwaves," *Complexity*, vol. 2019, 2019, doi: 10.1155/2019/6862031.

[5] P. Shenoy and D. S. Tan, "Human-Aided Computing: Utilizing Implicit Human Processing to Classify Images," in *Conference on Human Factors in Computing Systems - Proceedings*, 2008, pp. 845–854. doi: 10.1145/1357054.1357188.

[6] R. Kasper, K. Das, M. Eckstein, and B. Giesbrecht, "Decoding Information Processing When Attention Fails: An Electrophysiological Approach," in *Advances in Understanding Human Performance*, 2010, pp. 42–51. doi: 10.1201/ebk1439835012-6.

[7] A. List, M. D. Rosenberg, A. Sherman, and M. Esterman, "Pattern classification of EEG signals reveals perceptual and attentional states," *PLoS ONE*, vol. 12, no. 4, pp. 1–23, 2017, doi: 10.1371/journal.pone.0176349.

[8] M. S. N. Chowdhury, A. Dutta, M. K. Robison, C. Blais, G. A. Brewer, and D. W. Bliss, "Deep Neural Network for Visual Stimulus-Based Reaction Time Estimation Using the Periodogram of Single-Trial EEG," *Sensors (Basel)*, vol. 20, no. 21, 2020, doi: 10.3390/s20216090.

[9] M. C. Moulson, B. Balas, C. Nelson, and P. Sinha, "EEG Correlates of Categorical and Graded Face Perception," *Neuropsychologia*, pp. 3847–3853, 2011, doi: 10.1016/j.neuropsychologia.2011.09.046.

[10] M. R. Nuwer, "10-10 electrode system for EEG recording," *Clinical Neurophysiology*, vol. 129, no. 5, p. 1103, 2018, doi: 10.1016/j.clinph.2018.01.065.

[11] C. Jacques, O. d'Arripe, and B. Rossion, "The time course of the inversion effect during individual face discrimination," *Journal of Vision*, vol. 7, no. 8, pp. 1–9, 2007, doi: 10.1167/7.8.3.

[12] Y. Nihei, T. Minami, and S. Nakauchi, "Brain Activity Related to the Judgment of Face-Likeness: Correlation between EEG and Face-Like Evaluation," *Frontiers in Human Neuroscience*, vol. 12, no. 56, pp. 1–12, 2018, doi: 10.3389/fnhum.2018.00056.

[13] B. Rossion and I. Gauthier, "How Does the Brain Process Upright and Inverted Faces?," *Behav Cogn Neurosci Rev*, vol. 1, no. 1, pp. 63–75, 2002, doi: 10.1177/1534582302001001004.

[14] I. Muukkonen, K. Ölander, J. Numminen, and V. R. Salmela, "Spatio-temporal dynamics of face perception," *Neuroimage*, vol. 209, 2020, doi: 10.1016/j.neuroimage.2020.116531.

[15] R. Popa, "Emotion Detection Using EEG Signals," *International Journal of Progressive Sciences and Technologies (IJPSAT*, vol. 24, no. 1, pp. 175–182, 2020, doi: 10.52155.

[16] K. Dobs, L. Isik, D. Pantazis, and N. Kanwisher, "How face perception unfolds over time," *Nature Communications*, vol. 10, 2019, doi: 10.1038/s41467-019-09239-1.

[17] Y. Wang, S. Wang, and M. Xu, "The Function of Color and Structure Based on EEG Features in Landscape Recognition," *International Journal of Environmental Research and Public Health*, vol. 18, 2021, doi: 10.3390/ijerph18094866.

[18] R. Zafar, M. N. Malik, H. Hayat, and A. S. Malik, “Decoding Brain Patterns for Colored and Grayscale Images using Multivariate Pattern Analysis,” *KSII Transactions on Internet and Information Systems*, vol. 14, no. 4, pp. 1543–1561, Apr. 2020, doi: 10.3837/TIIS.2020.04.008.

[19] C. Or, T. Retter, and B. Rossion, “The contribution of color information to rapid face categorization in natural scenes,” *Journal of Vision*, vol. 19, no. 5, May 2019, doi: 10.1167/19.5.20.

[20] G. F. Woodman, “A brief introduction to the use of event-related potentials in studies of perception and attention,” *Attention, Perception, and Psychophysics*, vol. 72, no. 8, pp. 2031–2046, 2010, doi: 10.3758/APP.72.8.2031.

[21] C. Saavedra and L. Bougrain, “Processing Stages of Visual Stimuli and Event-Related Potentials,” *The NeuroComp/KEOpS’12 workshop*, 2012, Accessed: May 25, 2022. [Online]. Available: https://hal.inria.fr/hal-00756795/document

[22] S. Nagel and M. Spüler, “Modelling the brain response to arbitrary visual stimulation patterns for a flexible high-speed Brain-Computer Interface,” *PLoS ONE*, vol. 13, no. 10, 2018, doi: 10.1371/journal.pone.0206107.

[23] N. Kosmyna, J. T. Lindgren, and A. Lécuyer, “Attending to Visual Stimuli versus Performing Visual Imagery as a Control Strategy for EEG-based Brain-Computer Interfaces,” *Scientific Reports*, vol. 8, no. 1, 2018, doi: 10.1038/s41598-018-31472-9.

[24] C. Mazzi, G. Massironi, J. Sanchez-Lopez, L. de Togni, and S. Savazzi, “Face Recognition Deficits in a Patient With Alzheimer’s Disease: Amnesia or Agnosia? The Importance of Electrophysiological Markers for Differential Diagnosis,” *Frontiers in Aging Neuroscience*, vol. 12, 2020, doi: 10.3389/fnagi.2020.580609.

[25] I. Winkler, S. Debener, K.-R. Müller, and M. Tangermann, “On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP,” in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, 2015, pp. 4101–4105. doi: 10.1109/EMBC.2015.7319296.

[26] M. Klug and K. Gramann, “Identifying key factors for improving ICA-based decomposition of EEG data in mobile and stationary experiments,” *Eur J Neurosci*, vol. 54, no. 12, pp. 8406–8420, 2021, doi: 10.1111/ejn.14992.

[27] R. Maskeliunas, R. Damasevicius, I. Martisius, and M. Vasiljevas, “Consumer-grade EEG devices: are they usable for control tasks?,” *PeerJ 4:e1746*, 2016, doi: 10.7717/peerj.1746.

[28] T. Oates and Z. Wang, “Spatially Encoding Temporal Correlations to Classify Temporal Data Using Convolutional Neural Networks,” *JCSS*, 2015, doi: 10.48550/arXiv.1509.07481.

[29] Z. Wang and T. Oates, “Imaging Time-Series to Improve Classification and Imputation,” in *International Joint Conference on Artificial Intelligence*, 2015, vol. 2015-Janua. doi: 10.48550/arXiv.1506.00327.

[30] K. P. Thanaraj, B. Parvathavarthini, U. J. Tanik, V. Rajinikanth, S. Kadry, and K. Kamalanand, “Implementation of Deep Neural Networks to Classify EEG Signals using Gramian Angular Summation Field for Epilepsy Diagnosis,” *arXiv:2003.04534*, 2020, doi: 10.48550/arXiv.2003.04534.

[31] V. v. Grubov, A. E. Runnova, M. K. Kurovskaya, A. N. Pavlov, A. A. Koronovskii, and A. E. Hramov, “Demonstration of brain noise on human EEG signals in perception of bistable images,” *Dynamics and Fluctuations in Biomedical Photonics XIII*, vol. 9707, 2016, doi: 10.1117/12.2207390.

[32] R. Damaševičius, R. Maskeliūnas, M. Woźniak, and D. Połap, “Visualization of physiologic signals based on Hjorth parameters and Gramian Angular Fields,” *2018 IEEE 16th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, pp. 91–96, Feb. 2018, doi: 10.1109/SAMI43701.2018.

[33] F. Barbiero, “Convolutional Neural Network and Source Separation for bio-signals recognition and classification,” 2019. Accessed: Apr. 20, 2022. [Online]. Available: https://webthesis.biblio.polito.it/13656/

[34] B. Boashash, L. Boubchir, and G. Azemi, “A methodology for time-frequency image processing applied to the classification of nonstationary multichannel signals using instantaneous frequency descriptors with application to newborn EEG signals,” *Eurasip Journal on Advances in Signal Processing*, vol. 2012, 2012, doi: 10.1186/1687-6180-2012-117.

[35] Y. Gao, B. Gao, Q. Chen, J. Liu, and Y. Zhang, “Deep Convolutional Neural Network-Based Epileptic Electroencephalogram (EEG) Signal Classification,” *Frontiers in Neurology*, vol. 11, 2020, doi: 10.3389/fneur.2020.00375.

[36] Q. Liu *et al.*, “Spectrum Analysis of EEG Signals Using CNN to Model Patient’s Consciousness Level Based on Anesthesiologists’ Experience,” *IEEE Access*, vol. 7, pp. 53731–53742, 2019, doi: 10.1109/ACCESS.2019.2912273.

[37] F. Li, F. He, F. Wang, D. Zhang, Y. Xia, and X. Li, “A Novel Simplified Convolutional Neural Network Classification Algorithm of Motor Imagery EEG Signals Based on Deep Learning,” *Applied Sciences (Switzerland)*, vol. 10, no. 5, 2020, doi: 10.3390/app10051605.

[38] H. Zamanian and H. Farsi, “A New feature extraction method to Improve Emotion Detection Using EEG Signals,” *Electronic Letters on Computer Vision and Image Analysis*, vol. 17, no. 1, pp. 29–44, 2018, doi: 10.5565/rev/elcvia.1045.

[39] J. Cho and H. Hwang, “Spatio-Temporal Representation of an Electoencephalogram for Emotion Recognition Using a Three-Dimensional Convolutional Neural Network,” *Sensors (Switzerland)*, vol. 20, no. 12, 2020, doi: 10.3390/s20123491.

[40] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, “Deep learning-based electroencephalography analysis: A systematic review,” *Journal of Neural Engineering*, vol. 16, no. 5, Aug. 2019, doi: 10.1088/1741-2552/ab260c.

[41] H. Choi, S. Ryu, and H. Kim, “Short-Term Load Forecasting based on ResNet and LSTM,” in *2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, Dec. 2018, pp. 1–6. doi: 10.1109/SmartGridComm.2018.8587554.

[42] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.

[43] D. Sarwinda, R. H. Paradisa, A. Bustamam, and P. Anggia, “Deep Learning in Image Classification using Residual Network (ResNet) Variants for Detection of Colorectal Cancer,” in *Procedia Computer Science*, 2021, vol. 179, pp. 423–431. doi: 10.1016/j.procs.2021.01.025.

[44] S. Haradal, H. Hayashi, and S. Uchida, “Biosignal Data Augmentation Based on Generative Adversarial Networks,” *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 368–371, 2018, doi: 10.1109/EMBC.2018.8512396.

[45] S. M. Abdelfattah, G. M. Abdelrahman, and M. Wang, “Augmenting the Size of EEG datasets Using Generative Adversarial Networks,” *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–6, 2018, doi: 10.1109/IJCNN.2018.8489727.

[46] T. Luo, Y. Fan, L. Chen, G. Guo, and C. Zhou, “EEG Signal Reconstruction Using a Generative Adversarial Network With Wasserstein Distance and Temporal-Spatial-Frequency Loss,” *Frontiers in Neuroinformatics*, 2020, doi: 10.3389/fninf.2020.00015.

[47] K. G. Hartmann, R. T. Schirrmeister, and T. Ball, “EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals,” *arXiv*, 2018, doi: 10.48550/arXiv.1806.01875.

[48] N. K. N. Aznan, A. Atapour-Abarghouei, S. Bonner, J. D. Connolly, N. A. Moubayed, and T. Breckon, “Simulating Brain Signals: Creating Synthetic EEG Data via Neural-Based Generative Models for Improved SSVEP Classification,” *2019 International Joint Conference on Neural Networks (IJCNN)*, 2019, doi: 10.1109/IJCNN.2019.8852227.

[49] C. Esteban, S. L. Hyland, and G. Ratsch, “Real-valued (Medical) Time Series Generation with Recurrent Conditional GANs,” *arXiv*, 2017, doi: 10.48550/arXiv.1706.02633.

[50] I. J. Goodfellow *et al.*, “Generative Adversarial Networks,” *Advances in Neural Information Processing Systems*, vol. 3, no. 11, Jun. 2014, doi: 10.1145/3422622.

[51] M. Mirza and S. Osindero, “Conditional Generative Adversarial Nets,” *arXiv*, 2014, doi: 10.48550/arXiv.1411.1784.

[52] J. G. Moreno-Torres, J. A. Saez, and F. Herrera, “Study on the Impact of Partition-Induced Dataset Shift on k-Fold Cross-Validation,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 8, pp. 1304–1312, 2012, doi: 10.1109/TNNLS.2012.2199516.

[53] D. Berrar, “Cross-validation,” in *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, vol. 1–3, Elsevier, 2018, pp. 542–545. doi: 10.1016/B978-0-12-809633-8.20349-X.

[54] S. Raschka, “Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning,” *University of Wisconsin–Madison Department of Statistics*, 2018. https://ar5iv.labs.arxiv.org/html/1811.12808 (accessed May 19, 2022).

[55] T. G. Dietterich, “Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms,” *Neural Comput*, vol. 10, no. 7, 1998, doi: 10.1162/089976698300017197.

[56] Z. H. Zhou, “Cost-Sensitive Learning,” in *Lecture Notes in Computer Science*, vol. 6820, 2011, pp. 17–18. doi: 10.1007/978-3-642-22589-5_2.

[57] M. Hossin and M. N. Sulaiman, “A Review on Evaluation Metrics for Data Classification Evaluations,” *International Journal of Data Mining & Knowledge Management Process*, vol. 5, no. 2, pp. 1–11, Mar. 2015, doi: 10.5121/ijdkp.2015.5201.

[58] A. Tharwat, “Classification assessment methods,” *Applied Computing and Informatics*, vol. 17, no. 1, pp. 168–192, 2021, doi: 10.1016/j.aci.2018.08.003.

[59] K. H. Zou, A. J. O’Malley, and L. Mauri, “Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models,” *Circulation*, vol. 115, no. 5, pp. 654–657, Feb. 2007, doi: 10.1161/CIRCULATIONAHA.105.594929.

[60] C. Mazzi, S. Savazzi, J. Sanchez-Lopez, G. Massironi, and laura de Togni, “Face Recognition Deficits in a Patient With Alzheimer’s Disease: Amnesia or Agnosia?,” 2020. https://figshare.com/articles/Face_recognition_deficits_in_a_patient_with_Alzheimer_s_disease_amnesia_or_agnosia_/11913243/1 (accessed Dec. 16, 2020).

[61] D. Komolovaitė, R. Maskeliūnas, and R. Damaševičius, “Deep Convolutional Neural Network-Based Visual Stimuli Classification Using Electroencephalography Signals of Healthy and Alzheimer’s Disease Subjects,” 2022, doi: 10.3390/life12030374.

[62] F. I. Alarsan and M. Younes, “Best Selection of Generative Adversarial Networks Hyper-Parameters using Genetic Algorithm,” *SN Computer Science*, 2021, doi: 10.1007/s42979-021-00689-3.

[63] N. Waytowich *et al.*, “Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials,” *Journal of Neural Engineering*, vol. 15, no. 6, pp. 1–23, 2018, doi: 10.1088/1741-2552/aae5d8.

[64] A. Camuto, M. Willetts, U. Simsekli, S. Roberts, and C. Holmes, “Explicit Regularisation in Gaussian Noise Injections,” *Advances in Neural Information Processing Systems*, 2020, doi: 10.48550/arXiv.2007.07368.

[65] Z. You, J. Ye, K. Li, Z. Xu, and P. Wang, “Adversarial Noise Layer: Regularize Neural Network By Adding Noise,” *arXiv*, May 2018, doi: 10.48550/arXiv.1805.08000.

[66] A. Lewkowycz and G. Gur-Ari, “On the training dynamics of deep networks with L2 regularization,” *arXiv*, 2021, doi: 10.48550/arXiv.2006.08643.

[67] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014, Accessed: Apr. 04, 2022. [Online]. Available: http://jmlr.org/papers/v15/srivastava14a.html

[68] A. Gramfort *et al.*, “MEG and EEG data analysis with MNE-Python,” *Frontiers in Neuroscience*, vol. 7, 2013, doi: 10.3389/fnins.2013.00267.