# Brain-Computer Interface controlled simulated car in CARLA

Mangal Deep.B.M

7206937

Faculty of Computer Science

Dortmund University of Applied Sciences and Arts

Submitted in partial satisfaction of the requirements for the

Degree of M.Eng

in Embedded Systems for Mechatronics

*Supervisor*    Dr. Andreas Becker

27-January-2023

# Abstract

This is where your abstract will go. Usually this is written last, after writing the entirety of your thesis.

# Acknowledgements

Firstly, I want to thank Dad and Mom.[1] Here is another note.

---

[1]Here is a footnote

# List of Figures

# Table of Contents

# List of Abbreviations

BCI            Brain-Computer-Interface

CNN            Convolutional Neural Network

CSP            Common Spatial Patterns

CWT            Continuous Wavelet Transform

DCNN           Deep Convolutional Neural Network

DL             Deep Learning

DWT            Discrete Wavelet Transform

ECoG           Electrocorticography

EEG            Electroencephalography

EOG            Electrooculography

EP             Evoked Potential

FWHM           full width half maximum

ICA            Independent Component Analysis

LDA            Linear Discriminant Analysis

LSL            Lab Streaming Layer

MI             Motor Imagery

PCA            Principal Component Analysis

RNN            Recurrent Neural Network

ROS1           Robotic Operating System -1

| | |
|---|---|
| SMR | Sensorimotor Rhythm |
| SNR | Signal-to-Noise Ratio |
| SOTA | State-of-the-art |
| SSP | Signal Space Projection |
| STFT | Short-Time Frequency Transform |
| SVM | Support Vector Machine |
| TF | Time-Frequency |
| WSN | Wavelet Scattering Network |
| WST | Wavelet Scattering Transform |

# Chapter 1

# Introduction

Human computer interaction has been evolving over the past decades. With development in technologies, the physical contact between the computer and the human is decreasing rapidly. Advanced systems which work on speech and gesture control still requires a minimal effort from the user to interact with the machine. Though these effort seem to be mere, it is a challenging task for humans with disabilities. Systems which work on facial gestures bridge this gap to an extent but it does not completely understand what is the actual intent of the person. Brain Computer Interface paves way to encode the persons intent and thoughts without the need for any physical effort. It provides enormous capabilities for physically challenged people to express themselves just by their thoughts.

Autonomous vehicles are the future of mobility, several companies around the world invest and research on new technologies to solve new challenges that appears in developing level 5 autonomy. The level of human interaction with the vehicle has been decreasing with increasing safety. However including human in the loop is necessary at certain times to avoid any undesirable events. Level 5 autonomous vehicles is still a long way to go, but by bringing in a minimal interaction of the driver with the system, safety can be ensured. One of the ways of achieving it is interfacing the thought and decision process of the driver to the autonomous vehicle, a process commonly referred to as Brain-Computer-Interface (BCI).

Once shown in science fiction novels and movies, Brain-Computer Interface (BCI) has become widely researched and developed in the academic institutions and industries in the past decades. It has been applied and tested on mammals for a wide range

of applications. However it has its own challenges and limitations. With evolving technologies, new innovations and discoveries are made to understand and decode the brain waves better. The analysis of the brain signals have seen a shift in the paradigm with the introduction of machine learning techniques.

Modern BCI design involves understanding brain dynamics, recognizing the patterns in the brain waves and derive the statistics of the data obtained optimize it and use them as features to predict or classify the task. However it is very challenging to create a BCI system that can work on any person as the brain signals are task specific and brain signal signatures are very unique to a person. The cortex folding and the relevant functional maps are different across individuals. Even for the very same person, the brain dynamics are non-stationary at all time scales adding to it, it is almost impossible to place the electrodes exactly on the same location for every recording sessions. Further the psychological states of the user such as boredom, distraction... play a significant role in the quality of the signal measured. The Signal-to-Noise Ratio (SNR) in a brain signal is very poor, that makes it difficult to obtain required information from brain signal. It is harder to spatially measure the data from one region as large collection of neurons are involved in many different activity, not just one. These challenges can vary a lot depending on the methods used to record and analyse the brain signals as well as the task that needs to be achieved.

Given the challenges, the goal of this work is to steer a simulated car in the CARLA environment using brain signals obtained from the OpenBCI headgear. The brain signal from 16 channel OpenBCI headgear is fed through signal processing pipeline to extract the relevant Motor Imagery (MI) features and classify the users intention. The signal preprocessing pipeline constitutes reliable and conventional signal processing techniques as well as state-of-the-art deep learning techniques. Several tools, libraries and frameworks such as MNE, Numpy, Scipy, Scikit-learn and PyTorch are used to achieve the goal. The communication between the OpenBCI system and the signal processing pipeline is established using Lab Streaming Layer (LSL) and the steering information decoded by the signal processing pipeline is sent to CARLA through Robotic Operating System (ROS1) . The algorithms are packed into a ROS pack which consists of multi-threaded nodes enabling real-time information transfer

to CARLA. All the relevant code and references are made available in GitHub. Several open-source datasets are used to setup the basic brain signal processing pipeline and later tuned to work with the data obtained form OpenBCI headgear.

## OpenBCI headgear

OpenBCI is an open-source BCI platform that develops hardware and software for BCI scientific research. It provides electrodes, interface boards, graphical user interface and python API. In this work, the OpenBCI system consists of a 3D printed headgear with 16 EEG electrodes connected to Cython + Daisy board that interacts to the host machine wireless through an external dongle via serial communication. Cython is a 8-channel neural interface with a 32-bit processor which has a sampling frequency of 250Hz. It is topped with Daisy, an extension board which can add up to 8 more channels to the system. This leads to drop in the sampling frequency to 125HZ. With a WiFi module it is possible to achieve 1kHz. The potential difference between the electrodes and the reference points such as linked mastoids, inion or nasion. The measurements can be visualized on the host machine using OpenBCI GUI. The electrodes are placed in the internationally recognized 10-20 system. The measurement data can be streamed to the system through BrainFlow or LSL.

## CARLA

CARLA [1] is an open-source simulator to develop, train and validate autonomous driving systems. It supports ROS1 for establishing communication between CARLA server and clients that can feed into or extract data from CARLA environment.

## MNE

MNE [2] is an open-source python package for exploring, visualizing and analyzing neuro-physiological data. It provides support for various preprocessing, feature extraction and classification steps in brain signal analysis.

## Summary

# Chapter 2

# Formulation

## Introduction

The goal of this work is to control the simulated car in CARLA using motor intent from the users brain activity. The measured EEG contains several information and the motor intent of the user can be extracted from the Motor Imagery (MI) information. This chapter discusses different brain signal extraction and information that can be derived from it.

## 2.1 Brain Signal Acquisition/ Extraction

Neurons are the basic computational unit of the nervous system. The electric signal recorder as brain wave is the result of electro-chemical interactions that occur at the outer membrane of neurons. In case of an event, the raise and fall of potential at the membrane of the neurons is called as action potential. The electrical activity at the neurons enable us to record and decode the brain activities. There are many ways to record the electrical activity and are broadly classified into invasive and non-invasive recording. Some technologies can also be used to stimulate neurons or brain regions, that allows BCIs to send feedback to brain based on interactions with the world. This work is only on Electroencephalography, a non-invasive approach, hence the topic on invasive approaches and other non-invasive approaches are explained briefly.

### 2.1.1 Invasive Approaches

Approaches requiring some surgery where an electrode is placed in direct contact with required region in the brain are invasive methods of recording brain signals. These approaches are typically performed on animals such as monkeys, rats. In case of humans these approaches are carried on under strict clinical settings. As the recording sensor is in direct contact with the brain tissues, these provide higher quality signals with high SNR, higher spatial resolution and less spatial smearing compared to the non-invasive approaches.

#### 2.1.1.1 Electrocorticography (ECoG)

ECoG is an extra cortical invasive electro-physiological monitoring method [3]. Typically performed in a clinical setting, a strip of electrodes are placed on the surface of interest on the brain. It is the most commonly used invasive approach as it has many advantages compared to other invasive approaches.

### 2.1.2 non-invasive Approaches

Approaches that gather brain signals over the surface of the scalp, without any surgery or electrodes being inserted into the skull are termed non-invasive approaches. The signals could be collected by measuring the electrical or magnetic activity on the surface of the scalp. Some of the common non-invasive approaches are Electroencephalography (EEG) , functional magnetic resonance imaging (fMRI), functional near-infrared spectroscopy (fNIRS), magnetoencephalography (MEG), and electrooculography (EOG) .

#### 2.1.2.1 Electroencephalography (EEG)

EEG is the most commonly used non-invasive approach used to measure brain electrical activity on the surface of the scalp. The electrodes are usually placed according to internationally recognized 10-20 system [4]. The spatial resolution of the signals depend on the number of electrodes used. The temporal resolution depends on how many samples the system could measure in a second. Typically the temporal resol-

ution of the EEG signals is much better than the spatial resolution. In comparison to invasive approaches, the EEG systems have poor spatial and temporal resolution, and very poor SNR. Since the measurement is taken over the surface of the scalp, the obstruction of the skull and other tissues between cortex and the scalp act as a huge conductive surface leading to spatial smearing. However with other non-invasive approaches, EEG has higher temporal resolution, tolerant to noise and artifacts, low cost and no exposure to high intensity magnetic fields.

Spatial smearing is the effect by which all the electrodes tend to measure the same signal because of the conductive effect of the tissues between the cortex and scalp. EEG signals also very susceptible to other noises such as power-line, changing electrode impedance, eye movements, eye blinks, facial muscle movements and head movement. The brain signals measured from the EEG systems can be separated into frequency bands, where each frequency band represents to a specific brain state and level of awareness. The recorded brain signals contains information on various physiological, psychological, mental, sensory and cognitive activities. Hence it is required to analyse and extract the relevant information from the signal.

## 2.2 EEG Paradigms

The measured EEG data is rich in information relating to several human conscious and subconscious activities. In case of a controlled lab experiments, the brain signals are recorded with few stimuli and feedback requiring the focus of the user. Depending on the type of experiment and the task that is required to achieve, the brain signal analysis methods varies. These experiments are designed to utilize neurophysiological activities that broadly classified into Spontaneous EEG and Evoked Potential.

### 2.2.1 Spontaneous EEG

Spontaneous EEG also referred as Oscillatory Activity includes a wide range of task typically without any external stimulation such as mental task, motor imagery, sleeping, under fatigue stage. BCI systems that work with spontaneous EEGs are

also called Active BCI as they require consciously controlled thought independent from external events. BCI systems that work with spontaneous EEG are hard to train given their low SNR and variation in subjects.

### 2.2.1.1 Motor Imagery (MI)

Imagining a motor(physical) movement without performing an actual movement is termed Motor Imagery. It is one of the widely researched EEG paradigm that is used in several applications where the user is limited in their motor movement capabilities. The Sensorimotor rhythms (SMRs) are oscillatory events in the EEG signal originating from regions of the brain associated with control and action of a movement [5]. Gamma activity can be used in case of invasive BCI systems, but it s not suitable for non-invasive systems as they do not reach the scalp. It also helps in mental rehearsals as the person experience themselves performing the action. MI is due to two basic phenomena that occur in the brain onset of imagination - Event Related Desynchronization (ERD) and Event Related Synchronization (ERS). ERD/ERS refers to phenomena that the magnitude and the frequency distribution of the EEG signal power changes during a specific brain state. It is mostly observed in sensory, cognitive and motor tasks. During motor imagery contra-lateral ERD is observed in the mu band and after motor imagery ERS is observed in beta band.

Event Related Desynchronization (ERD) is due to activity of small set of neurons. A visual stimulation results in a short lasting attenuation or blocking of rhythm in the alpha band leading to power decrease of ongoing EEG signals. This decrease in the oscillatory is related to internally or externally paced events.

Event Related Synchronization (ERS) is due to activity of large set of neurons. The increase in oscillatory activity is again related to internally or externally paced events.It is characterized by short lasting amplitude enhancement.

## 2.2.2 Evoked Potential (EP)

Processing of physical stimulus rather than higher processes that might involve memory and attention. BCI systems that work with EPs could be also called Re-

active BCI as they rely on response to the stimulus provided. Few of the commonly researched EPs are described below.

### 2.2.2.1 Event Related Potential (ERP)

It is the most widely researched EPs and it is further classified based on the stimulus used to trigger the brain waves: visual evoked potential when stimulated visually, auditory evoked potential when stimulated with sound and somatosensory evoked potential when evoked with smell. The stimulus could be internal or external based on the BCI system. P300 is an important component in ERP and it is been widely used in many BCI systems existing in the market. It refers to the positive peak that appears 300ms after the onset of stimulus.

### 2.2.2.2 Error Related Potential (ErRP)

It is a result of an erroneous event generated by the error processing mechanism in the human brain. It provides a feature rich feedback to the BCI system and helps to tune the system in the desirable way.

## 2.3 Brain Signal Processing Pipeline

The brain signal is fed through several signal processing algorithms to extract the best possible information. These processing algorithms are specifically chosen to extract motor imagery data. The overall processing pipeline is given in the figure **??**. The first few steps in the processing pipeline is the same for both conventional and deep learning approaches. The variation is observed only at feature extraction and classification. These steps involved in the pipeline are explained in detail in the following sections and chapters. The data obtained needs to be processed effectively to get the best possible information as it influences the classification result dramatically. Most of the BCI systems in the literature use the following techniques to achieve the best results.

### 2.3.1 Data Extraction

The input to the pipeline could be an online/offline data from OpenBCI headgear or publicly available open-source datasets. The open-source data sets were used in order to design the pipeline and later the pipeline is calibrated to work with data from OpenBCI headgear.

#### 2.3.1.1 OpenBCI headgear

Before the experiment could be started several parameters are checked. In order to ensure the contact of electrodes on the users scalp, the impedance at the electrodes are monitored and adjusted until they fall into acceptable threshold i.e. less than $750K\Omega$. Next the frequency filters are configured to avoid DC offset and electric power line noises. However this filter is applied only to the visualizer and not to the recorded data. The electrodes are very susceptible to noises from electronic devices that appear around 25Hz even after filtering the DC offset and power line noises. Hence while performing the experiment the user is away from any electronic device. In experiments, it is also proved that it is not necessary when working with deep learning based signal processing pipeline.

Establishing a proper experiment is a key step in training the classifiers, By [**empty citation**] each trail is recorded for 12 seconds, the setup used to gather data for a three class system that consists of a window separated by a vertical line and two squares - yellow and blue, in the centre initially. The beginning of each trial is marked by change in color of the vertical line then the subject is requested to be in a idle state for the first 5 seconds. The motor imagery task to be performed is presented on the screen for 3 seconds with help of a yellow square by moving it to the left,right or standstill denoting only straight motion without steering, here the subject can prepare themselves. After which the subject either performs the instructed movement or imagines performing the movement for 3 seconds. The data collected so far could be wither used to train a model online or it can used as input to a presaved model and obtain the classification result. The result of the classifier is used to move the blue square accordingly. The answer presented to the subject stimulates feedback in

the user which is again captured in the recording for analyzing ErRP and to improve the classifier results.

LSL is used to establish communication between OpenBCI GUI and CARLA BCI Control ROS node. With the support of MNE, a mock LSL stream can be used to stream recorded data which helps in debugging the algorithm even in the absence of the OpenBCI headgear.

### 2.3.1.2 Open-source Data sets

BCI is a widely researched field for several decades, however the huge variation in the BCI experiments to obtain relevant information from the brain signal is a bottleneck in the publicly available dataset. This work required a multi-class motor imagery dataset recorded with 16+ channel BCI system and sampling rate of greater than 125Hz. Some of the datasets that are used in setting up the signal processing pipeline and used for comparative analysis are discussed further. The dataset thus used were re-sampled, channel picked in order to develop a signal processing pipeline compatible to the hardware in hand.

### 2.3.1.3 PhysioNet

PhysioNet dataset[6] consists of over 1500 one- and two-minute EEG recordings from 109 participants performing motor imagery task using 64-channel BCI2000 system at a sample rate of 160Hz. Each participants performed 14 experiments each involving either imagining or actually performing opening and closing fists and feet. The PhysioNet dataset followed the international 10-20 system making it easy to construct MNE Raw frame.

### 2.3.1.4 Berlin BCI Competition (BBCI)

BBCI [7] competition was held for few years where the competitors had to come up with the best performing signal processing pipeline for dataset. This includes a list of various EEG datasets, this work uses the motor imagery dataset from BBCI III IVa. The recordings were performed using 118-channel EEG system at a sampling rate of 1000Hz. It consists of 5 participants, each performing motor imagination of right

fist or foot. The electrode locations in the BCI system is arbitrary and the location info was provided with the dataset. It required manual assignment of locations to enable all the functionalities in MNE framework. From this point the word- dataset and data are used interchangeably and it is used to denote the data obtained from the OpenBCI headgear and the open-source datasets unless specified.

### 2.3.2   Artifact removal techniques

Any undesirable signals that originate outside the brain are termed artifacts. Some of the artifacts are listed in the section. Techniques that are used to remove such artifacts from relevant data are effective on one such artifact. Techniques described below are proved to be effective for this work.

#### 2.3.2.1   Band Pass

The information required to perform motor imagery classification is present in the low frequency region i.e 30Hz, hence band-passing the signal between 5Hz and 40Hz removes the DC offset and irrelevant information. It also avoids the need for a notch filter at 50Hz or 60Hz to remove the electric power line noise.

#### 2.3.2.2   Spatial Filtering

The potential that is measured at an electrode is overlap or a linear combination of several sources in the brain, this is primarily due to volume conduction. Given $\S_i$ from n independent sources, the sum of all independent sources can be written as 2.1.

$$\mathbb{C}^j = \sum_{n=i}^{n} w_i^j x_i = \mathbb{W}^{\mathsf{J}} \mathbb{X} \tag{2.1}$$

where $\mathbb{C}^j$ refers to measurement from arbitrary channel $j$, $\mathbb{W}^{\mathsf{J}}$ represents the weight matrix for channel $j$ and $\mathbb{X}$ represents independent sources. Inherently brain signals are low in SNR, improving the SNR would enable increased classification accuracy. Spatial filtering achieves this by performing one of the following - enhance local activity, reduce noise across channels, dimensionality reduction, identify hidden sources,

find projections that maximize discrimination between different classes. The following techniques perform one of the above mentioned operations. Spatial Filters also help to invert measurement to the original source.

**2.3.2.2.1  Bipolar Referencing**  Re-referencing the measurement from the electrodes is one of the common methods that help to achieve better SNR.

For a primitive system, Bipolar re-referencing will highlight the required features by finding the potential difference between two electrodes of interests. Consider measurement from channels $i$ and $j$

$$\mathbb{C}^i_{Bi} = \mathbb{C}^i - \mathbb{C}^j \tag{2.2}$$

**2.3.2.2.2  Laplacian Referencing**  Laplacian re-referencing enhances local activity at the electrode of interest by subtracting the potential of the channel of interest from the potentials of the adjacent four channels $\theta$. This is achieved by removing the muscle activity from the electrode of interest.

$$\mathbb{C}^i_{LP} = \mathbb{C}^i - \frac{1}{4} \sum_{i \epsilon \theta} \mathbb{C}^i \tag{2.3}$$

**2.3.2.2.3  Common Average Referencing (CAR)**  It is the most common method of re-referencing. It is very similar to Laplacian re-referencing, but instead of the adjacent electrodes, the average of potentials of all the electrodes is subtracted. However problems can be caused in calculation of average of electrode due to finite sample density and inadequate head coverage of EEG [8].

$$\mathbb{C}^i_{CAR} = \mathbb{C}^i - \frac{1}{N} \sum_{n=1}^{N} \mathbb{C}^i \tag{2.4}$$

**2.3.2.3  Independent Component Analysis (ICA)**

The potentials recorded using EEG electrodes appear the same and it is highly correlated. This measurement is highly redundant as it doesn't provide any distinctive fea-

tures that could enable efficient classification. Principal Component Analysis (PCA) enables dimensionality reduction that helps to remove the redundant information by finding the dimensions of highest variance, also called as principal components and transforming the measured data to the directions of N desired principal components that contains the most valuable information. With reduced dimensionality the classifier can be trained effectively as the principal components serve as feature vectors as they have the necessary information. Though the transformed data is decorrelated, the higher orders of the data are not independent, i.e. the transformed data is not completely independent. For analysis of brain signals, the measured data is assumed to be a linear combination of statistically independent signals from different regions in the brain, hence complete independence of the data is essential.

Consider $\mathbb{X}$ independent sources, the measurement $\mathbb{C}$ at the channels is mixed up by mixing matrix $\mathbb{M}$ 2.5.

$$\mathbb{C} = \mathbb{M}\mathbb{X} \tag{2.5}$$

ICA helps to recover the independent sources by finding the unmixing matrix $\mathbb{M}^{-\mathbb{K}}$ 2.6 in case where the $n$ independent sources and $N$ channels are the same.

$$\mathbb{X} = \mathbb{M}^{-\mathbb{K}}\mathbb{C} \tag{2.6}$$

However the independent sources $\mathbb{X}$ obtained from ICA can be less than, more than or equal to the number of measured channels and not orthogonal to each other unlike PCA. It is computationally efficient and the independent components serve as feature vector for classification or removes the artifacts from the measurement.

### 2.3.2.4   Signal Space Projection (SSP)

SSP helps in removing the artifacts in the signal by estimating a projection matrix based on measurements with and without signals of interest. The measurements are first taken without a subject to determine the directions of the noise vectors and the matrix formed by the noise vectors is used to project the measurement

onto the hyperplane orthogonal to the noise vectors to remove the noise from the measurement data. Noise reduction leads to loss of dimensions however it is relatively less compared to original signal space.

### 2.3.3 Epoch

After preprocessing the measurement, the data that is stored in the RAW format cannot be used further for feature extraction or classification as it consists of information relating to all the classes, hence the data has to be broken down into segments with a specific time bound before and after the event, grouped together based on the class. This process is referred to as epoching. This enables extraction of distinctive features that is most prevalent in all the segments in a particular class. The data after epoching is of the shape $Ep \times N \times T$, where $Ep$ is the number of epochs that is equal to the number of events, $N$ channels and $T$ epoched time points which is roughly between $-8$ and $4$ with the event trigger at the $0$ mark.

The frequency spectra of brain signals commonly have the $\frac{1}{f}$ structure, where the low frequency components dominate the results of the analysis. Power normalization resolves this by removing the $\frac{1}{f}$ and it is performed with the help of baselining the segments. Baseline is defined as a time interval usually before the occurrence of the event.

Power normalization is the ratio of activity (TF power) to baseline (TF power for a specific time window) 2.7.

$$10 \log_{10}(\frac{activity}{baseline}) \tag{2.7}$$

Apart from power normalization, baselining helps in separating task from background activity, normally distributes the data and makes it easier to compare the values across individuals and frequencies.

### 2.3.4 Feature Extraction

Several features such as temporal, statistical...can be extracted from noise-free epoched data. Different methods of extraction and analysis is detailed in section 3.1.

### 2.3.5 Feature Selection

With increasing number of channels and extraction of various features, the feature space is increasingly rich that can provide accurate results. However this increases the dimensionality of the feature space making it computationally expensive therefore lagging responses and over-fitting. Feature selection techniques helps in identifying the most discriminative features between the classes. [9] discusses about feature selection methods such as Minimum-redundancy and Maximum -relevance method that selected features such as band power, wavelet energy and kurtosis, Lasso regularization that selected features such as band power, wavelet energy and auto-regressive coefficients.

As this work deals only with 16-channel EEG system and less number of features, feature selection was not essential.

### 2.3.6 Classification

Machine Learning plays a predominant role in classification of the selected features. The models are trained on the selected features during the calibration and training phase. During online trials, the pre-trained classifiers are used in classification. Some of the commonly used classifiers are Support Vector Machines and Linear Discriminant Analysis, this is discussed further in section 3.2.

## Summary

From here the clean data is fed to conventional approaches and deep learning approaches separately.

# Chapter 3

# Signal Processing

## Introduction

The data after preprocessing is clean from noise and other artifacts, now the most informative and distinguishing features are extracted for classification. This chapter discusses about the conventional signal processing pipeline that is widely researched in the BCI community. These methods are proven to be robust and applied to many existing BCI systems.

Researches such as [5] reviews various SOTA signal processing pipelines for MI EEG based BCI systems. It discusses about different feature extraction, selection and classification techniques, and further about the benefits and challenges using them.

[10] studies integrating vehicle control systems (VCS) with 96-channels intra-cortical BCI and validated the concept with CARLA. The BCI signals were decoded and translated into acceleration, deceleration, turn right and turn left using a Python API. To demonstrate the potential of BCI for motion control of a car, 1:10 scale remote controlled car was driven with the same commands offline and prove it to work. Several performance statistics were considered in evaluating the proposed methodologies.

[11] studies the integration of BCI systems with ROS-based control algorithms for navigation and obstacle avoidance to ease cognitive workload. The proposed algorithm is applied to a semi-autonomous navigation system where the BCI system decodes EEG commands to take a left or right turn and the direction information is

used by ROS navigation stack to localize and move the robot with information also from different sensors and the provided static map.

## 3.1 Feature Extraction

Brain signals are transient as well as diffuse spike and action potential. It is rhythmic, hence could be sinusoidal or non-sinusoidal and includes small oscillations of semi periodic activation. Therefore the local information exists in time or frequency domain or both.

The brain signals could be time-locked i.e. the activity in the brain is time synchronized with the stimulus or the temporal dynamics is the same across every trials at the same instance. It could be phase-locked as well i.e. the phase angle time series is the same or similar on every trial. Depending on the type of the experiment used to gather data and the EEG paradigm, these parameters influence the choice of feature extraction mechanism.

### 3.1.1 Time Domain Analysis

Time domain analysis is useful in extracting temporal features in the preprocessed measurement. Some of the commonly used time domain analysis methods include signal averaging, Hjorth parameters, fractal dimensions, auto regressive models, Bayesian filtering, Kalman filtering, particle filtering, zero crossing, template matching, power and window detection.

In case of non-phase locked brain signals, Time domain analysis leads to loss of information on averaging several trials but frequency analysis doesn't lose information on averaging trials.

### 3.1.2 Frequency Domain Analysis

Transforming the preprocessed data into frequency domain provides more useful information about the phase and amplitude of the brain signal required for analysis. Imagined movements leads to oscillations in premotor and sensorimotor areas which

is observed as amplitude changes in the $\mu$, $\beta$ and $\gamma$ region. These features can be used to create feature vectors. Some of the common frequency domain analysis are Fourier transform, Short-Time Fourier transform and Welsh transform. These methods provide information on the spectral contents of the signal but not the temporal information of the event.

#### 3.1.2.1 Fourier Transform (FT)

Fourier transform decomposes signal into a weighted sum of sine and cosine waves of different frequencies. It works on the assumptions that the signal is infinite, periodic and stationary. Brain signals do not hold to these assumptions. These shortcomings can be over come by using short window periods for analysis, termed as Short-Time Frequency Transform (STFT) . Another commonly used method for frequency domain analysis is Welsh transform that provides smooth spectral decomposition. Phase content is not used in brain signal analysis in most cases as the signal could be non-phase locked. Hence amplitude information is mostly used in feature extraction.

Power spectrum 3.1 obtained by square of amplitude in different frequency components during course of the task is the feature of interest in most BCI experiments.

$$P(n) = A(n)^2 \tag{3.1}$$

where $A$ is the amplitude spectrum, $P$ is the Power Spectrum and $n$ denotes the index of the content.

### 3.1.3 Time-Frequency Analysis (TFA)

The method discussed above do not provide temporal information on occurrence of the event. Time-Frequency analysis on the signal provides temporal-spectral information. One of the most commonly used TFA method is wavelet transform.

#### 3.1.3.1 Wavelet Transform (WT)

Wavelet transform uses finite basis functions called wavelets that are scaled and translated copies of finite length waveform called mother wavelet. Wavelets divide

the signals of interest into different frequency components, each component can be studied at a resolution matched for its scale. A large scale component produces coarse resolution and small scale component produces fine resolution. The wavelet transform provides lower frequency resolution and higher temporal resolution at higher frequencies whereas higher frequency resolution and lower temporal resolution at lower frequencies.

Wavelet analysis is performed by convolution of wavelets of different scale and resolution with the signal of interest. The result of convolution contains components of the signal that are in the same frequency of the wavelet used. Scalogram provides graphical way to interpret the time and frequency information of the signal. The transform could be either discrete or continuous depending upon the scale and translation parameters. Continuous Wavelet Transform (CWT) provides more resolution than the Discrete Wavelet Transform (DWT).

A generic equation for a wavelet $\psi$ used in CWT can be written as 3.2.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi(\frac{t-b}{a}) \tag{3.2}$$

where $t$ denotes time, $b$ is the shift parameter that slides the waveform along the time axis and $a$ is the scale factor that is used to scale the frequency of the wavelet.

Generally for a given wavelet function $\psi(\frac{t}{a})$ when $a > 1$ the mother wavelet dilates and it helps in capturing low frequency component and when $0 < a < 1$ it compresses the mother wavelet, capturing the high frequency components in the signal of interest. Thus $a \propto \frac{1}{f}$.

Wavelets have band-pass characteristics, with sum of all points being zero 3.3 and its equivalent frequency can be determined by 3.4.

$$\int \psi(t) = 0 \tag{3.3}$$

$$F_{eq} = \frac{C_f}{a\delta_t} \tag{3.4}$$

where $F_{eq}$ is the equivalent frequency, $C_f$ is the centre frequency, $a$ is the scaling factor and $\delta_t$ is the sampling interval.

A generic equation for a wavelet used in DWT can be written as 3.5.

$$\psi_{a,b}(t) = \frac{1}{2^{\frac{j}{v}}}\psi(\frac{n-m}{2^{\frac{j}{v}}}) \tag{3.5}$$

where $j$ is the scaling parameter, $v(>1)$ denotes voices per octave (logarithmic unit of ratio between two frequencies) and $m$ is the translational parameter.

Various types of wavelets are used in research and one of the commonly used wavelet is the Morlet Wavelet. It the product of complex valued sine wave and Gaussian. In frequency domain, The amplitude spectrum of the complex morlet wavelet is not symmetrical and the power spectrum is independent of the phase relation between the wavelet and signal. The complex morlet wavelet can be written as 3.6.

$$\psi_\lambda(t) = e^{i2\pi ft - 0.5(\frac{t}{\sigma})^2} \tag{3.6}$$

where $\sigma = \frac{n}{2\pi f}$, $n$ is the number of cycles, $\lambda$ denotes a wavelet with specific scale and resolution parameters, $f$ is the frequency of the complex sine wave used.

The equation 3.6 can also be written as 3.7

$$\psi_\lambda(t) = e^{i2\pi ft}e^{-4\ln(2)\frac{t^2}{h^2}} \tag{3.7}$$

A wavelet transform can be written as 3.8.

$$\mathcal{W}x = \int x * \psi_\lambda(t)dt \tag{3.8}$$

where $h$ is the full width half maximum (FWHM) parameter. With higher $n$, better temporal resolution and with lower $n$ better spectral resolution is achieved. Most algorithms work by starting with low number of cycles and gradually increasing to higher number of cycles. However the results of the wavelet convolution are easily

interpretable only for stationary signals within the FWHM of the wavelet. With the right design of the wavelet this can be achieved.

### 3.1.4  Wavelet Scattering Transform (WST)

The wavelets are translated and convoluted with the signal of interest, as the wavelets commutes with translations the resulting expression becomes translation covariant. i.e. Shifting the signal also shifts wavelet coefficients, therefore the comparison between translated signals become difficult. In order to make proper representation of the signal the transform has to be translation invariant, stable under deformation and offer good structural information of all frequencies.

By WST non-informative variability in the signal such as translation, rotation and scaling are discarded. It can be suited for any application, with more data a pre-trained model is used as a pipeline and for small datasets wavelet scattering initialization could be performed. WST yields representations that are translation invariant and stable against time warping deformations. This is achieved by subsequent non-linearity and averaging steps after convolving with the wavelet.

As EEG signals can be influenced and varied by various factors Electrocardiography(ECG) provides more stable signals that can be used for analysis. In [12], the author performs emotion recognition from ECG signals using WST for feature extraction and evaluated the system for several classifiers.

CWT is the basic building block for WST. CWT measures similarity of the signal with wavelets of varying frequency and scale at each point in time. For a given wavelet $\psi$, it is first shifted by time $\tau$, the similarity if found by computing the inner product $\langle x, \psi \rangle$. This is repeated for all $\tau$ and $\psi$.

Any linear operation which is translation invariant of a wavelet coefficient will result in zero, which is not informative. Hence a non-linear invariant operand is required. Modulus being a non-linear, optimal contractive operator, when applied on CWT as given in 3.9, removes the phase information in the signal and makes it invariant to translations, that gives a measure on sparsity of the wavelet coefficients. The coefficients obtained are called as unaveraged coefficients.

$$\mathcal{U}x = |\mathcal{W}x|$$

$$\int |x * \psi_\lambda(t)|dt = \|x * \psi_{\lambda_1}\|_1$$

<div align="right">(3.9)</div>

where $\lambda_1 = 2^j$ is the centre frequency of the first order wavelets, with $j$ scale factor.

The unaveraged coefficients when convolved again with a low-pass or scaling filter(also called father wavelet) $\phi$ as given in 3.10 performs temporal averaging of the signal. The coefficients obtained are called Scattering coefficients. It imposes stability against time warping deformations. This however removes high frequency information, but can be recovered by performing WST again on the unaveraged coefficients.

$$\mathcal{S}x = \mathcal{U}x * \phi$$

$$\mathcal{S}x(t, \lambda_1) = \|x * \psi_{\lambda_1}\|_1 * \phi$$

<div align="right">(3.10)</div>

The above steps are performed recursively (usually 3 orders) on the results on the unaveraged coefficients and the features are extracted at each level by convolving the results with lowpass filter. Scatter representations consists of 0,1,2 order coefficients which are generated by composing wavelets in different sequences. Multiple wavelets composed together called filter banks capture high frequency components. Filter bank consists of several dilated and rotated wavelets with no orthogonality.

Consider family of wavelets $\psi_\lambda$. The dilated wavelet can be written as 3.11

$$\psi_\lambda(t) = 2^{\frac{-j}{Q}} \psi(2^{\frac{-j}{Q}} t)$$

<div align="right">(3.11)</div>

where $Q$ is number of band-pass filter in an octave.

$$\mathcal{W}x = \{x * \phi(t), x * \psi_\lambda(t)\}$$

<div align="right">(3.12)</div>

The second order coefficients obtained can be decorrelated to increase their invariance

through renormalization. Different features are obtained at each order of coefficients. The first order coefficients represents the frequency content (oscillation rate) in the signal, the second order coefficients represents rates of frequency change, treat every rate from first order as its own signal and repeat first order upon each modulus.

### 3.1.4.1 Exponential decay of scattering coefficients

Repeated application of modulo moves the energy contained in the high frequencies of the initial signal towards low frequencies. Convolution with the low-pass filter picks that energy and used in feature extraction. At each step of the scattering transform, energy in the lowest frequency bands is output by convolution with low-pass filter at its respective step. The remaining part of the energy is shifted towards the lower frequencies by applying modulus of wavelet transform.

$$\mathcal{W}x = \{x * \phi(t), x * \psi_\lambda(t)\} \tag{3.13}$$

If $|\hat{\phi}(w)|^2 + \sum_\lambda |\hat{\psi}_\lambda(w)|^2 = 1$ then $\mathcal{W}$ is unitary.

$$\|\mathcal{W}x\|^2 = \|x * \phi\|^2 + \sum_\lambda \|x * \psi_\lambda\| = \|x\|^2 \tag{3.14}$$

### 3.1.4.2 Neural Network representation of WST

Typical Convolutional Neural Network (CNN) involves, successive operations involving convolution, non-linearity, pooling/sub-sampling. WST performs the same operations, hence WST can be represented as a neural network called Wavelet Scattering Network (WSN) to extract the scattering coefficients. This can be highly efficient as the wavelets values for the filter bank is predetermined hence learning is not required and no complex architecture has to be designed. Kymatio is an Open-source WST implementation that leverages machine learning framework like PyTorch, Keras or Scikit-learn in the backend.

### 3.1.5 Spatial Analysis

The spatial filters invert the measurement to its original source. Spatial Analysis helps in mapping the source signals to the brain topography that can be used for further topographical studies. In general Patterns are not Filters. Patterns denote the propagation of a source to sensors which is used in forward model 3.15. Filters on the other hand denote weighting of EEG channels and it is used in backward model 3.16 to extract individual components.

$$x(t) = As(t) + n(t) \tag{3.15}$$

where $s$ can be a vector denoting sources. $n$ is a noise vector and $x$ is the measurement at the channel. A is the forward or mixing matrix in which every column corresponds to a spatial pattern.

$$\hat{s}(t) = W^T x(t) \tag{3.16}$$

where $W$ represents the unmixing matrix where every row of $W^T$ corresponds to spatial filter.

The spatial map of a filter is hard to interpret, but patterns are easy to interpret and hence patterns corresponding to a given spatial filter is found.

#### 3.1.5.1 Common Spatial Patterns (CSP)

CSP is a supervised feature extraction method for classification algorithms which works only for motor-imagery information in EEG. It learns to optimally discriminate band power features using multiple electrodes placed at optimal regions. Variance of the band pass signal is proportional to the band power of the frequency band. CSP leverages this by finding patterns where the variance of filtered data from one class is maximized while variance of filtered data from other class is minimized. The resulting feature vector enhance discriminability between different classes. However it works on the assumptions that frequency band and time window of the measurements are

known, source activity constellation differs between two classes and the band passed signal is jointly Gaussian within the time window.

[13] studies different variants of CSP that could achieve maximum accuracy with least number of EEG channels. This helps removing noisy channels and avoid over-fitting. It revealed that a conventional CSP without any regularization applied performs better on cross-subject validation. [9] proposed motor imagery classification pipeline using channel selection, band passing and CSP based extraction of 38 features and Gaussian Naive Bayes (GNB) classifier.

Consider measurement data set $\{\mathbb{X}_c^i\}_{i=1}^k$ for trial $i$ belonging to class $c \in \{1, 2\}$. Each $\{\mathbb{X}_c^i\}$ is a $N * T$ matrix, where N is the number of channels and T is the number of samples per channel. The goal of CSP is to find $W$ given by $N * M$, consisting M spatial filters i.e. each column is a spatial filter, that transforms the measured signals 3.17.

$$x_{CSP}(t) = W^T x(t) \tag{3.17}$$

where $x(t)$ is vector of input signals at time $t$ from all channels. CSP can formally be defined for a two class problem as 3.19.

$$
\begin{aligned}
J(w) &= \frac{W X_1 X_1^T W^T}{W X_2 X_2^T W^T} \\
&= \frac{W C_1 W^T}{W C_2 W^T}
\end{aligned}
\tag{3.18}
$$

where $C_c$ is the spatial covariance matrix for class $c$, $W$ is the spatial filter to optimize and $X_c$ is the multichannel EEG signal from class $c$. The problem could be solved to either minimize $J(w)$ where variance in $X_1$ is minimized and maximized in $X_2$ or maximize $J(w)$ where variance in $X_1$ is maximized and minimized in $X_2$.

This can be computed by Generalized Eigen Value Decomposition (GEVD) of $C_1$ and $C_2$. The Eigen vectors corresponding to the largest eigen value will maximize $J(w)$. and eigen vectors corresponding to the smallest eigen value will minimize $J(w)$.

$$WC_1W^T = \Lambda_1$$
$$WC_2W^T = \Lambda_2$$

<div align="right">(3.19)</div>

where $\Lambda_1$ and $\Lambda_2$ are diagonal matrices containing eigen values such that $\Lambda_1 + \Lambda_2 = \mathbf{I}$

$$WC_1W^T = \Lambda W^T(C_1 + C_2)W = \mathbf{I}$$

<div align="right">(3.20)</div>

Typically 3 CSP filter pairs i.e. minimization and maximization are used which results in 6 feature vectors that forms the filter matrix $W$. Once filter is obtained, the prediction function using CSP can be defined as 3.21.

$$y = sign(\theta log(var(WX)) + b)$$
$$= sign(\theta log(WCW^T) + b)$$

<div align="right">(3.21)</div>

The application of log on the spatially filtered data makes it a Gaussian distribution. The classifier applied can be either linear or nonlinear. After applying CSP the data is rotated and placed orthogonal.

CSP is simple, computationally efficient with high classification performance. But it is not robust to noise and non-stationaries, prone to over-fitting and requires many training examples. Several variants of CSP are researched and developed such as Filter Bank CSP, Regularized CSP, Invariant CSP... to overcome the disadvantages of vanilla CSP.

## 3.2 Classification

With the required features extracted, the vectors can be forwarded to the classifier as a last step in the BCI pipeline. Some of the most commonly used classifiers include SVM, LDA... These could also be neural networks, but neural network could also perform feature extraction, hence it would not be efficient to use neural network just for classification. However several literature have used neural networks only as a

classifier. [14] discusses the machine learning methods, their application to BCI, and validating them for BCI. It discusses about SVM and how regularization is beneficial for BCI data. In case of nonlinear data, it can be converted to a linearly separable feature space and efficiently used for classification without any complexity. [15] discusses about how machine learning is applied to motor imagery with feedback.

### 3.2.1 Support Vector Machines (SVM)

SVM is the most commonly used classifiers as they provide highly accuracy with less compute power. The aim of SVM as a classifier is to find the hyperplane such that it separates the support vectors with a high margin.

### 3.2.2 Linear Discriminant Analysis (LDA)

LDA is considered to be optimal classifier given the following assumptions are true:

1. Features of each class are Gaussian distributed.

2. Gaussian's of all classes have the same covariance matrix.

3. True class distributions are known.

Given two class distributions $\mathcal{N}(\mu_1, \epsilon)$ and $\mathcal{N}(\mu_2, \epsilon)$ , then LDA is defined by normal vectors

$$w = \epsilon^{-1}(\mu_2 - \mu_1) \tag{3.22}$$

and

$$b = w^T(\mu_2 + \mu_1)/2 \tag{3.23}$$

But as the true class distribution is not known, the distributions are estimated, which is usually different from the true value and it determines the performance of the classifier. Shrinkage-LDA overcomes this problem by shrunk covariance matrix

that is determined by optimal shrinkage parameter given by analytical and cross-validation methods.

# Summary

# Chapter 4

# Deep Learning

## Introduction

Deep learning approaches are used in several industries and are replacing many existing algorithms as they are inherently adaptive to changes in trends as observed in the data. In BCI, the signal processing methods used are tailored for very specific application or EEG paradigm, EEG signals are exposed to various artifacts that needs to be eliminated separately and requires expertise in relevant domain for feature engineering. Deep Learning can help researchers to elevate this problem by building representations from input data with or without prior knowledge of ground truth. Deep learning models can be broadly classified as Discriminative, Representative, Generative and Hybrid models. This work focuses on classifying the motor intent of the user online, with a system pre-trained on data obtained in the same session, hence discriminative models are used. Discriminative models can be further classified into Recurrent Neural Network (RNN) and Convolutional Neural Network. These networks are capable of extracting different information from the EEG data which serves as feature vectors. [16] and [3] provides in-depth survey on methods and applications of various deep learning techniques used in research and industries.

In [17], the authors review more than hundred publications that apply DL to BCI. It reveals that inter subject studies are the most researched, CNNs are most commonly used approach for EEG signal analysis and most of the publications work on the raw EEG data without any preprocessing. [18] reviews various advancements and applications of machine learning in BCI. [19] developed a single system for diagnosing

multiple neurological disorder. DWT were used to separate the EEG signals into sub-bands and then statistical features were extracted from each sub-band, which is then later used in the classifier.

## 4.1  Spatial Information

Convolutional Neural Networks are very often applied in machine vision systems, used in extracting different spatial features in the image. A typical CNN includes layers such as input layer, convolutional layer, pooling layer, fully connected layer, output layer and activation layer in order. Batch-norm layer can also be used between convolution layer and pooling layer which provides uniformly distributed data and avoiding over or undershoot of the activation. Dropout layers are generally used between stacks of above mentioned layers and in the fully connected layer that provides regularization effect to the network and prevents over-fitting. The spatial information is extracted as features using a series or stacks of layers between input and the last pooling layer. Classification is performed in the fully connected layers.

[20] uses Deep Convolutional Neural Networks (DCNN) to classify motor imagery tasks. The raw EEG data is preprocessed to remove any artifacts and time-frequency (TF) information from the noise-free EEG data is extracted using time-frequency transforms such as STFT and CWT. The TF information is forwarded to a DCNN for further feature extraction and classification. The authors used transfer learning on pre-trained AlexNet to adapt the BCI EEG data. [21] studies motor-imagery EEG signal classification using SVM and DCNN. It also revealed that data preprocessing for artifact removal influences the classification results drastically.

[22] proposed a new DCNN architecture that extracts low level features and classifies the motor intent from the time-frequency information in motor imagery EEG data. It is found that the accuracy of the classification is highest in the very start of the experiment and gradually dropping down to the end of experiment.

[23] developed a Computer Aided Diagnosis(CAD) method to detect Autism spec-

trum disorder using DWT to decompose EEG signals into approximations and details coefficient for each EEG sub-band, extract statistical features such as mean, variance, skewness, kurtosis...or using Shannon entropy values and, classify using Artificial Neural Network (ANN). The ANN structure used was a simple FCNN with one hidden layer. The study found that using DWT and Shannon entropy achieved better accuracy.

[24] proposed a CNN model for a four class EEG MI dataset and compared it to LDA based classifier and results showed that CNN performed better in extracting and classifying EEG data. The CNN uses mean of specified number of samples as the filter, that slides to get a running average of the EEG data for each channel.

[25] studied the classification performance on a 2 class data using CNNs and also conventional signal processing pipeline that uses features such as power, CSP or auto-regression coefficients. The results showed that the CNNs performed better than the conventional signal processing pipeline. The authors proposed 5-layer CNN model 4.1 that could observe the spatial and temporal information using a column and row filters in individual layers.
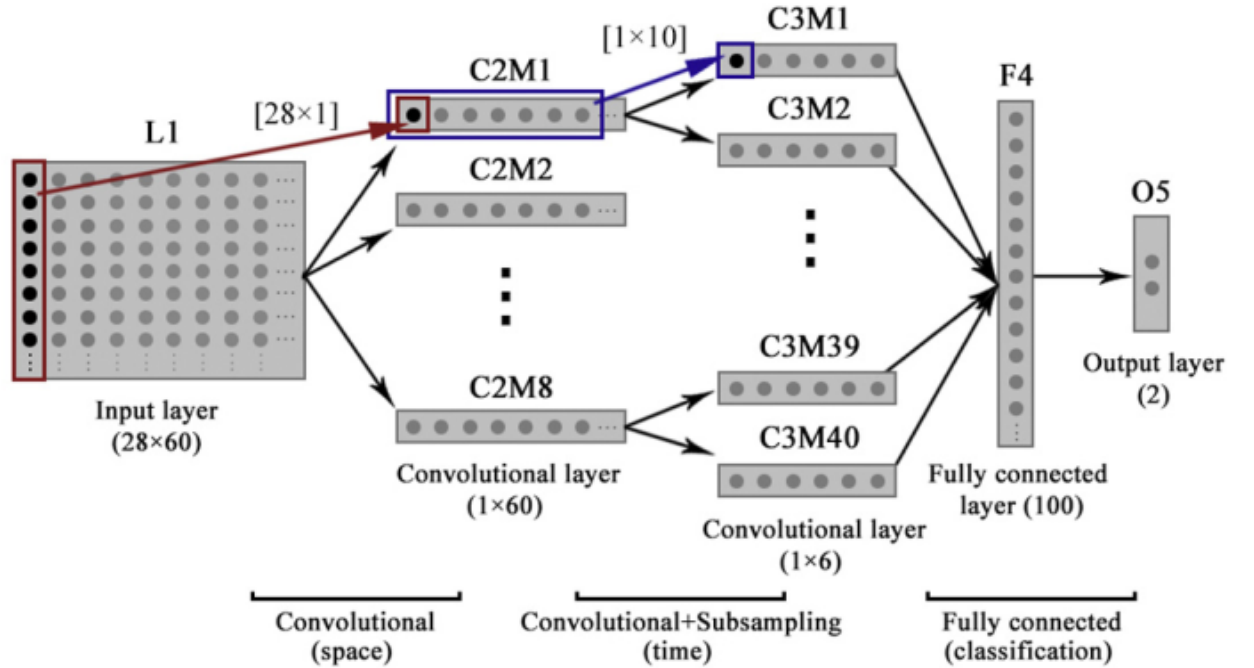


Figure 4.1: CNN architecture proposed by [25](source)

## 4.2 Temporal Information

Recurrent Neural Networks are vastly used in time series applications. They are inherently good at capturing dynamic information in a serial data. RNNs have the problem of vanishing or exploding gradients which are overcome by replacing RNN nodes by LSTM cells. LSTM cells have a more complex internal structure that enables to remember data over a long or short time. A typical LSTM cell is shown in figure 4.2.
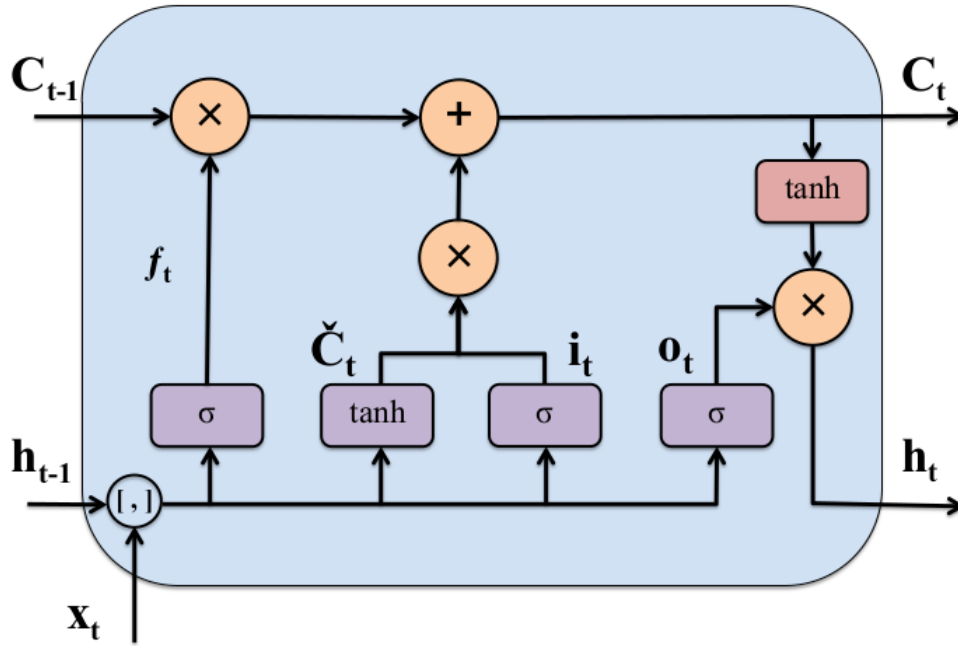


Figure 4.2: A LSTM cell structure. Source [26]

Attention mechanism is generally used at the end of an RNN model to find and weight importance of each discriminative feature, that assures high classification scores. It is used in Natural Language Processing (NLP) [26] to find the contribution of each word to improve the scores. It amplifies the result by aggregating the hidden states and weighting their relative importance.

[27] proposed a solution for classification of hand movements from EEG using deep attention-based LSTM network by first extracting time and frequency domain features from the EEG signals and passing them to the LSTM input layer. The attention layer captures the importance of the EEG information varying through time where

discriminative information with higher importance are assigned higher scores, which results in higher classification results. Time and frequency domain features extracted are listed in 4.3. The last hidden state of the LSTM is multiplied with trainable weights to capture more discriminative task related features which forms the attention mechanism. The network uses 297 features from 7 time steps in each of the 2 second time segment. The entire architecture is given in the figure 4.4.

TABLE II
TIME AND FREQUENCY DOMAIN FEATURES

| Method | Formula |
|---|---|
| Mean | $\mu = \dfrac{1}{N} \sum_{i=1}^{N} x_i$ |
| Variance | $\sigma^2 = \dfrac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$ |
| Skewness | $S = \dfrac{\dfrac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^3}{(\dfrac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu)^2)^{3/2}}$ |
| Kurtosis | $K = \dfrac{\dfrac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^4}{(\dfrac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2)^2} - 3$ |
| Zero-crossing | $zc = \sum_{i=1}^{N-1} \mathbb{1}_{\mathbb{R}_{<0}}(x_i x_{i-1})$ |
| Absolute area under signal | $simps = \int_a^b |f(x)| dx$ |
| Peak to Peak | $pk2pk = max(\mathbf{x}) - min(\mathbf{x})$ |
| Amplitude spectrum density | $\hat{X}(\omega) = \dfrac{1}{\sqrt{T}} \int_0^T x(t) \exp^{-i\omega t} dt$ |
| Power spectrum density | $S_{xx}(\omega) = \lim_{T\to\infty} E\left[|\hat{X}(\omega)|^2\right]$ |
| power of each frequency band | $P = \dfrac{1}{\pi} \int_{\omega_1}^{\omega_2} S_{xx}(\omega) d\omega$ |

Figure 4.3: The table of features used by[27](source)

## 4.3 Spatio-Temporal Information

[28] introduced cascade and parallel convolutional recurrent neural network to learn the spatio-temporal dynamics of EEG data. The method involves converting 1-dimensional EEG sequences to 2-dimensional EEG meshes, the values 2D meshes corresponds to the amplitude from each electrode. At time $\sqcup$ the 1D data from all the $n$ electrodes $r = \S_i, i \epsilon n$ is converted to 2D mesh with $n$ points and the vacant spaces are set to zero. The 2D meshes are created for each time instance $\sqcup$, for a specific time interval $\int$ and passed into a cascaded convolutional recurrent neural network 4.5.
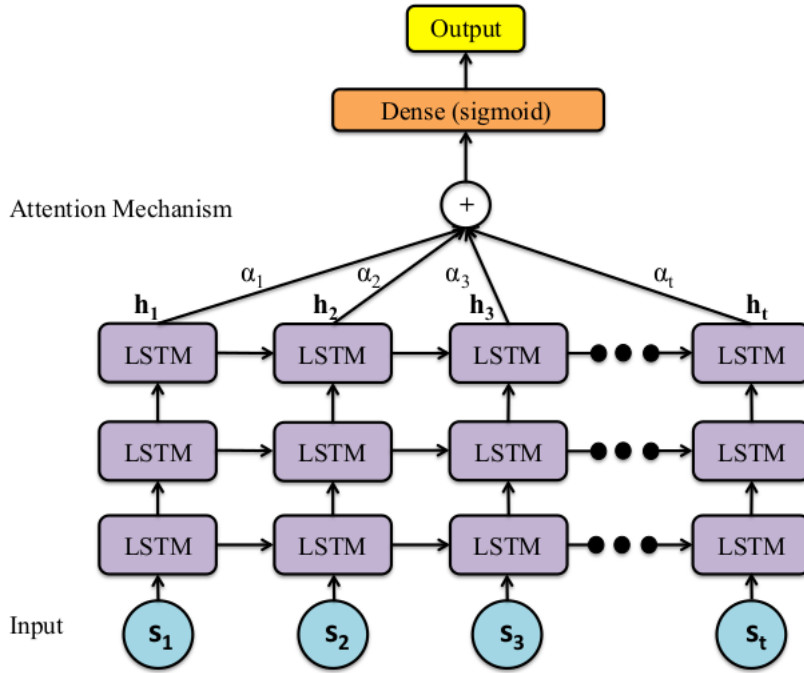
Figure 4.4: A LSTM based architecture proposed by [27](source)

The network first performs spatial feature extraction by a series of $\int$ parallel convolution layers followed by a linear layer that fattens the information. The temporal features are extracted from the data by passing the flattened data to the LSTM layers and the output of the last LSTM cell is used to predict the movement intention using softmax as activation layer. The [28] also suggests using CNNs and RNNs separately and later concatenate the outputs of the last layers to predict the motor intention of the user using softmax activation layer.

[29] provides systemic review of several Hybrid Deep Learning models used in BCI research. The study reveals that CNN-RNN hybrid deep learning architectures with Adam optimizer are the most widely researched for extracting spatial-temporal features from the EEG data and the average accuracy is 80 percentage across several datasets.

[30] extracts domain-invariant multi-level spatial-temporal features to tackle domain differences. This is done by minimizing the source and target distribution distance.

While the above techniques proposed 2D convolutional neural networks, [31] proposed 1D CNN-LSTM architecture for automatic recognition of epileptic seizures
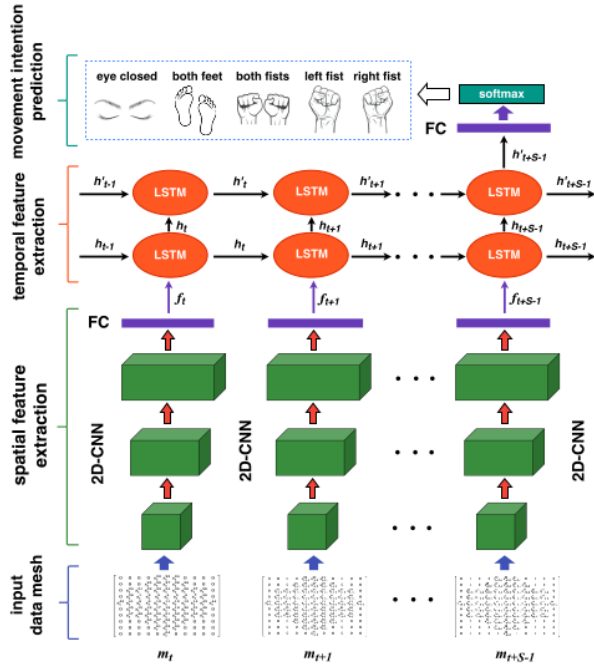
Figure 4.5: A Cascaded convolutional recurrent neural network based architecture proposed by [28](source)

through EEG signal analysis. Here the EEG data is preprocessed and normalized before the high level spatial information is extracted by the 1D-CNNs and the temporal features by the LSTM layer figure 4.6.
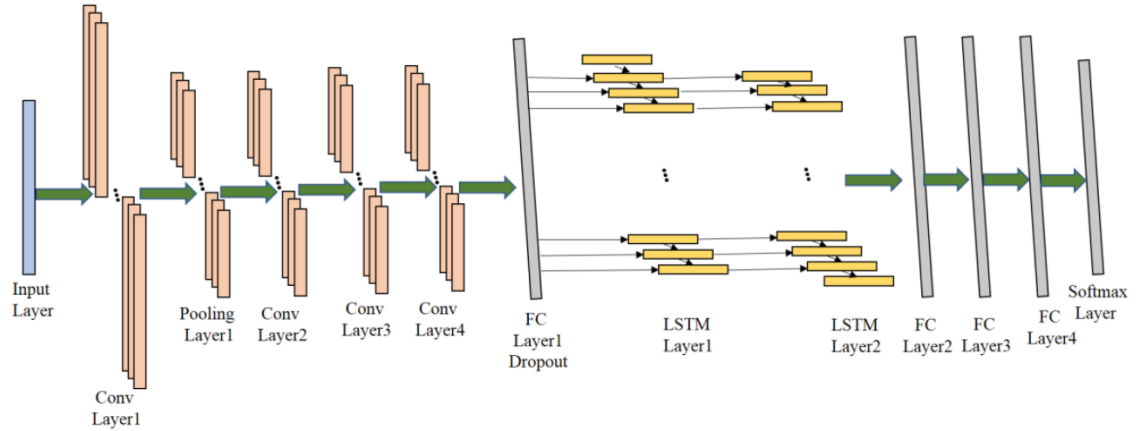


Figure 4.6: A CNN-LSTM architecture proposed by [31](source)

# Summary

# Chapter 5

# Comparison and Benchmark

Introduction

5.1    Spatial Information

5.2    Temporal Information

5.3    Spatio-Temporal Information

Summary

# Chapter 6

# Challenges and Conclusions

Introduction

6.1   Challenges

6.2   Conclusion

Summary

# Chapter 7

# Reflective Analysis

# References

[1]  A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez and V. Koltun, 'CARLA: An open urban driving simulator,' in *Proceedings of the 1st Annual Conference on Robot Learning*, 2017, pp. 1–16.

[2]  A. Gramfort *et al.*, 'MEG and EEG data analysis with MNE-Python,' *Frontiers in Neuroscience*, vol. 7, no. 267, pp. 1–13, 2013. DOI: `10.3389/fnins.2013.00267`.

[3]  X. Zhang and L. Yao, *Deep Learning For Eeg-based Brain-computer Interfaces: Representations, Algorithms And Applications*. World Scientific Publishing Company, 2021, ISBN: 9781786349606. [Online]. Available: `https://books.google.de/books?id=7YdGEAAAQBAJ`.

[4]  Wikipedia contributors, *10–20 system (eeg) — Wikipedia, the free encyclopedia*, [Online; accessed 11-December-2022], 2022.

[5]  N. Padfield, J. Zabalza, H. Zhao, V. Masero and J. Ren, 'Eeg-based brain-computer interfaces using motor-imagery: Techniques and challenges,' *Sensors*, vol. 19, no. 6, 2019, ISSN: 1424-8220. [Online]. Available: `https://www.mdpi.com/1424-8220/19/6/1423`.

[6]  S. G, M. DJ, H. T, B. N and W. JR., *Bci2000: A general-purpose brain-computer interface (bci) system*, 2004. DOI: `10.1109/TBME.2004.827072`.

[7]  G. Dornhege, B. Blankertz, G. Curio and K.-R. Müller, *Boosting bit rates in non-invasive eeg single-trial classifications by feature combination and multi-class paradigms*, 2004.

[8]  M. Iftikhar, S. A. Khan and A. Hassan, 'A survey of deep learning and traditional approaches for eeg signal processing and classification,' in *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, vol. 0, 2018, pp. 395–400. DOI: `10.1109/IEMCON.2018.8614893`.

[9]  S. Sreeja, J. Rabha, K. Nagarjuna, D. Samanta, P. Mitra and M. Sarma, 'Motor imagery eeg signal processing and classification using machine learning approach,' in *2017 International Conference on New Trends in Computing Sciences (ICTCS)*, 2017, pp. 61–66. DOI: `10.1109/ICTCS.2017.15`.

[10] C. Dunlap *et al.*, 'Towards a modular brain-machine interface for intelligent vehicle systems control – a carla demonstration,' Oct. 2019, pp. 277–284. DOI: `10.1109/SMC.2019.8914317`.

[11] G. Beraldo, M. Antonello, A. Cimolato, E. Menegatti and L. Tonin, 'Brain-computer interface meets ros: A robotic approach to mentally drive telepresence robots,' *2018 IEEE International Conference on Robotics and Automation (ICRA)*, May 2018. DOI: `10.1109/icra.2018.8460578`. [Online]. Available: `http://dx.doi.org/10.1109/ICRA.2018.8460578`.

[12] A. Sepúlveda, F. Castillo, C. Palma and M. Rodriguez-Fernandez, 'Emotion recognition from ecg signals using wavelet scattering and machine learning,' *Applied Sciences*, vol. 11, p. 4945, May 2021. DOI: `10.3390/app11114945`.

[13] N. Masood, H. Farooq and I. Mustafa, 'Selection of eeg channels based on spatial filter weights,' in *2017 International Conference on Communication, Computing and Digital Systems (C-CODE)*, 2017, pp. 341–345. DOI: `10.1109/C-CODE.2017.7918954`.

[14] K.-R. Muller, M. Krauledat, G. Dornhege, G. Curio and B. Blankertz, 'Machine learning techniques for brain-computer interfaces,' *IEEE Transactions on Biomedical Engineering*, 2004.

[15] G. Dornhege, J. del R. Millán, T. Hinterberger, D. J. McFarland and K.-R. Müller, 'The berlin brain-computer interface: Machine learning-based detection of user specific brain states,' in *Toward Brain-Computer Interfacing*. 2007, pp. 85–102.

[16] X. Zhang, L. Yao, X. Wang, J. Monaghan and D. McAlpine, 'A survey on deep learning based brain computer interface: Recent advances and new frontiers,' *CoRR*, vol. abs/1905.04149, 2019. arXiv: `1905.04149`. [Online]. Available: `http://arxiv.org/abs/1905.04149`.

[17] Y. Roy, H. Banville, I. M. Carneiro de Albuquerque, A. Gramfort, T. Falk and J. Faubert, 'Deep learning-based electroencephalography analysis: A systematic review,' *Journal of Neural Engineering*, vol. 16, May 2019. DOI: `10.1088/1741-2552/ab260c`.

[18] S. Rasheed, 'A review of the role of machine learning techniques towards brain–computer interface applications,' *Machine Learning and Knowledge Extraction*, vol. 3, pp. 835–862, Oct. 2021. DOI: `10.3390/make3040042`.

[19] F. A. Alturki, K. AlSharabi, A. M. Abdurraqeeb and M. Aljalal, 'Eeg signal analysis for diagnosing neurological disorders using discrete wavelet transform and intelligent techniques,' *Sensors (Basel, Switzerland)*, vol. 20, no. 9, E2505, Apr. 2020, ISSN: 1424-8220. DOI: `10.3390/s20092505`.

[20] S. Chaudhary, S. Taran, V. Bajaj and A. Sengur, 'Convolutional neural network based approach towards motor imagery tasks eeg signals classification,'

*IEEE Sensors Journal*, vol. 19, no. 12, pp. 4494–4500, 2019. DOI: `10.1109/JSEN.2019.2899645`.

[21] N. D. Skomrock *et al.*, 'A characterization of brain-computer interface performance trade-offs using support vector machines and deep neural networks to decode movement intent,' *Frontiers in Neuroscience*, vol. 12, 2018, ISSN: 1662-453X. DOI: `10.3389/fnins.2018.00763`. [Online]. Available: `https://www.frontiersin.org/articles/10.3389/fnins.2018.00763`.

[22] X. Xiao and Y. Fang, 'Motor imagery eeg signal recognition using deep convolution neural network,' *Frontiers in Neuroscience*, vol. 15, Mar. 2021. DOI: `10.3389/fnins.2021.655599`.

[23] A. Sivasangari, K. Sonti, G. P. Kanmani, Sindhu and D. Deepa, 'Chapter 12 - eeg-based computer-aided diagnosis of autism spectrum disorder,' in *Cognitive Systems and Signal Processing in Image Processing*, ser. Cognitive Data Science in Sustainable Computing, Y.-D. Zhang and A. K. Sangaiah, Eds., Academic Press, 2022, pp. 277–292, ISBN: 978-0-12-824410-4. DOI: `https://doi.org/10.1016/B978-0-12-824410-4.00010-6`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/B9780128244104000106`.

[24] H. Abdulkarim and M. Al-Faiz, 'Online multiclass eeg feature extraction and recognition using modified convolutional neural network method,' *International Journal of Electrical and Computer Engineering*, vol. 11, pp. 4016–4026, May 2021. DOI: `10.11591/ijece.v11i5.pp4016-4026`.

[25] Z. Tang, C. Li and S. Sun, 'Single-trial eeg classification of motor imagery using deep convolutional neural networks,' *Optik*, vol. 130, pp. 11–18, 2017, ISSN: 0030-4026. DOI: `https://doi.org/10.1016/j.ijleo.2016.10.117`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0030402616312980`.

[26] Y. Zhang, J. Wang and X. Zhang, 'YNU-HPCC at SemEval-2018 task 1: BiLSTM with attention based sentiment analysis for affect in tweets,' in *Proceedings of the 12th International Workshop on Semantic Evaluation*, New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 273–278. DOI: `10.18653/v1/S18-1040`. [Online]. Available: `https://aclanthology.org/S18-1040`.

[27] G. Zhang, V. Davoodnia, A. Sepas-Moghaddam, Y. Zhang and A. Etemad, 'Classification of hand movements from eeg using a deep attention-based lstm network,' *IEEE Sensors Journal*, vol. 20, no. 6, pp. 3113–3122, 2020. DOI: `10.1109/JSEN.2019.2956998`.

[28] D. Zhang, L. Yao, X. Zhang, S. Wang, W. Chen and R. Boots, 'Eeg-based intention recognition from spatio-temporal representations via cascade and parallel convolutional recurrent neural networks,' *CoRR*, vol. abs/1708.06578, 2017. arXiv: `1708.06578`. [Online]. Available: `http://arxiv.org/abs/1708.06578`.

[29]     N. Abo Alzahab *et al.*, 'Hybrid deep learning (hdl)-based brain-computer interface (bci) systems: A systematic review,' *Brain Sciences*, vol. 11, p. 75, Jan. 2021. DOI: `10.3390/brainsci11010075`.

[30]     W. Xu, J. Wang, Z. Jia, Z. Hong, Y. Li and Y. Lin, 'Multi-level spatial-temporal adaptation network for motor imagery classification,' in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 1251–1255. DOI: `10.1109/ICASSP43922.2022.9746123`.

[31]     G. Xu, T. Ren, Y. Chen and W. Che, 'A one-dimensional cnn-lstm model for epileptic seizure recognition using eeg signal analysis,' *Frontiers in Neuroscience*, vol. 14, 2020, ISSN: 1662-453X. DOI: `10.3389/fnins.2020.578126`. [Online]. Available: `https://www.frontiersin.org/articles/10.3389/fnins.2020.578126`.