UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA MAGISTRALE IN
…Control System Engineering…

"Cognitive State Classification In VR: The Relation Between Subjective Quality Of Experience, Movement And Video Quality In Virtual Reality"

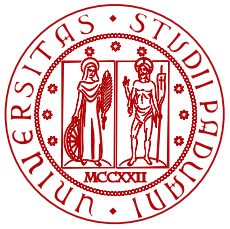
Relatore: Prof. / Dott …Luca Schenato…

Laureando/a: …Yeganeh Jabbari...

Correlatore: Prof./Dott…Klass Bombeke…

ANNO ACCADEMICO: 2023 – 2024

Data di laurea : 7th March 2024



Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

Master Thesis in Coontrol System Engineering

Cognitive State Classification In VR: The Relation Between Subjective Quality Of Experience, Movement And Video Quality In Virtual Reality

Master Candidate

Yeganeh Jabbari

Student ID 2006038

Supervisor

Prof. Luca Schenato

University of Padova

Co-supervisor

Prof. Klaas Bombeke

imec-mict-UGent, Ghent University

Academic Year
2023/2024

Expressions of Gratitude

I extend my deepest gratitude to those who have made this thesis possible. First and foremost, my heartfelt appreciation goes to my family. To my parents for their unwavering love and support throughout my academic journey and to my sister Niki, whose encouragement and belief in my abilities have been a constant source of strength.

I am profoundly grateful to Professor Luca Schenato at the University of Padova for his guidance and invaluable insights, which have significantly enriched my research experience. His mentorship has been instrumental in my academic development.

Special thanks are due to my esteemed supervisor, Professor Klaas Bombeke, whose expertise and attention to detail have greatly enhanced the quality of my work. His patience and willingness to educate me, even when I took missteps, have been truly inspiring. I am also indebted to Dott. Aleksandra Zheleva and Dott. Jonas De Bruyne at Ghent University and imec, for their constructive feedback, unwavering support, and wise counsel. Their commitment to excellence and innovation has profoundly influenced my growth as a researcher.

I also express my deep appreciation for my friend Sepideh, whose support, camaraderie, and unwavering faith in me have been of immense comfort during this challenging but rewarding journey.

To all, I express my sincere appreciation and gratitude for accompanying me on this scholarly voyage. Your support has been a beacon of light in moments of doubt and has consistently guided me toward realizing my academic goals.

Abstract

This thesis investigates the impact of varying video resolutions in Virtual Reality (VR) on user experience and cognitive load, employing advanced techniques like electroencephalography (EEG) and machine learning (ML). Grounded in the pioneering work of Zheleva et al. (2020) on the Quality of Experience (QoE) in VR, the study methodically examines four levels of video quality, from high to low resolution, and their effects on user perception and cognitive processing.

By integrating objective electroencephalography (EEG) measurements and subjective user feedback, the research analyzes how different video qualities are experienced in virtual reality regarding their effect on sensory immersion, simulator sickness, and narrative engagement of the user. Employing sophisticated ML algorithms, including Multi-Layer Perceptron (MLP) and Long Short-Term Memory (LSTM) models, the thesis aims to identify EEG markers that indicate cognitive load variations in response to these video quality changes.

This work advances our understanding of VR's technological and cognitive dimensions. It contributes to the broader cognitive science and digital media fields, highlighting the nuanced interplay between virtual environments and human cognitive processes.

Sommario

Contents

List of Figures

List of Tables

List of Algorithms

List of Code Snippets

List of Acronyms

CSV Comma Separated Values

VR Virtual Reality

VE Virtual Environment

ML Machine Learning

DL Deep Learning

CLT Cognitive Load Theory

QoE Quality of Experience

EEG Electroencephalogram

AR Augmented Reality

AI Artificial Intelligence

HD High Definition

RFE Recursive Feature Elimination

MLP Multi-layer Perceptron

LSTM Long Short Term Memory

ANNs Artificial Neural Networks

RNN Recurrent Neural Network

TSP Time Series Prediction

PSD Power Spectral Density

CFS Correlation-based Feature Selection

ICA Independent Component Analysis

UX User Experience

ReLU Rectified Linear Unit

BN Batch Normalization

fMRI Functional Magnetic Resonance Imaging

HMDs Head-Mounted Displays

MCN Modified Combinatorial Nomenclature

ERP Event Related Potentials

LOOCV Leave One Out Cross-Validation

1 Introduction

1.1 Motivation

As virtual reality (VR) technology advances, users are fully submerged in virtual worlds that provide a powerful sensation of presence, surpassing traditional media such as video games and movies [59]. This technology stimulates several senses and frequently persuades the brain that the experience is authentic, beyond simple visual and auditory involvement. Combining virtual reality (VR) with artificial intelligence (AI) can potentially build environments more identical to reality, providing increasingly realistic experiences. To overcome obstacles like motion sickness [58] and improve head-mounted displays (HMDs) display resolution, constant innovation is essential to this path toward previously unattainable realism. These enhancements are crucial because they solve the common problems seen in poor virtual reality experiences, which can significantly reduce the sense of immersion and realism. By focusing on these aspects, VR can transcend its current limitations, offering an experience that is not only immersive but also comfortable and visually compelling for all users.

Therefore, the development of VR technology depends on our ability to comprehend the slight variations in visual resolution and how they affect the user experience. A central question of our research is whether Machine Learning (ML) and Deep Learning (DL) algorithms can effectively analyze empirical data across different resolutions to identify these variations and uncover user patterns and preferences. This inquiry is pivotal, as it explores the potential of ML and DL in enhancing the user experience (UX) research of VR, especially in optimizing VR

content for improved immersion, visual comfort, and overall content perception. Specifically, in VR streaming, particularly in super high-definition (HD), which demands significant bandwidth, our research examines the role of UX studies in identifying the optimal streaming resolution. This resolution must balance user satisfaction with effective bandwidth management. We are exploring how machine learning (ML) algorithms may develop into sophisticated adaptive streaming algorithms that can dynamically change resolution in response to user preferences and real-time bandwidth circumstances with the help of insights from UX research. This approach aims to balance maintaining high streaming quality and ensuring efficient content delivery, providing a smooth, immersive VR experience that responsibly manages network resources and addresses individual user needs.

Within user experience research, our work examines the intricate connection between virtual reality (VR) video quality and electroencephalography (EEG) data. This part of the study is critical because it enables us to understand how users receive and handle information in VR settings, providing insights into their mental involvement and comfort levels. While integrating EEG and VR opens avenues for various applications, from entertainment to medical advancements, our focus is on evaluating cognitive load in VR settings. This approach provides a deeper understanding of user interaction within virtual worlds and contributes to our scientific knowledge of brain-behavior in response to virtual stimuli.

Specifically, the study's comprehensive design attempts to uncover the relationship between EEG data and virtual experiences, offering an understanding of how various VR video resolutions impact users' brain activity. The present research explores certain facets of human cognition inside virtual worlds to determine how different degrees of digital inputs individuals engage in VR correlate with different brain activity patterns. This goes beyond a simple user experience assessment to propose a more nuanced perspective on cognitive engagement within VR settings. By enhancing our understanding of these mental states, our research contributes to the evolving field of human-computer interaction. Such advancements could have significant implications for how future generations interact with technology, potentially initiating a shift in how humans engage with digital realms.

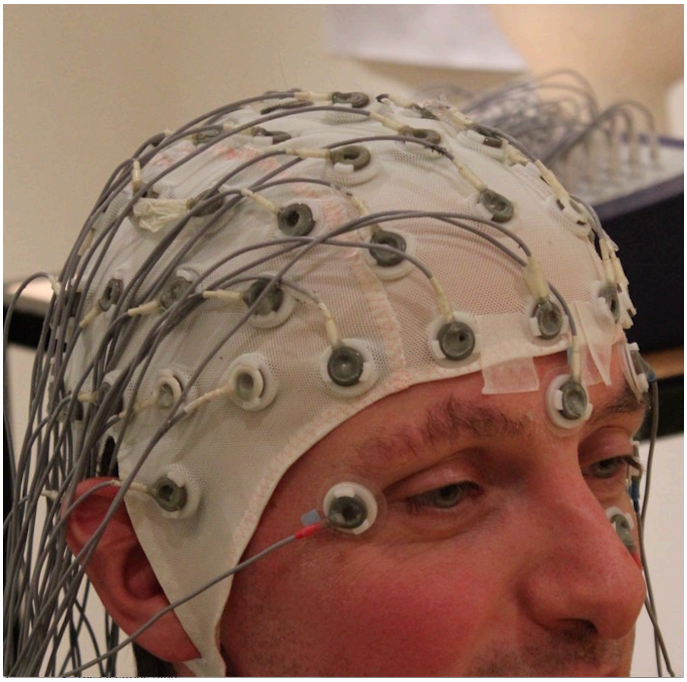


Figure 1.1: a cap for electroencephalograms. Several electrodes spaced uniformly over the skull are used to read the readings. The picture extracted from https://www.flickr.com/photos/tim_uk/8135755109/

1.2 Outline

VR, a key player in advancing modern technology and cognitive science, is significantly shaped by its historical development and contemporary technological innovations [1] The intricacy of cutting-edge VR systems, which can provide life-like interactions and navigations to replicate real-world experiences, distinguishes this progression [3]. VR, typified by its computer-generated simulations, immerses users in experiences mainly through visual and aural cues, improving our understanding of these complex, rapidly evolving fields [2].

Furthermore, the progression of VR technology is exemplified by state-of-the-art devices like the HTC Vive Pro Eye _1_ or Meta _2_ which augment realism and immersion through features like 6 degrees of freedom_3_ field of view and head tracking, crucial for a high-quality VR experience [5-7] . In line with these advancements, the concept of Quality of Experience (QoE) in UX of VR, as defined

1https://www.vive.com/sea/product/vive-pro-eye/overview/
2https://www.meta.com/be/en/quest/quest-3/
3https://www.motive.io/blog/dof-in-vr/

by the International Telecommunications Union, extends beyond mere functional metrics. It encompasses subjective user perceptions and interpretations, signifying overall satisfaction influenced by the user, system, and contextual factors [8-10].

Importantly, EEG is an instrument for the assessment of UX in VR. More specifically, EEG provides a window into the user's cognitive state by recording brain activity during VR interactions, capturing shifts in attention, engagement, and even discomfort [56]. These recordings serve as a basis for interpreting the specifics of the brain's interaction with VR surroundings, providing a detailed comprehension of the user's QoE.

ML and DL techniques are crucial to dealing with large amounts of EEG data and deciphering its complexities. With their ability to handle enormous amounts of information, these sophisticated algorithms reveal complex relationships and patterns that traditional analytic techniques could miss. Innovation and the development of VR systems are sparked by the use of ML and DL, which improve the accuracy of our comprehension of QoE in VR. Based on state-of-the-art brain analytics, this integrated approach is essential to advancing VR design. It guarantees applications are not just cognitively tuned but also personalized to the interests and requirements of individuals.

Notably, our study is a follow-up to the investigation of Zheleva et al. [29], which provided critical insights into the impact of video quality on QoE in VR. Utilizing the HTC Vive Pro Eye head-mounted display (HMD), participants experienced the animated feature "INVASION" in various video quality settings, highlighting the profound influence of video quality on perceptual and cognitive experiences in VR. The study utilized the Balanced Latin squares technique to avoid any order effects in the data. Building on these findings, the current study delves into the impact of video quality variations on the user experience, focusing on cognitive and sensory aspects.

The video quality levels explored in the study were:

- High-quality (Q1) - 2469 x 2743 pixels
- Medium quality 1 (Q2) - 1808 $\times$ 2009 pixels
- Medium quality 2 (Q3) - 1169 $\times$ 1298 pixels
- Low quality (Q4) - 512 $\times$ 549 pixels

They employed a dual approach to estimate QoE: objective measurement using EEG caps with 64 electrodes alongside subjective assessment through participant

questionnaires [29]. These questionnaires allowed participants to provide their perspectives on various aspects of the VR experience, such as perceived video quality, cognitive load, sensory immersion, and additional factors like simulator sickness and narrative immersion.

The central objective of this thesis is to conduct a thorough analysis of EEG data gathered during the investigation mentioned above by utilizing an amalgamation of progressive statistical, ML, and DL methodologies. The primary emphasis is on identifying EEG markers that consistently reflect users' cognitive load as they experience varying levels of video quality in VR. This study underscores the utilization of ML algorithms, particularly the Long Short-Term Memory (LSTM) and Multi-Layer Perceptron (MLP) models, due to their proficiency in detecting complex patterns embedded within the high-dimensional, often noisy EEG data. These advanced computational techniques are anticipated to uncover nuanced yet crucial correlations and patterns that might remain undetected through conventional analytical methods.

1.3 Limitations

This Master's thesis project navigates several challenges, primarily focusing on data collection constraints. The dataset is from 30 participants and examines four distinct resolution film types. Given this limited participant pool, common in such studies, only 28 are used for model training, while the remaining two serve as test data. This scenario complicates the classification task and underscores the difficulties inherent in research involving high-end EEG systems in VR, which is expensive and time-consuming. A key objective of this project is to evaluate the efficacy of machine learning (ML) methodologies in the context of small EEG sample sizes. Such an approach holds significant potential for small companies or research institutes seeking efficient quality testing methods without extensive data collection.

Regarding computational resources, the project is bound by the capabilities available at imec-mict-UGent. The constraints of about 16 GB of memory and a 250 GB disk capacity limit the complexity of the neural network models that can be developed and the number of parameters and trials that can be conducted. Although ML cloud infrastructure could serve as a viable alternative, the project adheres to these specific limitations due to budgetary and logistical reasons.

Lastly, the focus of this project is on utilizing sensor-level data. This data type

represents the direct, unprocessed information gathered from EEG readings and VR system outputs, offering an immediate and detailed view of the participants' physical and sensory experiences. This project emphasizes the straightforward examination of raw data. This method may not go into the more complex, interpretative parts of data analysis, but it is appropriate given the resources accessible and the project's defined objectives. Despite these drawbacks, we're committed to getting the most out of this unprocessed data and will use machine learning methods to draw valuable conclusions from the collection. The goal of utilizing the inherent value of unprocessed sensor data is to establish a standard for carrying out influential and significant research in this field.

2 Background

This chapter introduces the pertinent theory underlying the work done in this Master's thesis project and expands on the central investigated concepts.

2.1 Virtual Reality

Virtual reality (VR) is a multifaceted concept primarily understood as integrating virtual objects within a virtual environment [3]. A broader definition views VR as a computer-generated simulation of real-world scenarios to deliver immersive experiences primarily through auditory and visual stimuli [2].

It is an environment created by a computer designed to mimic being in three dimensions and feel like the user is there. Desai, Ajmera & Mehta (2014) [30] claim that this sense of real life inside a digitally created environment allows people to alter, engage with, and observe their environment actively. This functionality was further explained by Brooks (1999) [31], who highlighted the many gestures and actions that users may employ to engage with the VR world.

Additionally, advanced VR systems feature high-quality virtual environments (VEs), facilitating user interaction and navigation that closely mimic real-life experiences [3]. Ivan Sutherland [4][32], a pioneer in computer graphics and virtual reality, had a visionary concept that significantly influenced the development of VR. In this hypothetical space, objects like balls would possess realistic weight and texture, enabling actions such as throwing or bouncing. Even more complex items, like handcuffs, could be simulated with convincing realism. This vision underscores VR's ability to replicate the tactile feel of physical objects with im-

pressive accuracy [32].

To be more precise, virtual reality falls into two main categories: immersive and non-immersive [11]. Immersive VR, often experienced through a head-mounted display (HMD), is notable for its capacity to replace the user's external sensory inputs with those of a virtual environment [33]. This technology involves deep immersion, interactivity, and user involvement, fully absorbing the user into the virtual world and detaching them from the real world [33]. Also, HMDs enhance the immersive experience, allowing users to navigate VR environments in sync with their head and eye movements, effectively blocking the real world. This capability is not limited to exploration but extends to interacting with virtual objects and characters, as highlighted by Sanchez-Vives & Slater (2005) [34].

In contrast, non-immersive VR provides a virtual experience that does not wholly isolate users from their physical environment, allowing for an interactive yet less encompassing form of engagement [33]. Comprehending this distinction is critical to understanding the spectrum of virtual experiences available with current technology [33].

With the burgeoning popularity of VR, various innovative devices have emerged, each contributing uniquely to the VR experience. The HTC Vive Pro Eye_1_ is a notable example; it is a lightweight goggle device optimized for virtual environments. It offers users 6 degrees of freedom_2_ and advanced head-tracking features to intensify immersion [5-6]. This device utilizes a combination of a magnetometer, accelerometer, and gyroscope for precise head movement tracking, ensuring that the virtual imagery aligns seamlessly with the user's movements. In the same realm, Oculus _3_has significantly contributed to the VR industry, with its devices celebrated for their innovative design and user-friendly interface. Both Oculus and the HTC Vive Pro Eye (Figure2.1) exemplify the technological advancements in VR, providing highly immersive experiences and setting new standards in the field.

1https://www.vive.com/sea/product/vive-pro-eye/overview/

2https://www.motive.io/blog/dof-in-vr/

3https://www.counterpointresearch.com/insights/oculus-captures-half-xr-headset-market-2020/

Figure 2.1: HTC Vive Pro Eye and Oculus head-mounted display

2.2 Quality of Experience

Quality of experience (QoE) [35] is more than just a matter of performance metrics or functionality; it immerses itself into the subjective realm of the user's perception and interpretation of events. This interpretation views an event as observable and characterized by its qualitative, temporal, and geographical features [35]. Perception, a fundamental aspect of QoE, involves a complex and multi-layered process. It is not just about passively receiving stimuli but actively identifying and interpreting them. This dynamic nature of perception means that QoE extends beyond the initial sensory experience, including interpreting and analyzing the meaning experienced [35].

Furthermore, the International Telecommunications Union (ITU-T) [9] broadens this understanding by defining QoE as the overall level of customer satisfaction with a service or application subjectively experienced by the user. Complementing this view, Reiter and colleagues emphasize that QoE is shaped by a combination of factors related to the user, such as visual acuity, the system including audio or visual distortions, and the context, like the environmental setting [161].

Additionally, the 2020 Qualinet white paper [36] examines the convergence of VR and QoE. It suggests that the complex phenomenon of QoE acquires additional layers in the context of VR, attributed to its immersive nature. This paper focuses on how individual differences in perception among users can influence QoE in VR environments [36]. A vital aspect of this, especially in VR, is the 'sense of being there' or presence. Presence is a user-centric characteristic of QoE that allows users to perceive the virtual environment and its events as authentic, becoming a foundational element for immersive media experiences [36]. This discourse underlines that aspects related to the user, system, service, application, or context all

significantly shape the overall QoE in VR environments.

As we delve deeper into the nuances of QoE within VR, the critical role of the user's cognitive processes becomes increasingly apparent. The immersive nature of VR does not merely alter perception; it significantly impacts cognitive functioning and the experience of cognitive load (CL). This observation naturally leads us to cognitive load theory (CLT), which offers a framework to comprehend how immersive experiences in VR can influence cognitive processes. Bridging VR's immersive characteristics with the implications on cognitive load is crucial for a comprehensive understanding of the user experience in VR.

2.3 Cognitive load

In our study, we focus on how Quality of Experience (QoE) in Virtual Reality (VR), can be significantly influenced by Cognitive Load (CL). Cognitive load encompasses how much information our brain can handle at one time [65]. This is significant in virtual reality, as viewers face intricate, multi-faceted conditions.

For instance, the brain's working memory functions like a short-term storage space, capable of simultaneously handling approximately 7 bits of information [23-24]. In the context of Virtual Reality (VR), this capacity is frequently challenged to its maximum due to the intricate nature of virtual environments. For instance, when learning the grammar of a new language in VR, the demands on working memory are significantly higher than for more straightforward tasks such as vocabulary memorization [37]. This heightened demand is primarily attributed to the immersive nature of VR environments, which often necessitate the simultaneous processing of visual, auditory, and sometimes haptic stimuli. Such multisensory engagement intensifies the load on working memory, particularly when grappling with abstract concepts like grammatical rules.

On the other hand, long-term memory acts like a vast archive, helping to reduce the load on working memory. It does this through 'cognitive schemas' — mental frameworks that help process complex information as single, manageable units, easing cognitive strain [37-38]. Over time, as users become more familiar with VR tasks, these cognitive schemas become more ingrained, reducing the mental effort required for similar functions in the future [40].

Breaking down cognitive load further, it consists of intrinsic, extraneous, and germane load [65]. Intrinsic load is about the inherent complication of the material relative to the user's knowledge [16]. Apt load is the effort to make sense of this

complexity and integrate new information [41]. On the other hand, the extraneous load is any additional mental effort caused by the way information is presented, like poor video quality in VR, which doesn't contribute to learning [43].

It is relevant to our VR study because the quality of experience (QoE) in virtual reality (VR) encompasses more than visual clarity and immersion. It also concerns the effects of intrinsic, relevant, and extrinsic cognitive load on the user's capacity for experience processing and enjoyment. An excessive cognitive load may adversely affect a user's experience, which would reduce the effectiveness and enjoyment of VR interactions. Thus, our study's focus on recognizing and controlling cognitive load is essential to ensuring that VR experiences are enjoyable and supportive of learning or engagement. Using CL, we can create VR experiences that align with consumers' cognitive abilities, improving the overall quality of experience in the medium.

2.4 Electroencephalography (EEG)

Electroencephalography (EEG), a sophisticated electrophysiological technique that measures ionic currents along cell membranes, capturing excitatory and inhibitory synaptic activities. This functionality enables EEG to detect activities at varying depths within the brain, with extracellular potentials generated during this process amalgamating to form the EEG signals. These signals are distinguished by their rapid amplitude fluctuations, typically occurring within time constants of less than a second [69].

The recordings of spontaneous electrical activity, also known as spontaneous EEG activity, show distinct waveforms that predominate over an extensive frequency range [64]. The random EEG activity is generally classified according to the following frequency bands: delta (δ, < 4 Hz), theta (θ, 4-8 Hz), alpha (α, 8-13 Hz), beta (β, 13-30 Hz), and gamma (γ, > 30 Hz). A specific distribution over the scalp and a clear biological significance distinguish the rhythmic activity at these frequency bands. For example, alpha waves are usually measured from the occipital cortex region in an awake person with the eyes closed, and they diminish with eye opening or mental activity. Furthermore, it is common for EEG rhythms to exhibit a reduced amplitude as frequency increases [64].

It is crucial to analyze the frequency components of its continuous signal to glean more subtle insights from the EEG data. These components are reflective of the synaptic connections' activity between neurons. Research by Pfurtscheller and

Da Silva (1999) [70] has revealed that neuronal networks within this framework display pattern fluctuations, oscillating between synchronization and desynchronization. When these oscillations achieve specific frequencies, they give rise to 'alpha waves,' detectable through EEG methodologies [71].

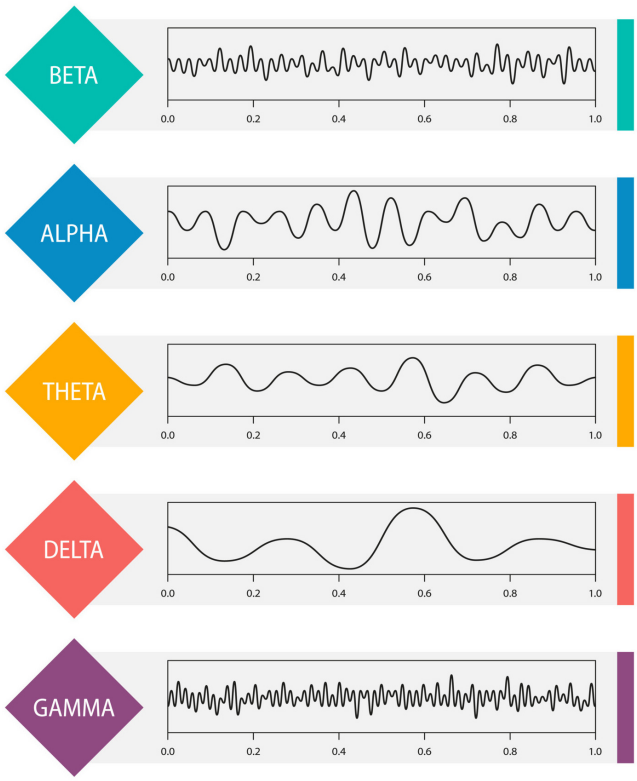


Figure 2.2: Rhythmic EEG activity patterns

2.4.1 Measuring Cognitive Load with EEG

EEG, a noninvasive and time-tested technique has become a cornerstone in measuring brain activity, particularly in real-world environments [45]. Esteemed for its ability to capture the brain's electrical dynamics through scalp-placed electrodes, EEG adeptly registers electrical signal variations corresponding to fluctuating cognitive stimulus levels. This characteristic, as highlighted in the works of Anderson and Bratman [64], Klimesch [47], and the study by Tromp et al., which investigates language processing in naturalistic environments using VR and EEG [186], makes EEG particularly well-suited for assessing cognitive load. The advent of advanced, wireless EEG systems, such as the ANT Neuro [187], has further increased ecological validity by minimizing equipment size and enabling simultaneous data collection from multiple subjects

However, the application of EEG in natural settings faces certain limitations. Standard EEG research necessitates data gathering in controlled experimental setups, requiring numerous trials to ensure accuracy. Furthermore, EEG's spatial resolution _4_, the scale at which the smallest unit of an image can identify distinct objects [66] is relatively low (in the centimetre range), posing challenges in precisely pinpointing activated brain regions. Additionally, EEG is susceptible to motion artefacts, such as blinking or movement, which can introduce noise into the data, sometimes overwhelming the actual neural activity signal [48]. External disturbances like electrical interference and physiological factors, including respiration and heartbeats, can also contaminate the EEG signal.

Despite these challenges, EEG remains crucial for continuous cognitive load monitoring. Modern EEG systems have sophisticated software that aids signal denoising, artefact elimination (e.g., eye blinks), and data analysis. Notably, EEG's high temporal resolution_5_ can detect shifts in cognitive activity at the millisecond level [66] and allows it to reflect a participant's mental state [48] consistently.

A robust theoretical and methodological framework is vital for interpreting EEG data accurately and reliably. Notable in this context is Basar's theory of neural oscillations [28], which sheds light on the functional significance of alpha

4Spatial resolution describes the smallest unit of an image that may be used to differentiate objects or the smallest linear or angular distance that can be used to identify neighbouring items in an image.

5Temporal resolution is the time required to return to and collect data for a particular place. In remote sensing, its duration is determined by the orbital parameters of the sensor platform and the sensor itself.

and theta oscillations as described by Klimesch [47]. Additionally, the event-related (de-)synchronization method developed by Pfurtscheller and Aranibar [50] significantly enhances the precision and efficacy of continuous EEG in measuring cognitive load.

Moreover, the continuous EEG signal, or spontaneous EEG, is characterized by oscillations at various frequencies. These oscillations, crucial in representing and transferring information within and across neuronal assemblies, have been emphasized in research like that of Klimesch et al. [26].

In addition, Basar, who proposed a comprehensive theory of neural oscillations, is a notable contributor to this renewed focus [28]. His extensive work underscores the functional significance of the brain's electrical activities, suggesting EEG's ability to detect brain wave rhythms and proposing that these rhythms serve as an "alphabet for brain functions" [27]. These rhythmic potential changes provide direct, measurable indices of various brain activities, encompassing functions like sensory registration, perception, movement, and cognitive processes such as attention, learning, and memory [28]. This perspective has shifted the perception of EEG from a tool primarily measuring responses to specific stimuli to one that also captures neuronal networks' broader and dynamic interactions.

To better understand the intricacies of obtaining and interpreting EEG data, the following section will expand on the technical and theoretical background of the EEG method.

2.5 Other methods

A brief overview of several techniques for measuring brain activity is provided below.

- MEG: Magnetoencephalography (MEG), Like EEG, measures the magnetic fields in the brain caused by electrical currents. There are usually hundreds of channels in MEG. Moreover, the magnetic measurements exhibit less distortion than the EEG's electric potential due to the skull. MEG must be protected against magnetic fields, such as the Earth's. It can thus only be applied in regulated settings, such as clinical settings.

- fMRI: Brain activity is calculated by practical magnetic resonance imaging, which looks for variations in blood flow. It takes advantage of the idea that blood flow is enhanced in engaged brain regions. A powerful magnetic field is used to align the oxygen nuclei, and a second field is used to identify the nuclei and detect blood flow. The approach has a high spatial resolution, identifying the place in three dimensions, but a low temporal resolution.

- Invasive methods: Measurements taken beneath the cranium are part of invasive procedures. One advantage is that the cranium does not distort the signals. However, it is a hazardous surgery that should be reserved for dire circumstances.

2.5.1 Montage

The international 10-20 System in EEG research is paramount, as it provides a standardized framework for placing scalp electrodes (i.e., a montage) on a global scale. This system is intricately detailed in Figure 2.3, serving as a cornerstone for uniformity in research and clinical applications. Its methodology involves strategically positioning electrodes in proximity to the cerebral cortex. The critical aspect of this system is the measured distances between adjacent electrodes, which are set at either 10% or 20% of the skull's total front-back or right-left distance. This approach is thoroughly documented in the work of Herwig et al., 2003 [132], who provided an in-depth analysis of the system's underpinnings.

Further refinement of this system is evident with the introduction of the Modified Combinatorial Nomenclature (MCN). The MCN enhances the precision of electrode placement through the use of specific numbers (1, 3, 5, 7, 9 for the left hemisphere; 2, 4, 6, 8, 10 for the right hemisphere) to indicate percentages (10%, 20%, 30%, 40%, 50%) of the inion-to-nasion distance. This is clearly illustrated in Figure Figure2.4. The adoption of the MCN allows for the integration of additional

electrodes at 10% divisions, thereby enriching the EEG recording by incorporating intermediate sites between the standard electrodes of the 10-20 system. In the MCN, each electrode placement is denoted with a letter that corresponds to different cerebral regions: Temporal (T), Parietal (P), Occipital (O), Frontal (F), and Central (C).

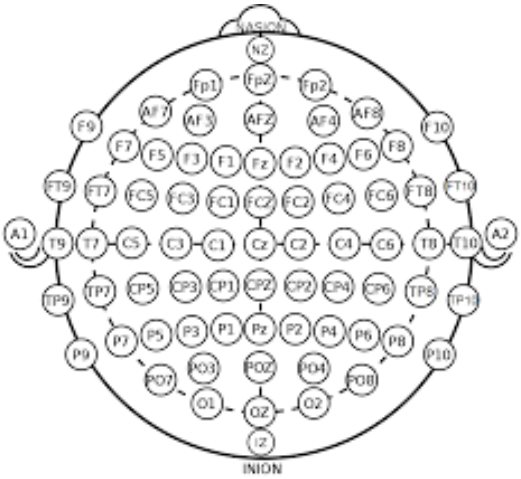


Figure 2.3: The Modified Combinatorial Nomenclature. When recording a more detailed EEG using more electrodes, extra electrodes are interpolated using the 10% division, which fills intermediate sites halfway between those of the existing 10–20 system.

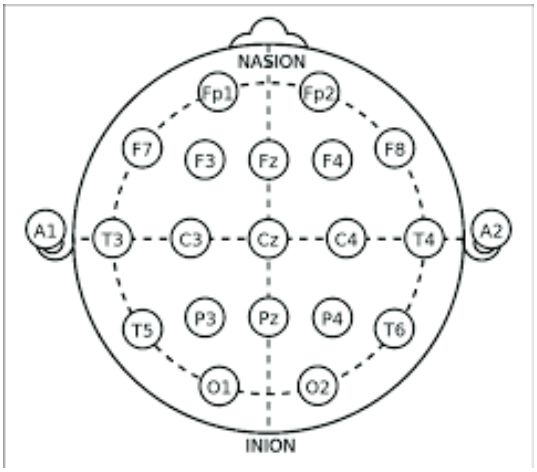


Figure 2.4: The international 10–20 EEG Placement System.

The significance of these standardized electrode placements must be balanced in the context of EEG data analysis. EEG studies form channels by measuring the differences in electrical potentials between these strategically positioned

electrodes. This methodology is essential in translating the rigorous standards of electrode placement, as set forth by the 10-20 and MCN systems, into actionable neurological insights. The work of Hu et al., [133] provides a comprehensive discussion on this topic, delving into the types of montages utilized in EEG. These include bipolar and referential montages, distinguished by their specific combinations of channels based on electrode potentials. This nuanced approach underscores electrode placement's complexity and critical nature in pursuing accurate and insightful EEG data interpretation.

2.6 EEG features

In analyzing EEG data to assess various cognitive workload states, the features associated with cognitive workload can be categorized into several domains: time domain, frequency domain, time-frequency domain, and nonlinear dynamics features. These categories encompass a range of EEG characteristics, each offering unique insights into brain operation and cognitive load [73].

An EEG experiment allows for a broad and multifaceted understanding of brain activity. Additionally, the classification of these features can be based on cognitive or perceptual responses. Such a classification helps distinguish between brain activity types and their corresponding cognitive or perceptual implications. This multi-domain feature extraction and classification approach is crucial in accurately interpreting EEG data, particularly about cognitive workload and its impact on brain function.

Time Domain

Obtaining the time domain analysis is simple and intuitive. Its goal is to find the change in signal amplitude or other attributes concerning time. Event-related potentials (ERP) [60], statistics features (e.g., mean, standard deviation, variance, kurtosis, skewness), higher-order crossing analysis [74], and the Hjorth parameter are the most common time domain features. Since the time domain analysis was initially created and included most EEG data, many researchers continue to use it. ERP properties include time-averaged EEG peaks that are time-locked to distinct stimuli [75].

Frequency Domain

The frequency domain analysis is presented to display the frequency information of EEG signals under the premise that EEG signals are stationary. First, a signal is translated from the time domain into the frequency domain. From there, the frequency band can be broken down into several sub-bands, including delta (0.1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (30-80 Hz), which are all strongly associated with neuronal activity in humans. In short, deep sleep, drowsiness, relaxedness, engagement, awareness, and active states are mostly linked to delta, theta, alpha, beta, and gamma bands, respectively [78]. Depending on how the researcher defines each frequency domain, its range might vary slightly. These decomposition strategies rely on the Fourier transform [79]. Numerous research studies have demonstrated correlations between different effort levels and power changes in EEG frequency bands. The power fluctuations of the alpha band (decreased) and theta band (raised), in particular, may act as discriminant markers for workload estimates with increasing workload [76][80]. Aside from the alpha and theta bands, powers in the delta, beta, and gamma bands have been linked to varied workloads [81-83]. Zarjam et al. [80] investigated the working memory workload discriminating capabilities of distinct frontal frequency bands and discovered that lower delta power was substantially correlated with increased workload. Zarjam et al. [80] also discovered a correlation between rising effort and rising theta band power across prefrontal regions and falling alpha band power in parietal areas [84].

Furthermore, frequency domain properties such as power spectrum, energy, event-related synchronization/desynchronization [85], and power spectral density (PSD) may be estimated using these frequency sub-bands. PSD, or power distribution as a function of frequency, is the most commonly used feature [60].

Time-Frequency Domain

Since EEG signals are non-stationary and can contain time and frequency information, time-frequency domain research investigates how the spectrum changes with time. Typically, it employs sliding windows and assumes that the signals within the window are stationary before computing the features using frequency domain algorithms. Some related techniques are the fixed-window short-time Fourier transform [86], wavelet transform (which includes both discrete and continuous wavelet transform [83]), and empirical mode decomposition (a nonlinear

technique) that adaptively adjusts the window size based on the wavelet function.

The short-time Fourier transform separates signals into small sequential data frames (by shifting windows) and then performs a quick Fourier transform to each frame [.87]. For example, a short-time Fourier transform is applied to 128 Hz down-sampled data (5 minutes) with 40 s windows and 35 s overlap, and power features of seven frequency bands are calculated, yielding 133 features (of 19 channels) as classifier inputs [86]. The lengthy data sample is utilized with the longer windows in this case. In most situations, the window length of the Fourier transform is changed from 0.5 to 10 seconds for cognitive activities, depending on the duration of the data segmented. The discrete wavelet breaks down the signal transform into precise information and a rough estimate. It utilizes scaling and wavelet functions connected to low-pass and high-pass filters, respectively [77]. It produces the approximation and detail coefficients as features. The choice of acceptable wavelets and the number of scales is critical.

In essence, the time-domain signals are translated into the frequency domain using both the frequency and time-frequency domain, and the feature types (such as PSD) determined by the two are identical. The time-frequency domain is anticipated to extract more valuable information than the frequency domain since it contains some time-related information.

Nonlinear Dynamics

EEG signals have been considered nonlinear and nonstationary for a considerable time [88]. Non-linearity indicates that brain oscillation is not a linear combination of frequency band components. Nonstationary means that the frequency band components fluctuate in amplitude or form as time passes [89]. Consequently, nonlinear dynamic features are utilized to describe electroencephalography's irregular and nonlinear properties. Complexity and entropy measures are the most widely utilized nonlinear characteristics for workload analysis [90]. The complexity analysis primarily indicates the degree of unpredictability in time series, such as Lempel-Ziv Complexity [91] and its variations. Entropy metrics, including Shannon, approximation, and permutation entropy [53], represent signal variations and unpredictability. For example, Shannon entropy calculates the probability density based on the probability distribution of amplitude values. Approximate entropy evaluates the predictability of future amplitude values of a signal based on prior amplitude values. Interested readers might see [88-90] for further details.

2.7 Machine Learning

Machine learning, integral to artificial intelligence and data science, empowers computer systems to develop new abilities, autonomously evolving through experience and data. This field is defined by its capability to improve performance metrics in various tasks through training [92]. A broad range of techniques exists within machine learning, each suited to specific applications. These include supervised learning, unsupervised learning, semi-supervised learning, active learning, transfer learning, multi-task learning, and reinforcement learning. Among these, supervised learning is particularly prominent [92].

In supervised learning [92], models are trained to predict outputs from input data where most or all data labels are known. This approach includes classification and regression tasks, employing a k-nearest neighbour, support vector machine, and linear discriminant analysis.

Traditional machine learning models use fixed-length feature vectors, such as Random Forest, AdaBoost, SVM, XGBoost, and ANN. These models can predict unseen examples by identifying patterns in the training data, which is typically labelled.

Advancements in cognitive load identification mark a significant leap in machine learning applications. Researchers now utilize physiological data, including EEG, fNIRS, and ECG, to build models that predict cognitive load levels. Machine learning and deep learning algorithms interpret these signals and categorize cognitive load levels. Despite the progress, challenges such as high individual variability, noise interference, and limited model generalization highlight the need for more robust and accurate models [93].

2.8 Deep Learning

Deep learning, a technical subset of machine learning, employs artificial neural networks (ANNs) modelled after the biological brain's structure and functionality. These ANNs are adept at handling and learning from various unknown inputs [106]. Mirroring biological neurons, ANNs consist of interconnected neurons that process input values. Their non-linear nature enables them to discern complex relationships and patterns between inputs and outputs, making them particularly effective for predicting cryptocurrency prices based on time-series data [107].

Parameters and hyperparameters are critical to the functionality of ANNs,

which significantly influence the learning process and overall model behaviour [108]. These include:

Number of Hidden Layers

The number of hidden layers in an ANN impacts its speed and generalization capabilities. A lower count leads to faster operation and broader applicability, aligning with the preference for simpler models that accurately classify data. Increasing the number of layers generally improves accuracy, but additional layers yield only a few improvements beyond a certain point, rendering a heavily layered ANN inefficient.

Activation Functions

These are essential for calculating neuron outputs based on inputs and connection weights [109]. They play a pivotal role in the network's ability to process data non-linearly.

Learning Rate

This determines how much the model's weights are adjusted during training. Lower learning rates extend the training duration but ensure steady progress, whereas more significant rates might hasten the process at the cost of potential instability [110].

Number of Epochs

This represents the total number of times the neural network will train on the entire dataset. More epochs typically mean higher accuracy, but beyond a certain point, additional epochs yield negligible improvements in accuracy, making further training time inefficient [111].

Batch Size

This parameter defines the number of inputs in each dataset sample processed by the neural network [111]. It influences the efficiency and effectiveness of the training process.

Optimizers

These algorithms update the neural network's parameters, like weights and learning rate, to minimize loss - the discrepancy between predicted and actual outputs during training. Optimizers aim to fine-tune the network for the most accurate results [112].

2.9 Time Series Classification

Artificial Neural Networks (ANNs) [101] are renowned for their ability to discern patterns and elements within data. This capability is primarily attributed to their proficiency in handling high-dimensional data and their differentiable structure, although these are not their only distinguishing characteristics. Certain ANN types are found to be particularly effective in the realm of time series classification.

Key discriminative models in this domain include Multi-Layer Perceptrons (MLP), One-dimensional Convolutional Neural Networks (CNN), Fully Convolutional Neural Networks (FCNN), Encoder models, Multi-Scale CNN, Time Le-Net, Multichannel Deep CNN, Residual Networks, Long Short-Term Memory (LSTM) networks, Gated Recurrent Neural Networks (RNN), GRUs (Gated Recurrent Units), S4 models, and Transformers. These models are referenced across various sources, including [101-105]. Among these, MLP and LSTM are particularly notable for their specific features and capabilities, which make them well-suited to the challenges of processing and classifying time-sequenced data, a focal point of my thesis.

2.9.1 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks, a type of recurrent neural network (RNN) _6_ architecture, are highly esteemed in deep learning, particularly for their prowess in sequence prediction tasks [113]. These networks are distinct from traditional neural networks due to their feedback connections, enabling them to process entire sequences of data rather than isolated data points. This capability is crucial for accurately understanding and predicting patterns in sequential data like time series. An LSTM unit operates in three key stages [115] :

- Determining whether to retain or discard information from the previous timestamp
- Learning new information from the current input
- Transmitting updated information to the subsequent timestamp

This cycle is considered a single-time step in LSTM processing.

Each LSTM unit comprises three gates: (1) the Forget gate, (2) the Input gate, and (3) the Output gate - which regulates the flow of information within the memory cell or LSTM cell [116]. These gates function similarly to layers of neurons in a traditional feedforward neural network, with each neuron featuring a hidden layer and a current state.

The traditional ILSTM architecture uses sigmoid and tanh functions internally within the LSTM cells for gating and state regulation. The output layer in the model is a linear layer without an explicit activation function because CrossEntropyLoss is used, which expects raw scores and applies the necessary transformations internally. Systematic experimentation was conducted, including variations in the number of sub-layers within the LSTM layers, the sizes of the hidden states, and the dropout probabilities. These experiments aimed to identify improvements in the ILSTM architecture, potentially enhancing the model.

6A recurrent neural network (RNN) is one of the two broad types of the artificial neural network [114-113], characterized by the direction of the flow of information between its layers.

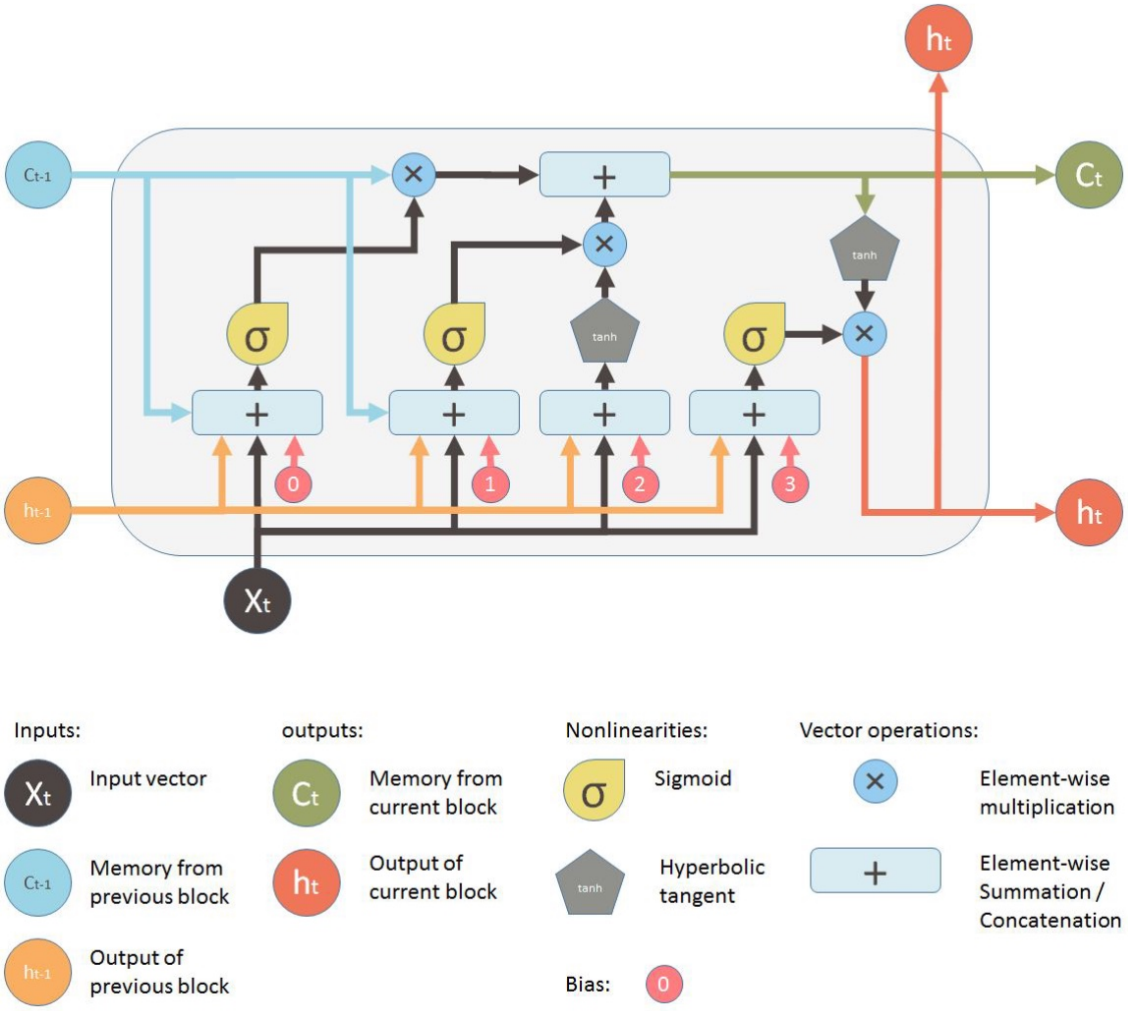


Figure 2.5: LSTM method

The fundamental equation for an LSTM is as follows:

$$\vec{c}_0 \leftarrow \vec{0} \tag{2.1}$$

$$\vec{h}_0 \leftarrow \vec{0} \tag{2.2}$$

$$\vec{f}_t \leftarrow \sigma_s(\vec{W}_f \vec{x}_t + \vec{U}_f \vec{h}_{t-1} + \vec{b}_f) \tag{2.3}$$

$$\vec{i}_t \leftarrow \sigma_s(\vec{W}_i\vec{x}_t + \vec{U}_i\vec{h}_{t-1} + \vec{b}_i) \tag{2.4}$$

$$\vec{o}_t \leftarrow \sigma_s(\vec{W}_o\vec{x}_t + \vec{U}_o\vec{h}_{t-1} + \vec{b}_o) \tag{2.5}$$

$$\vec{h}_t \leftarrow \vec{o}_t \odot \sigma_h(\vec{c}_t) \tag{2.6}$$

$$\vec{c}_t \leftarrow \vec{f}_t \odot \vec{c}_{t-1} + \vec{i}_t \odot \vec{c}_t \tag{2.7}$$

$$\vec{\tilde{c}}_t \leftarrow \sigma_t(\vec{W}_c\vec{x}_t + \vec{U}_c\vec{h}_{t-1} + \vec{b}_c) \tag{2.8}$$

- $\vec{x}_t$:= The input vector
- $\vec{f}_t$:= Forget gate activity $\in (0,1)^h$
- $\vec{i}_t$:= Input gate activity $\in (0,1)^h$
- $\vec{o}_t$:= Output gate activity $\in (0,1)^h$
- $\vec{h}_t$:= Hidden state (output vector) $\in (-1,1)^h$
- $\vec{\tilde{c}}_t$:= Cell input activation vector $\in (-1,1)^h$
- $\vec{c}_t$:= Cell state vector $\in \mathbb{R}^h$
- $\vec{W}$:= Input Weights $\in \mathbb{R}^{h \times d}$
- $\vec{U}$:= Recurrent Weights $\in \mathbb{R}^{h \times h}$
- $\vec{b}$:= Bias $\in \mathbb{R}^h$
- σ_s := Sigmoid function
- σ_t := Hyperbolic tangent function

2.9.2 Multilayer Perceptron (MLP)

The Multi-Layer Perceptron (MLP), A predominant type of supervised neural network, it is intricately designed with three fundamental layers: (1) input, (2) hidden, and (3) output. Each layer comprises nonlinear computational elements, often called neurons or processing units, allowing for the seamless flow of information from the input to the output layer via the hidden layer. This intricate architecture of MLPs has established them as the neural network of choice for a broad range of approximation tasks in static and dynamic contexts.

In static scenarios, MLPs are traditionally trained using a gradient-descent-based supervised procedure; a method made highly efficient by the specific topology of the MLP. This process, known as the backpropagation algorithm, has been a cornerstone since its inception [117]. In dynamic contexts, MLPs have demonstrated their prowess in complex tasks such as identifying and controlling nonlinear systems and time series prediction. These applications have been successfully explored through feedback architectures in supervised neural models [118-121], notably in addressing the Time Series Prediction (TSP) problem, where the objective is to forecast future values of a series beyond a known index.

In the MLP (Multilayer Perceptron) architecture, the model distinctly employs nonlinear activation functions and regularization techniques, differentiating it from ILSTM architectures. Specifically, the MLP utilizes Rectified Linear Unit (ReLU) activations in intermediate layers, introducing essential non-linearity that enables the network to capture and model complex data patterns. This is crucial, as without such nonlinear activations (see Appendix 7.1), an MLP would only perform linear transformations, limiting its ability to solve complex, nonlinear problems.

Additionally, implementing dropout in a model is a regularization strategy to prevent overfitting, ensuring the model generalizes well to unseen data. The final output layer is a linear transformation, a common characteristic in MLPs, particularly when paired with loss functions like CrossEntropyLoss that process raw scores. This configuration, along with systematic experimentation on layer sizes and dropout rates, underlines the typical structure of an MLP and its capability to adapt and learn from diverse data sets.

An MLP can be defined mathematically as follows:

$$\vec{h}_0 \leftarrow \vec{x} \tag{2.9}$$

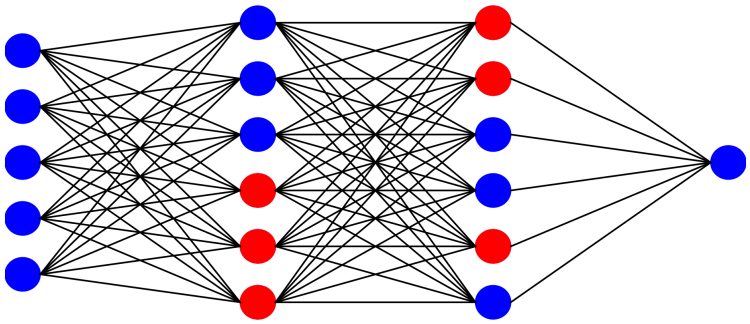


Figure 2.6: MLP method with 0.5 dropout rate

$$\vec{h}_n \leftarrow g(\vec{W}\vec{h}_{n-1} + \vec{\theta}_n) \tag{2.10}$$

$$\vec{y} \leftarrow \vec{h_N} \tag{2.11}$$

- $\vec{x}$:= The input vector that is fed to the network.
- n := n $\in [0, N]$
- N := The number of layers in the network (N $\in \mathbb{Z}^+$)
- $\vec{h_n}$:= The vector of neuron state values in network layer n
- g := Activation function)
- $\vec{y}$:= The output vector of the network
- $\vec{\theta_n}$:= The threshold vector of the network layer n
- $\vec{W_n}$:= The weight matrix of network layer n

2.10 Performance Evaluation

In evaluating the implementation of machine learning (ML) algorithms, we adopted a comprehensive approach using a set of critical metrics. These include Accuracy, Error Rate, Recall (also known as Sensitivity), Specificity, Precision, and the F1-Score. Our methodology is aligned with the practices outlined in previous studies [142-143].

The evaluation process began with calculating Sensitivity (or Recall), Specificity, Precision, and Accuracy. Sensitivity [143], also referred to as Recall, is the metric that determines the proportion of actual positives correctly identified. This is crucial for understanding how effectively the model detects positive cases and is computed as per formula 2.12. On the other hand, specificity focuses on the proportion of true negatives correctly identified [143], providing insight into the model's ability to recognize negative instances, as per formula 2.13.

Precision is another critical metric, representing the proportion of true positives among all optimistic predictions [143]. This measure, calculated using formula 2.14, helps assess the Accuracy of tAccuracydels' optimistic predictions. Lastly, Accuracy, calculatedAccuracyformula 2.15, gives an overall picture of the model's performance by measuring the proportion of total accurate predictions (positives and negatives).

This multifaceted approach ensures a thorough and balanced assessment of the ML models, adhering to established standards in the field [142-143].

$$\text{Recall} = \frac{\text{True Positives}}{\text{False Negatives} + \text{True Positives}} \tag{2.12}$$

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{False Positives} + \text{True Negatives}} \tag{2.13}$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{2.14}$$

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{True Positives} + \text{False Positives} + \text{True Negatives} + \text{False Negatives}} \tag{2.15}$$

The F1-Score emerges as a crucial metric in machine learning because it effectively combines Precision and Recall into a singular, comprehensive measure [144]. This integration is vital because neither Accuracy nor Recall alone can provide a complete picture of an algorithm's performance. For instance, a model may exhibit high Precision but low Recall, indicating accurate but incomplete optimistic predictions. Conversely, a model with high Recall but low Precision correctly identifies most positive cases but with numerous false positives.

The F1-Score addresses this imbalance by harmonizing these two metrics into one unified score [145]. It offers a more balanced view of a model's performance, especially when the trade-off between Precision and Recall is significant. The computation of the F1-Score, as outlined in formula 2.16, is particularly valuable for binary or multiclass classification tasks. It involves a calculation considering Accuracy and Recall, providing a single, more holistic measure of an algorithm's efficacy.

By integrating Precision and Recall, the F1-Score thus serves as a reliable indicator of a model's overall accuracy and completeness in predictions, making it an indispensable tool in evaluating machine learning algorithms.

$$F1\text{-}Score = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2.16}$$

In multiclass classification problems, metrics like Accuracy, Recall, Precision, and the F1-Score must be calculated individually and then averaged for each class. This approach ensures a comprehensive assessment of the model's performance across all classes, adapting these metrics from their original binary classification context to suit the complexity of multiclass scenarios.

Accuracy

Accuracy over a dataset is the percentage of correct outputs.

2.10.1 Loss Function

Artificial neural networks (ANNs) serve as function approximators, learning through identifying inaccuracies in input classification. Central to this process are loss functions, which quantify the deviation of ANNs' predictions from actual values [101]. Selecting an appropriate loss function depends on the specific problem, such as linear regression, binary classification, or multi-class classification. Distinctively, regression problems focus on predicting continuous quantities, whereas classification problems target discrete class labels. In the case of this thesis, the problem is a multi-class classification.

Categorical Cross-Entropy Loss

This is used for multi-class classification problems where the outputs are probabilities of each class, and the actual output is one of these classes. It is a generalization of binary cross-entropy to multiple classes.

$$H(p, q) = -\sum_{x} p(x) \log q(x) \tag{2.17}$$

where p is the true probability for x, and q is the model output.

3

Methodology

This chapter presents a full description of the dataset used in this work, demonstrating the procedures used in its collection, preparation, and structure before its use in this thesis (Figure3.1). This preliminary step is crucial because it provides us with clean data that is best suited for our next model assessment.

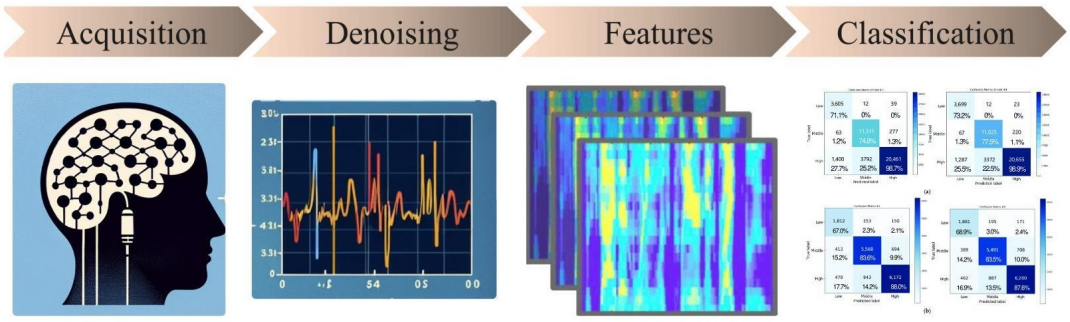


Figure 3.1: Stages of EEG Data Processing for Brain Activity Analysis

3.1 Participants

In our study, we initially recruited 34 participants with an average age of 24.24 years (SD = 3.93), predominantly comprising females (24 out of 34). Notably, most of them, accounting for 51.51%, had prior experience with Virtual Reality, as detailed in Figure 3.2. Our recruitment strategy primarily involved a social media

campaign. The selection process was guided by stringent criteria established by the study's EEG methodology, which included the exclusion of individuals with a history of head trauma or concussions, those with thick, curly hair or dreadlocks, and left-handed persons. During the research, we encountered challenges with EEG signal quality, primarily due to the mobile nature of recording and noise. Consequently, this led to excluding four participants whose EEG signals were significantly compromised.

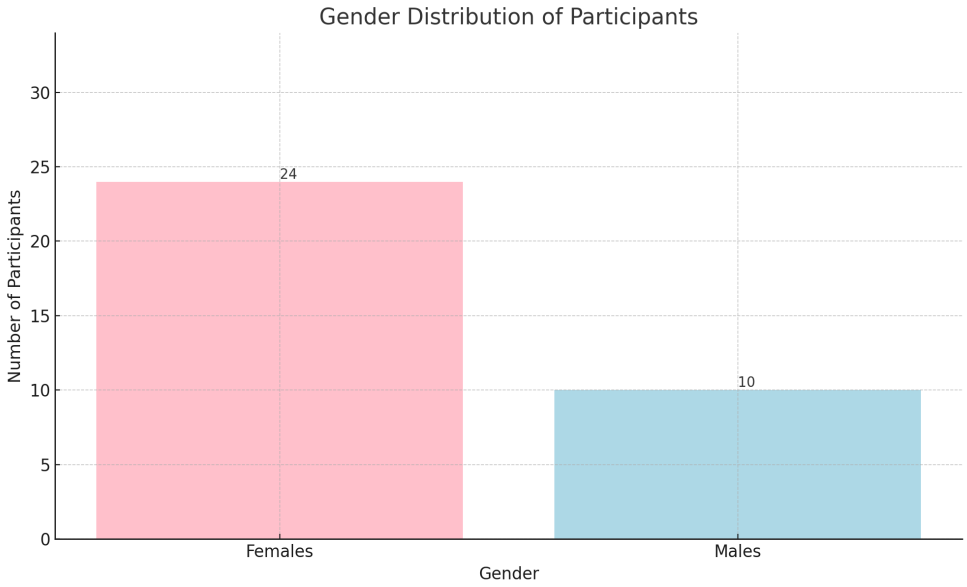


Figure 3.2: Demographic and Experience Profile of 34 Participants

3.2 Data Acquisition

Our study utilized a within-subject design where participants experienced the VR movie INVASION!, developed by Baobab Studios Inc. and narrated by Ethan Hawke _1_. The viewing was through the HTC Vive Pro Eye head-mounted display (HMD), offering a resolution of 2880 × 1600 pixels and a 90Hz refresh rate. The movie starts with a view of Earth from space, transitioning to a scene on a frozen lake, featuring interactions with animated characters like a white bunny and aliens.

Participants viewed INVASION! in four resolution levels: high-quality (2469 ×

1https://www.baobabstudios.com/invasion

2743 pixels), two medium qualities (1808 × 2009 pixels and 1169 × 1298 pixels), and low quality (512 × 549 pixels), based on the highest resolution playable without jitter, then linearly degraded.

Due to EEG requirements, screening criteria excluded individuals with specific physical conditions. After an online pre-test questionnaire, participants attended a 1.5-hour lab session. Here, they were fitted with a 64-electrode EEG cap, ensuring all electrodes had an impedance below 20Ω. The EEG data, including vertical eye movement (VEOG), was recorded at 1024 Hz, using the ANT Neuro cap _2 _, referenced to CPz. The ANT Neuro mobile EEG set and eego™ software facilitated recording without online filtering. This methodology aligns with the original QoMEX paper [29] standards, crucial for detailed brain activity analysis in VR environments. Post-experience, participants completed a questionnaire for each video version, followed by a debrief and compensation.

Figure 3.3: Minimalist Depiction of EEG Data Collection in a Lab

2https://campaign.ant-neuro.com/unleash-the-power-of-high-density-eeg?source={google}&gad_source=1&gclid=CjwKCAiA1-6sBhAoEiwArqlGPgGXDegcqYjqvtEgiDCCwKwh6Y_FHEdSt7dSZXgZJaGbGcw-m1YeohoCxLcQAvD_BwE

3.2.1 VR Equipment

For a visually stimulating participant experience, we employed the HTC Vive Pro Eye head-mounted display (HMD)_3_. With a refresh rate of 90Hz and a resolution of 2880 × 1600 pixels, this HMD produced an immersive viewing experience. Participants did not utilize the HMD controls since they were not required to engage with the VR experience.

3.2.2 VR Content

Participants viewed "INVASION!", a five-minute animated film produced by Baobab Studios Inc. with narration by Ethan Hawke _4_. "INVASION!" opens with a compelling picture of Earth seen from orbit, engrossing viewers in the story right away. The narrative then shifts to an interactive scene on a frozen lake where viewers meet and converse with a white rabbit. This interactive feature is enhanced with a transforming element that makes viewers envision themselves as bunnies when they gaze at their bodies, drawing them even more into the experience. With the arrival of two extraterrestrial entities and the magnificent spacecraft's landing, the story of "INVASION!" abruptly shifts. The introduction of a second rabbit character at this critical time adds complexity and excitement to the tale, increasing participant involvement. Our study presented four different visual quality iterations of "INVASION!" to each participant:

- Q1 (high quality, 2469 × 2743 pixels),
- Q2 (medium quality 1, 1808 × 2009 pixels),
- Q3 (medium quality 2, 1169 × 1298 pixels),
- Q4 (poor quality, 512 × 549 pixels).

To ensure seamless viewing, we first identified the highest resolution that could be played without video jitter and then created lower quality levels by reducing the resolution linearly. We employed balanced Latin squares to control for potential biases due to the viewing order. This methodical approach guaranteed that each participant's experience was consistent and unbiased across the various iterations of the movie.

3https://www.vive.com/sea/product/vive-pro-eye/overview/
4https://www.baobabstudios.com/invasion

3.2.3 Measurement Instrument

The study commenced with participants completing an online pre-test questionnaire, which assessed factors such as age, gender, previous Virtual Reality (VR) experience, and immersive tendencies [96]. This initial screening was crucial for gathering baseline data on the participants' backgrounds.

Following their VR experience, participants were asked to complete a comprehensive post-test questionnaire. This questionnaire consisted of subjective opinions on the video quality [97], assessments on narrative and sensory immersion [98], a presence questionnaire to evaluate their sense of being 'in' the VR environment [99], and the Karolinska Sleepiness Scale to measure their alertness [100]. These tools were essential for understanding the participants' experiences and reactions to the VR content.

For the EEG recording, the study employed a 64-electrode EEG ANT Neuro cap, recording EEG signals at 256 Hz relative to the CPz reference point. In offline analyses, an electrode was positioned below the left eye to scan vertical eye movements (VEOG), later called FP1. The EEG signals were captured using the eegoTM software on a Microsoft Surface Go tablet. Notably, no internet filtering was applied during this process, ensuring the purity and integrity of the data collected.

3.3 Data Pre-processing

Ensuring the integrity of our research, the preprocessing of EEG data stands as a cornerstone, addressing the susceptibility of EEG recordings to a spectrum of artefacts. These artifacts are broadly classified into physiological—stemming from intrinsic body activities like heartbeats, muscle contractions, and eye movements—and non-physiological, arising from external factors such as electrical interference and issues with electrode placement. Through a rigorous preprocessing regimen, the EEG data is subjected to band-pass filtering within the 1-50 Hz range to remove extraneous line noise and slow drifts. Concurrently, Independent Component Analysis (ICA) is employed to isolate and exclude eye blinks and muscle artifacts [122-125][130][155][177], while additional measures, such as handling and discarding of malfunctioning EEG channels and the strategic placement of electrodes using a standard layout, are taken to uphold data quality. Despite these efforts, challenges persist, notably in the form of muscular artefacts, particularly

from the frontalis and temporalis muscles, which may lead to the inadvertent loss of crucial brain information while attempting to purify the signal [127].

To combat non-physiological artefacts that significantly alter the data [124][128], the preprocessing protocol incorporates the use of the MNE software package for in-depth analysis [129]. This includes channel interpolation, data segmentation in alignment with specific events, and the computation of power spectral density for each segment with Welch's method, collectively enhancing the robustness and dependability of the findings [131].

In parallel, the machine learning facet of the study is pivotal, deploying tailored algorithms to navigate through and categorize the prepared feature datasets. Each participant's data is precisely segmented to coincide with individual VR video presentations, a critical step to guarantee the precise interpretation and classification by the learning algorithms, thereby deriving meaningful insights and conclusions.

This methodological rigor extends to the dataset "other_participants _dataJASP.csv," where specific participants (p6 and p26) were selectively excluded to verify the robustness of the final model, and participant p14 was omitted upon being flagged as an outlier. In a bid to render the data machine-learning-ready, the 'Video_quality' column was transcribed into a numerical format using the LabelEncoder, and the StandardScaler was applied for feature normalization. This holistic approach to EEG artefact management and meticulous data preparation is indispensable for the preservation of the EEG data's accuracy and the reliability of subsequent neuroscience research findings.

3.4 Data Analysis

Effective computational models in the field of EEG data processing are largely dependent on both sensible feature selection and the requirement for feature reduction. The characteristics that are described in this section were carefully chosen because they were recognized in academic literature for their resilience and importance in representing intricate brain patterns, as well as their inherent capacity to capture important EEG signal qualities. The choice was supported by early testing to validate their unique effectiveness and improve classification model accuracy, as well as a comprehensive review of prior research. Parallel to this, a large number of possible features—particularly in intricate EEG analysis—make feature reduction necessary. Although practical constraints force a reduction approach to eliminate superfluous parts, optimal classifiers must still be able to identify the importance

of each feature. This improves classifier accuracy and operational efficiency, which are essential for the effectiveness of real-time applications.

3.4.1 Features Extraction

- Hjorth parameters is created by Bo Hjorth In 1970 [152]. Their composition consists of three variables that are both time- and frequency-domain calculable.

- Activity: The variance of the signal var(y(t)), which is equivalent to the power.
- Mobility: Defines the average frequency and may be computed as:

$$\text{mobility}(y(t)) = \sqrt{\frac{\text{var}\left(\frac{d}{dt}y(t)\right)}{\text{var}(y(t))}} \tag{3.1}$$

- Complexity: Represents the variation in frequency and is computed as:

$$\text{complexity}(y(t)) = \frac{\text{mobility}\left(\frac{d}{dt}y(t)\right)}{\text{mobility}(y(t))} \tag{3.2}$$

-Mean is a statistical measure that represents the central point of a data set by averaging its values.

-Standard deviation, on the other hand, is the square root of the variance and provides a measure of dispersion in the same units as the data, which makes it more interpretable. It is instrumental in understanding the variability of the dataset and is widely used to gauge the degree of variation or dispersion of a set of values.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{3.3}$$

-Variance is a fundamental statistical measure that quantifies the dispersion within data points. It represents the expected value of the squared deviation of each data point from the mean, serving as a precise indicator of the data's spread. In simpler terms, Variance measures the degree to which each number differs from the mean and, thus, from every other number in the set.

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 \tag{3.4}$$

-Kurtosisis a statistical measure that describes the tails of a probability distribution about the normal distribution [151]. It assesses the extremity of tail

outliers, with high Kurtosis indicating a distribution with significant tail data suggesting outliers (heavy-tailed) and low Kurtosis indicating a distribution with fewer tail data suggesting fewer outliers (light-tailed). In essence, Kurtosis is an indicator of the presence or absence of outliers in a dataset. A distribution with a kurtosis similar to that of a normal distribution is referred to as mesokurtic; one with higher Kurtosis is leptokurtic, and one with lower Kurtosis is platykurtic. The uniform distribution, with its lack of tails, represents an extreme case of low Kurtosis.

$$\text{Kurtosis} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^4}{(N-1)S^4} \tag{3.5}$$

-Skewness is a statistical metric that quantifies the asymmetry of the likelihood distribution of a real-valued random variable around its mean [151]. It is a descriptor of the distribution's shape, with the skewness value being positive for a distribution that tails off to the right, negative for one that tails to the left, zero for a perfectly symmetrical distribution, and potentially undefined for distributions without a well-defined mean or variance. This parameter is pivotal in statistical analysis, as it provides insights into the nature of the probability distribution beyond the standard measures of central tendency and variability.

$$\text{Skewness} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^3}{(N-1)S^3} \tag{3.6}$$

-Energy of a time series signal is the sum of the squares of its amplitude values over time [153]. This concept, drawn from the field of signal processing, encapsulates the signal's power over a finite duration and is a crucial parameter in analyzing the signal's strength and potential impact.

$$E = \sum_{i=1}^{N} \text{signal}[i]^2 \tag{3.7}$$

N is the total number of values in the signal.

-Power Spectral Density (PSD) outlines a signal's power spread across frequencies[153-154]. Welch's method refines the periodogram approach to estimate spectral density by segmenting the signal into overlapping parts and applying windows to each, thereby computing modified periodograms [155]. This process reduces variance significantly, yielding a reliable and less noise-sensitive PSD estimation.

This refined method enhances the accuracy and consistency of PSD estimates,

ensuring a stable representation of the signal's energy distribution across the frequency spectrum [155], and is preferred in advanced spectral analysis.

$$PSD(f) = \frac{1}{N \cdot L} \sum_{k=1}^{K} \left| X_k(f) \right|^2 \tag{3.8}$$

After computing the PSD across all frequencies, the function limits the PSD and frequency array to include only frequencies up to 100 Hz. This is done using a boolean mask (freqs <= 100), which selects only the elements of freqs and PSD where the condition is proper.

-Shannon entropy, often referred to as Shannon's entropy, is a measure from information theory that quantifies the hesitation involved in predicting the value of a random variable [51]. Claude E. Shannon introduced this concept in his seminal 1948 paper "A Mathematical Theory of Communication."

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i) \tag{3.9}$$

Parameter	Units	Description
Mean	μV	The average value of the signal.
Standard deviation	μV	Quantifying signal variation and dispersion.
Variance	μV^2	Square of the standard deviation.
Kurtosis	Unitless	Assessing distribution tails in statistics.
Skewness	Unitless	Measuring mean asymmetry in distributions.
Hjorth parameters	Unitless	Statistical analysis of time series.
Shannon entropy	Bits	Measure of the randomness in the signal.

Table 3.1: Time Domain Feature

Parameter	Units	Description
Energy	μV^2	Total quantity of energy in the signal.
PSD	$\mu V^2/Hz$	Assessing frequency-based power in signals.

Table 3.2: Frequency Domain Feature

3.5 Feature Reduction

In our EEG signal analysis, a key challenge was the abundance of potential features, necessitating effective feature reduction strategies to enhance classifier accuracy and operational efficiency, crucial for real-time applications. We addressed this through two primary strategies: feature selection and feature extraction, both tailored specifically to our data set.

For feature selection, we adopted both Forward and Backward Selection methods. Initially, our approach involved Forward Selection, where we started with no features and sequentially added them, evaluating each addition based on improvements in the classifier's performance. This method was instrumental in identifying the most impactful features from our EEG data. Subsequently, we employed Backward Selection, beginning with the entire set of features and methodically eliminating the least beneficial ones. This two-pronged approach in feature selection was pivotal in retaining a subset of original features that were most relevant for our analysis, aligning with standard practices as described by Huang (2015) [54].

In the realm of feature extraction, we focused on transforming the original feature space into a dimensionally reduced form. This was essential for managing the high computational complexity we faced with our large set of EEG features. By distilling the essence of the original data, we could significantly streamline the data analysis process. Our approach here was mainly supervised, utilizing known labels of our training samples to guide the transformation.

Throughout this process, we were acutely aware of the trade-offs between accuracy and computational feasibility. Hence, our focus often shifted to heuristic methods, balancing speed and precision. This was especially true in scenarios where the sheer volume of features could potentially impede operational speed. Our methods, while sometimes embracing the inherent randomness of non-deterministic approaches, were consistently aimed at achieving near-optimal solutions in the context of EEG signal classification.

3.5.1 Correlation-Based Feature Selection

In data analysis and machine learning, the strategic selection of features is paramount for optimizing model performance and enhancing interpretability. Among the various methodologies available, the Correlation-Based Feature Selection (CFS),

a brainchild of Hall in 1998 [52], stands out as a groundbreaking approach. Central to this technique is the conviction that an exemplary subset of features should exhibit robust correlations with the target class and maintain minimal inter-correlations among themselves. This guiding principle is succinctly articulated by Hall: "Good feature subsets contain features highly correlated with the class, yet uncorrelated with each other." [52].

The rationale behind CFS is deeply rooted in the notion that each feature within a predictive model should contribute distinct and complementary insights regarding the class. This approach recognizes that when features exhibit high inter-correlations, they are more likely to be redundant, offering limited new understanding and potentially escalating the risk of overfitting. Conversely, features indicating solid correlations with the class bolster the model's predictive accuracy.

To estimate the efficacy of a feature subset, CFS employs a specialized metric, denoted as M_S, which is calculated considering two pivotal factors: the mean class-feature correlation ($\vec{r_{cf}}$) and the average feature-feature correlation ($\vec{r_{ff}}$). The formula is defined as follows:

$$M_s = \frac{\overline{kr_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}} \tag{3.10}$$

k signifies the count of features within the subset. The numerator of this equation encapsulates the collective strength of correlation between the class and the features, endorsing subsets where each feature significantly aids in class prediction. The denominator, grounded in the average feature-feature correlation, is a deterrent against feature redundancy. Accordingly, a higher value of MS indicates a more favorable feature subset, achieving an optimal balance between high relevance to the class and low inter-feature redundancy.

Thus, Correlation-Based Feature Selection epitomizes a strategic, discerning approach to feature selection, prioritizing subsets of features that are individually predictive and collectively diverse. This method is crucial in refining model accuracy while curtailing the likelihood of overfitting. It underscores the intricate balance necessary in the feature selection process, emphasizing the significance of synergistic interaction and distinctiveness among features in constructing robust, predictive models.

3.5.2 Correlation-Based Feature Extraction

In the field of EEG-based machine learning analysis, one of the most formidable challenges encountered is the 'curse of dimensionality.' This term encapsulates the complexities that arise when dealing with high-dimensional EEG data, which is often compounded by the limitation of having only a small sample size available for analysis, as highlighted by Mwangi et al. [140]. The critical task here is the extraction of a concise set of discriminative features from the extensive EEG dataset, a step that is imperative for improving both the classification performance and the interpretability of the resulting models, as noted by Tu et al. [141].

In our research, we have addressed this challenge by implementing the JASP software [57]. This innovative approach entailed segmenting the EEG dataset according to electrode placements, thereby facilitating a detailed and precise analysis of the features. Our methodology involved the strategic application of ANOVA, utilizing 'Video_quality' as a consistent fixed factor across various experimental conditions, namely baseline, q1, q2, q3, and q4.

This meticulous process of feature extraction, which involved carefully selecting both dependent and fixed parameters and applying ANOVA, allowed us to effectively distil significant features from the high-dimensional EEG dataset. By doing so, we were able to confront and mitigate the challenges posed by the curse of dimensionality. Our approach was not merely about selecting the most relevant features; it was about transforming the raw, high-dimensional data into a more manageable and informative subset, aligning with our overarching goal of enhancing the performance of our predictive models.

The successful implementation of this technique illustrates a significant stride in the realm of EEG data analysis. Through strategic feature extraction, we have demonstrated that it is possible to overcome the inherent difficulties of high-dimensional data spaces, thereby paving the way for the development of more accurate, efficient, and interpretable machine learning models in the field of neuroinformatics and beyond.

Cases	Sum of Squares	df	Mean Square	F	p	η^2
Video_quality	$6.112 \times 10^{+14}$	4	$1.528 \times 10^{+14}$	6.780	$< .001$	0.153
Residuals	$3.381 \times 10^{+15}$	150	$2.254 \times 10^{+13}$			

Table 3.3: ANOVA - C5_energy

Video_quality	Mean	SD	SE	Coefficient of variation
baseline	$2.655 \times 10^{+6}$	$3.196 \times 10^{+6}$	573938.306	1.204
q1	$8.716 \times 10^{+6}$	$6.598 \times 10^{+6}$	$1.185 \times 10^{+6}$	0.757
q2	$6.835 \times 10^{+6}$	$5.059 \times 10^{+6}$	908702.326	0.740
q3	$5.379 \times 10^{+6}$	$4.077 \times 10^{+6}$	732283.367	0.758
q4	$5.521 \times 10^{+6}$	$4.088 \times 10^{+6}$	734233.241	0.740

Table 3.4: Descriptives - C5_energy

One example among the various features analyzed for each electrode 'C5_energy' was used as the dependent variable, representing. The findings, demonstrated by a 'p value' of less than 0.001 and a significant effect size ($\eta^2 = 0.153$), revealed that 'C5_energy' is considerably influenced by the levels of 'Video_quality.' This result exemplifies the effectiveness of our feature selection process, emphasizing the discriminative features that contribute to the robustness and interpretability of the machine learning models we aim to build.

3.5.3 Selection of Significant Features

After an in-depth review of Correlation-Based Feature Extraction, we have compiled a comprehensive list of critical features for our study. These features have been meticulously selected based on their statistical significance in relation to the dependent variables, which is outlined in the table below.

The guiding principle for our selection was the p-value; a feature was deemed significant if it was associated with a p-value below the threshold of 0.05, a standard recognized in statistical analyses to differentiate between random fluctuations and true associations.

However, our investigative process revealed a notable exception in this dataset: the P8 and C4 electrodes. Despite rigorous analysis, these electrodes did not present any features that met our significance criterion. This lack of significance could stem from various factors, including the specific physiological functions or regional brain activities that these electrodes monitor, or it might reflect the unique characteristics of our dataset and the phenomena under investigation.

This finding prompts a discussion about the potential limitations of the P8 and C4 electrodes in our current research framework. It raises questions about whether these electrodes capture data relevant to our dependent variables or if alternative analytical methods might unveil subtle, yet significant, features. It's

also worth considering this outcome when interpreting our results, as it highlights the complex and nuanced nature of EEG data and its relationship with cognitive and neural processes.

In the table summarizing our discriminative feature selection, we include a note on the P8 and C4 electrodes, specifying the absence of significant features. This transparency is crucial, as it provides a complete picture of our feature selection process, reflecting both the strengths and the limitations of our dataset. The table continues to detail the features that have shown a substantial correlation with our outcome variables across various frequencies and dimensions, including measures like power spectral density (PSD) at multiple frequencies, standard deviation (std), variance, and other statistical metrics.

Electrode	Features
VEOG	psd : 1HZ , 3-59HZ, 62HZ,63HZ; std; skewness; shannon entropy
Fpz	mobility; complexity
Fp2	psd: 2 Hz
F7	energy
F3	energy; psd: 2-5HZ, 13HZ, 14HZ; std
Fz	psd : 2-4Hz, 12-16HZ, 27-29HZ, 32-49HZ, 51-62HZ; std
F4	energy; psd: 3-6HZ
F8	energy; psd: 2-9HZ
FC5	energy; psd: 2-8 HZ
FC1	energy; 12-62HZ; std; variance
FC2	activity; energy; psd:1-6HZ, 12-15HZ, 29-62HZ; std; variance
FC6	energy; psd: 3-8 HZ, 14HZ
M1	psd: 3-9Hz, 14HZ
T7	activity; energy; psd: 1-5HZ, 13-16HZ, 38-52HZ; std; variance
C3	activity; energy; psd: 1HZ; std; variance
Cz	activity; energy; psd: 1-4 HZ, 12-22HZ, 27-62HZ; std; variance
T8	activity; energy; psd: 1-31HZ, 33HZ, 40-41HZ, 47-49HZ, 51HZ; std, variance
M2	psd: 3-4Hz; std
CP5	std
CP1	energy; psd: 2-6HZ; std
CP2	psd: 24-46Hz, 48-55Hz, 57-60Hz, 62Hz; std
CP6	psd: 4Hz
P7	psd: 2-8Hz, 11-17Hz; std
P3	psd: 13-62Hz; std
Pz	psd: 2-5Hz, 14-17Hz, 20-62Hz; std
P4	psd: 12-62Hz; std
POz	psd: 8-62Hz; std
PO8	psd: 4-26Hz, 42-51Hz, 56Hz, 59-63Hz; std
O1	activity; psd: 1-19Hz; std; variance
O2	activity; psd: 4-25Hz, 43-52Hz, 55-56Hz, 58-63Hz; std; variance

Table 3.5: Selected Features

Electrode	Features
AF7	energy; psd: 2-7Hz; activity; energy
AF3	psd: 1-9Hz, 12-15Hz; std; variance
AF4	energy; psd: 1-8Hz; mean; std
AF8	energy; psd: 2-9Hz
F5	energy; psd: 2-7Hz
F1	psd: 2-5Hz, 13-16Hz, 44-45Hz; std
F2	energy; psd: 3-5Hz, 12-14Hz
F6	energy; 1-7Hz
FC3	psd: 13-14Hz
FCz	activity; energy; psd:3-4Hz, 12-18Hz, 26-62Hz; std; variance
FC4	energy; psd: 2-6Hz;
C5	activity; energy; psd: 2-7Hz; std; variance
C1	activity; energy; psd: 1Hz, 5-6Hz, 12-16Hz, 21-27Hz, 29-62Hz; std; variance
C2	activity; energy; psd: 1-6Hz, 13-14Hz, 24-62Hz; std; variance
C6	activity; energy; psd: 2-8Hz, 13-14Hz; mean; std; variance
CP3	psd: 25-26Hz, 29Hz, 35-37Hz, 39-40Hz, 42-43Hz, 45-46Hz, 48-49Hz
CP4	psd: 3-7Hz; std
P5	psd: 8Hz, 12-62Hz, std
P1	psd: 2-8Hz, 13-62Hz, std
P2	activity; energy; psd: 3-9Hz, 12-62Hz; std; variance
P6	activity; psd: 1-9Hz, std; variance
PO5	activity; psd: 1-9Hz, 11-25Hz, 58Hz, 62Hz; std; variance
PO6	activity; psd: 3-27Hz, 38-63Hz; std; variance
PO7	activity; psd: 1-18Hz; std; variance
PO3	activity; psd: 1-9Hz, 12-63Hz, std, variance
PO4	psd:14-63Hz
FT7	activity; energy; psd: 3-7Hz, 9Hz, 13-14Hz; std; variance
FT8	activity; energy; psd: 2-15Hz; std; variance
TP7	activity; energy; psd: 1-9Hz, 14-63Hz; std; variance
TP8	activity; psd: 2-9Hz, 13-15Hz; std; variance
Oz	activity; psd: 3-24Hz; std; variance
CPz	activity; complexity; energy; psd: 1-5Hz, 30-60Hz; std; variance

3.6 Artificial Neural Networks

Two model types have been the primary focus of this thesis project: MLP and LSTM. It should be noted that the LSTM and MLP models can only process univariate data, which limits the number of sensors they can process at once in this situation. Since the univariate models are trained using all of the available sensor data, they must handle all of the sensors equally. This has the advantage of providing many more training data points.

3.6.1 LSTM

The use of a Long Short-Term Memory (LSTM) model, a particular kind of recurrent neural network (RNN) that is very useful for sequential data processing, is the basis of this study. Our video quality evaluation system is based on this LSTM model, which has adjustable parameters such as input size, hidden size, output size, number of layers, and dropout rate. An LSTM layer is part of the model's architecture, and it is in charge of processing input sequentially and preserving data across time using hidden states. Because of this characteristic, it's especially appropriate for situations where context knowledge is essential. A completely linked layer that processes the output further comes after the LSTM layer. In this project, we will use an LSTM model to detect video quality using a dataset including a range of characteristics, representing significant progress in the industry, since the introduction of machine learning, particularly deep learning models such as LSTM, has revolutionized various sectors.

```
1 class LSTM(nn.Module):
2     def __init__(self, input_size, hidden_size, output_size,
    num_layers=1, dropout_rate=0.5):
3         self.init__ = super(LSTM, self).__init__()
4         self.hidden_size = hidden_size
5         self.num_layers = num_layers
6         self.lstm = nn.LSTM(input_size, hidden_size, num_layers,
    batch_first=True, dropout=(dropout_rate if num_layers > 1 else 0))
7         self.fc = nn.Linear(hidden_size, output_size)
```

Code 3.1: LSTM Model

3.6.2 MLP

In this project, we meticulously implemented a Multi-Layer Perceptron (MLP) for the classification of video quality, showcasing the model's adeptness in managing complex data patterns. The architecture of the MLP is thoughtfully constructed, comprising several key components: fully connected (linear) layers, batch normalization layers, ReLU activation functions, and dropout layers, each playing a pivotal role in the model's performance.

The fully connected layers, often referred to as linear layers, form the foundational backbone of the MLP. Their primary function is to execute linear transformations on the input data, effectively mapping the incoming data to a desired output space. This transformation is fundamental to the learning capabilities of the network.

To enhance the stability and speed of the training process, batch normalization layers are integrated into the architecture. These layers normalize the inputs of each layer, ensuring consistent distribution of the data as it passes through the network. This normalization is crucial as it mitigates the problem of internal covariate shift, thereby accelerating the convergence of the model during training.

The introduction of ReLU (Rectified Linear Unit) activation functions is a strategic decision to infuse non-linearity into the MLP. Non-linearity is a critical aspect of deep learning models, enabling them to learn and represent complex patterns and relationships in the data that linear models cannot capture. The ReLU function, in particular, is favored for its simplicity and effectiveness in introducing this non-linearity.

Furthermore, to combat the risk of overfitting, dropout layers are judiciously employed in the network. These layers randomly deactivate a subset of units in the network during training. This random deactivation prevents the model from becoming overly reliant on any specific feature, thereby ensuring a more generalized and robust learning.

The holistic approach in the implementation of this MLP, from rigorous data preprocessing to thorough model evaluation, underscores the meticulousness and strategic planning inherent in the development of a robust machine learning model. Each component of the MLP is carefully selected and integrated, reflecting the depth of understanding required to effectively navigate the complexities of video quality classification. This project not only demonstrates the technical proficiency involved in model construction but also highlights the importance of each step in

the journey from data to decision-making.

```
class MLP(nn.Module):
    def __init__(self, input_size, output_size, dropout_rate=0.5):
        super(MLP, self).__init__()
        self.fc1 = nn.Linear(input_size, 128)
        self.bn1 = nn.BatchNorm1d(128)
        self.relu1 = nn.ReLU()
        self.dropout1 = nn.Dropout(dropout_rate)
        self.fc2 = nn.Linear(128, 32)
        self.bn2 = nn.BatchNorm1d(32)
        self.relu2 = nn.ReLU()
        self.dropout2 = nn.Dropout(dropout_rate)
        self.fc3 = nn.Linear(32, output_size)
```

Code 3.2: MLP Model

3.7 Overfitting

In the training of Artificial Neural Network (ANN) classifiers, the goal is typically to construct a model that efficiently categorizes inputs into distinct classes by learning the most salient features relevant to the task. A common challenge encountered is overfitting, where the model, rather than acquiring a general understanding of the data, learns patterns specific to the training set that do not generalize well to unseen data. This issue can manifest when the model gives undue importance to less significant features that do not robustly represent the underlying problem.

To prevent overfitting, it is crucial to provide a rich and diverse training dataset. Such a dataset should encapsulate a wide array of features representative of the classes in question. This diversity ensures that the model is exposed to the various aspects of the data, prompting it to learn more substantial and generalizable features that are truly indicative of each class. Consequently, this practice enhances the model's predictive performance on new, unencountered data, reinforcing its practical applicability.

3.7.1 Prevention

- Dropout : Dropout[185] is a straightforward yet powerful technique for decreasing overfitting. It requires the model to acquire generic characteristics by excluding some neurons from the computation, i.e., disregarding their output. This prevents so-called specialized neurons from allowing the ANN to memorize the training data rather than acquiring the necessary general characteristics. The neurons to be dropped are chosen at random, with a fixed probability represented by p_d, which is specified as a hyperparameter (Figure3.4).

- L-Regularization: This regularisation technique, which is often referred to as weight decay, involves adding a multiple of the model weights' norm to the loss function as a penalty. For weight decay, the following two norms are frequently used:

  –
  $$L_1 \text{ norm} : \lambda \sum_{i=1}^{k} |w_i| \tag{3.11}$$

  –
  $$L_2 \text{ norm} : \lambda \sum_{i=1}^{k} w_i^2 \tag{3.12}$$

  Weight decay has been shown to increase weight sparsity, decrease overfitting, and increase the model's generalizability [184].

- Batch normalization: By normalizing the input data to the ANN's layers, it lessens the change in internal covariance. Furthermore, batch normalization has been shown to accelerate training by allowing for faster learning rates in addition to reducing overfitting[183]. Because trainable parameters determine the normalization effect, the model may be trained to deactivate the effect.

- Hyperparameter Optimization with Optuna: Optuna is used to determine the ideal hyperparameters for the model [139], including thelearning rate and dropout rate. By avoiding poor hyperparameter selections that might result in overfitting, Optuna reduces the chance of overfitting, as explained in detail in the section 3.7.2.

- Data Standardization: The use of StandardScaler [182] from sklearn. preprocessing for feature normalization is a crucial step that ensures all input features contribute equally to the model training, avoiding bias towards features with higher magnitude.

Table 3.6: Overfitting Prevention Techniques Utilized in MLP and LSTM Models

Model	Dropout	L-Regularization	BN	Optuna	Data Standardization
MLP	✓	✓	✓	✓	✓
LSTM	✓	✓		✓	✓

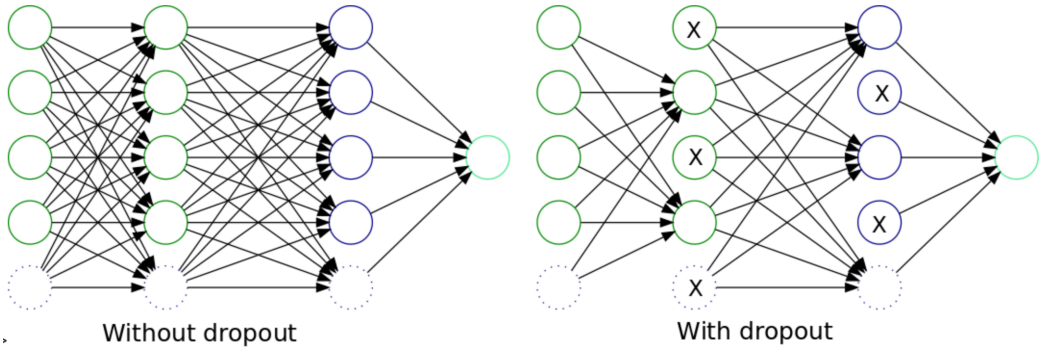


Figure 3.4: a fully linked multilayer neural network, as well as a network with and without dropouts. the picture extracted from:
https://www.oreilly.com/library/view/machine-learning-for/9781786469878/252b7560-e262-49c4-9c8f-5b78d2eec420.xhtml

3.7.2 Hyperparameter Optimization Techniques

Hyperparameter tuning is a critical process in the development of machine learning models, aiming to identify the ideal parameters that enhance precision and accuracy [138]. Recognized as a complex aspect of model building, it involves an iterative search for the optimal parameters, often requiring multiple attempts to find the 'sweet spot' that bolsters a model's performance. Traditional methods like grid and random search are commonly used for this purpose. Grid search systematically determines the optimal hyperparameters, while random search explores the parameter space randomly until certain criteria are met.

To aid in this process, Optuna, an open-source optimization framework developed by Preferred Networks, a Japanese AI company, has been introduced [139]. It is designed to automate hyperparameter optimization, efficiently navigating through the trial-and-error phase. Optuna stands out for its framework-agnostic capability, compatible with various Python-based frameworks such as Keras, Scikit-learn, and Pytorch. Its use is not limited to machine learning; it can be applied in any sector with a definable objective function.

Optuna utilizes an imperative, define-by-run API, which allows for dynamic alteration of the search space for hyperparameters, thus enhancing flexibility [139]. It employs Leave-One-Out Cross-Validation (LOOCV), ensuring each participant's data is used as a test set once, which maximizes the model's exposure to different data scenarios and enhances its robustness. The optimization process primarily focuses on tuning learning rate and dropout rate, significantly impacting the training

dynamics and the model's generalization ability.

The hyperparameter tuning process with Optuna is conducted outside the model, before entering the training phase, resulting in a set of optimized hyperparameters ready for integration into the training stage. This is crucial due to the involvement of numerous parameters, lengthy training times, and the use of multiple folds to minimize information leakage.

By integrating traditional optimization techniques with Optuna's advanced capabilities, developers can achieve a balance between inquiry and exploitation in their search for effective hyperparameters. This synergy not only saves time and resources but also significantly propels the performance of machine learning models toward their maximum potential.

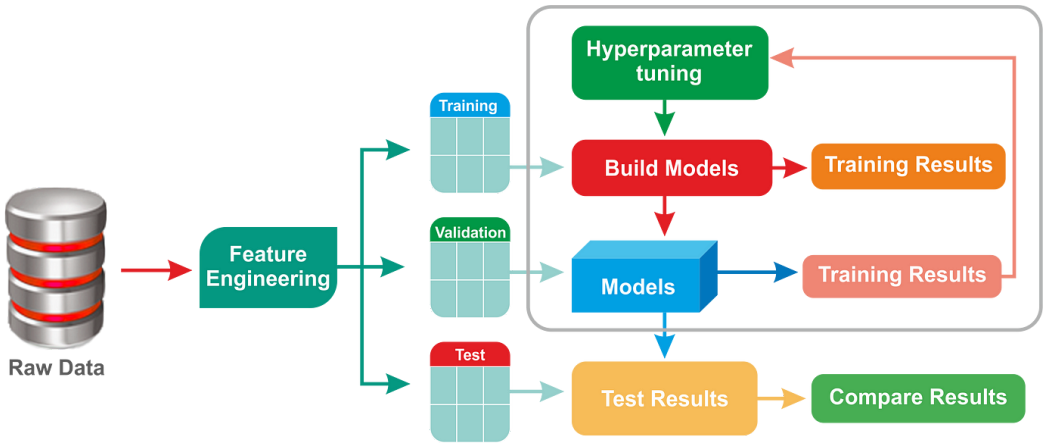


Figure 3.5: Photo credit: Optuna hyperparameter tuning source:https://medium.datadriveninvestor.com/hyperparameter-tuning-with-keras-tuner-3a609d3fd85b

3.8 Classification Fundamentals

A classifier for machine learning is a function that takes the values of several aspects of brain activity as independent variables or predictors from a condition and predicts the class to which the condition belongs. There are a few fundamental ideas in machine learning that must be covered:

- Class: the grouping to which anything is assigned. A class is a collection of patterns that have similar characteristics and typically come from the same source.
- Pattern: a list of an object's characteristics combined with the object's class information.
- Sample: An object's specific pattern is called a sample.
- Feature: a group of variables that includes information about an object's characteristics and ability to discriminate.
- Feature vector: A set of K such characteristics for a single sample, arranged in some manner into a K-dimensional vector, is called a feature vector.
- Feature space: the K-dimensional space in which the feature vector is located.

In an EEG experiment, characteristics might be derived from time, frequency, or spatial domains, and the class could be cognitive or perceptual responses. A sample (for example, a trial or subject) is represented by the vector X $\in$ R $^{N\times K}$, and its class label is represented by the vector y $\in$ R $^{N\times 1}$. A classifier may learn numerous parameters and estimate the weights of each feature using the training data to establish correlations between features and class labels. As a result, the classifier is officially a function f that predicts label y = f(X) given a sample X. The function f can be either classification, where the output is a discrete integer corresponding to a restricted number of classes, or regression, where the result is a continuous variable. The machine learning classifier has to be trained to understand the relationship (f) between features and the appropriate class labels. Following training, the classifier may be applied to test data to assess whether the features include discriminative information between sample classes. If the trained classifier accurately recorded the connections between features and labels, it could predict new classes of samples in the test data. Assuming that the training and test data are separately taken from a "sample distribution," we refer to the training and test data as $X_{training} \in R^{N\times K}$ and $X_{test} \in R^{L\times K}$, respectively. In this case, the class labels are represented by column vectors $y_{training}$ and y_{test}, features are the columns, while the samples are the rows. Prediction error, often

known as accuracy, measures how well a classifier performs in classification. It is the percentage of test data samples that are appropriately categorized.

3.8.1 Data Split

In machine learning, a classifier's primary objective is to estimate the correct class for a given feature vector, utilizing knowledge acquired during its training phase. This involves learning the mapping relationship between feature vectors and their corresponding class labels. The connection between these vectors and labels forms a decision border in the feature space, distinguishing patterns of different classes.

Before training, it is essential to organize the data into different sets:

- a training set (data with known class labels used for training the classifier)
- a testing set (data with known class labels used to evaluate the classifier's performance)
- sometimes, a field data set (data with unknown class labels for classification by the trained classifier)

The size of these samples plays a crucial role. Generally, a larger sample size improves the test power for the significance of accuracy and helps the classifier accurately estimate its training parameters.

Thanks to advancements in cross-validation techniques, all data can now be utilized for training and testing. An example is the "leave-one-out" cross-validation (LOOCV), as described by Tu et al. [137]. In LOOCV with N samples, the data is divided into N-1 training samples and one test sample. This process is repeated N times, ensuring each sample is used as a test sample once.

Moreover, other cross-validation methods are widely used in fields like EEG investigations. For instance, K-fold cross-validation, where K represents the number of splits (like 5 or 10), is notable. It's crucial to include samples from all classes in the training data of each fold to prevent the classifier from being unable to predict missing classes.

A successful classifier is not just about training performance (its ability to correctly identify the classes of training data) but also about generalization performance - the ability to classify unseen test data accurately. An ideal machine learning classifier should have a decision boundary that optimizes generalization performance, not just perfect training performance. Over-fitting, a condition where the classifier learns noise or random errors instead of the underlying relationship,

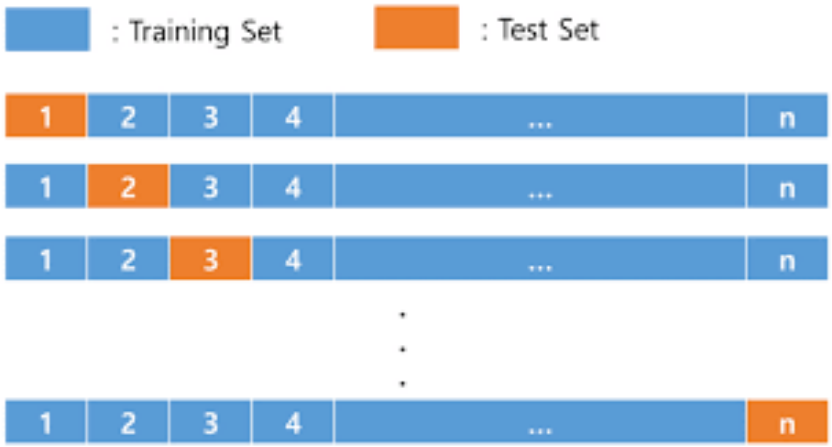


Figure 3.6: Leave One Out cross-validation

often arises from striving for perfect separation in training data. This can lead to a classifier that performs well on training data but poorly on unseen data.

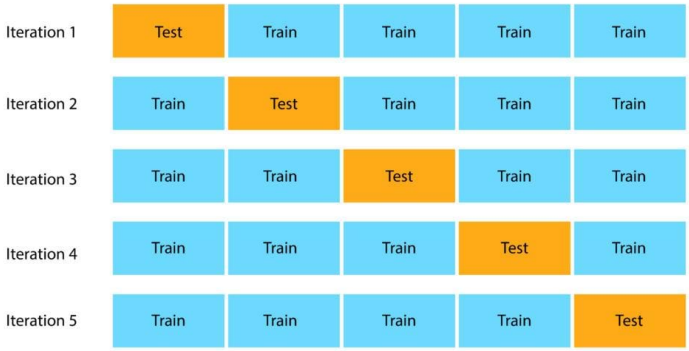


Figure 3.7: 5-fold cross-validation

3.8.2 Training and evaluation

In the training process of both the Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM) models for classification tasks, several critical steps are undertaken to enhance their performance. The initial phase of the MLP model's training involves the use of the AdamW optimizer and CrossEntropyLoss, focusing on refining predictions through a series of forward and backward passes, loss computation, and weight updates. Post the hyperparameter optimization using Optuna, the MLP model enters a conclusive training phase where both training and validation losses are closely monitored and depicted through plots, providing insights into the model's learning trajectory. Additionally, for each iteration of Leave-One-Out Cross-Validation (LOOCV), a confusion matrix is generated, offering a detailed view of the model's behavior across various classes, and a heatmap of this matrix is produced, presenting a visual summary of the model's accuracy and misclassification areas.

Similarly, the LSTM model employs the AdamW optimizer and CrossEntropyLoss with the primary goal of minimizing the loss function. Its performance is evaluated using precision, recall, and F1 scores, along with a confusion matrix for each LOOCV iteration to visualize the model's accuracy across different classes. The final phase of the LSTM model's training includes plotting the average training and validation losses over LOOCV iterations, which highlights trends in the model's development and provides a comprehensive understanding of its classification abilities and areas for improvement.

It's important to note that in this specific code, a traditional separate validation set is not explicitly created. However, the testing set in each LOOCV iteration effectively acts as a validation set, as it is used to validate the model's performance on unseen data. The model is not trained on this data, and its performance here gives insights into its generalization capabilities.

```
    for epoch in range(50): # Training loop
        total_train_loss = 0
        for inputs, targets in train_loader:
            optimizer.zero_grad()
            outputs = model(inputs)
            loss = criterion(outputs, targets)
            loss.backward()
            optimizer.step()
            total_train_loss += loss.item()

```

```
    model.eval()
    total_val_loss = 0
    y_true, y_pred = [], []
    with torch.no_grad():  # Evaluation loop
        for inputs, targets in test_loader:
            outputs = model(inputs)
            loss = criterion(outputs, targets)
            total_val_loss += loss.item()
            _, predicted = torch.max(outputs.data, 1)
            y_true.extend(targets.cpu().numpy())
            y_pred.extend(predicted.cpu().numpy())
```

Code 3.3: Training and Evalution loop

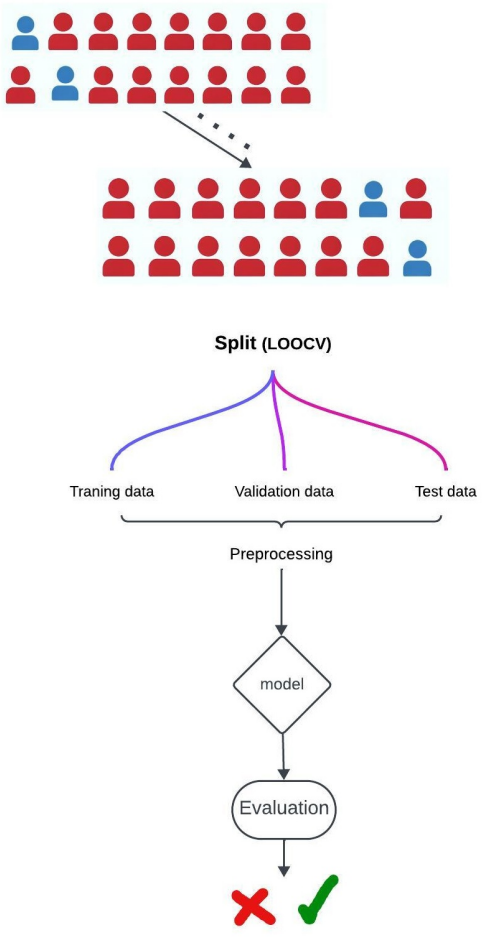


Figure 3.8: :1)Illustration of the model-data pipeline. The cross and check marks denote whether the model's evolution based on unseen data is acceptable or not.2)Schematic of Leave-One-Out Cross-Validation (LOOCV) in EEG data analysis

Below is a confusion matrix representing the classification accuracy of the LSTM model across different classes. It is generated after LOOCV iterations, illustrating the true versus predicted labels.

3.9 Implementation

Throughout this research, a versatile Python code foundation was diligently created, which expedited the process of carrying out and recording different tests. Beyond the particular outcomes produced by this study, this code base has inherent value and is also flexible for use in other experiments. It is evidence of the project's ongoing commitment to experimental and research approaches.

The principal frameworks used in the implementation of this code base were Numpy [178], Pandas [179-180], PyTorch [181], Optuna [139] and Sklearn [182]. These instruments played a crucial role in constructing a sturdy and effective system that could accommodate various experimental requirements and open the door for further investigations and discoveries.

4

Results

This chapter details the pivotal results of the thesis, adhering to the methodologies delineated in Chapter 3. It delves into the intricacies of the ANN architectures as specified in Section 3.6 and the nuanced approach to data segmentation described in Section 3.8.1. We have relied on two critical metrics in assessing our models: the F1 score and accuracy. The F1 score is averaged across the final epoch of all cross-validation folds to ensure a thorough evaluation. Simultaneously, accuracy is measured to provide a holistic view of the model's performance on unseen data, reflecting its practical predictive capabilities.

The designation of the 'best' model or hyperparameter set is reserved for those configurations that attain the highest average F1 score, with the selection process rigorously expounded in Section 3.7.2. To appraise the model's applicability in real-world contexts, we conducted tests on data from unseen participants, using accuracy as the decisive metric. This approach illuminates the model's adaptability to broader scenarios beyond the training environment.

Confusion matrices serve as a visual conduit, elucidating the superior model and hyperparameter choices based on F1 scores while demonstrating which configurations excel in accuracy concerning the unseen test data. Section 2.10 offers a compendium of F1 scores and accuracies for a more expansive analysis, meticulously cataloguing the performances of various models and parameter sweeps. This structured presentation affords a transparent view of each model's capabilities, ensuring a clear understanding of their performance in various settings.

4.1 Long Short Term Memory (LSTM)

This section explores the information obtained by using the Long Short-Term Memory (LSTM) network. The outcomes presented here demonstrate how well the LSTM can represent and understand the intricate time-series data that is associated with changes in cognitive state as a result of different VR video qualities.

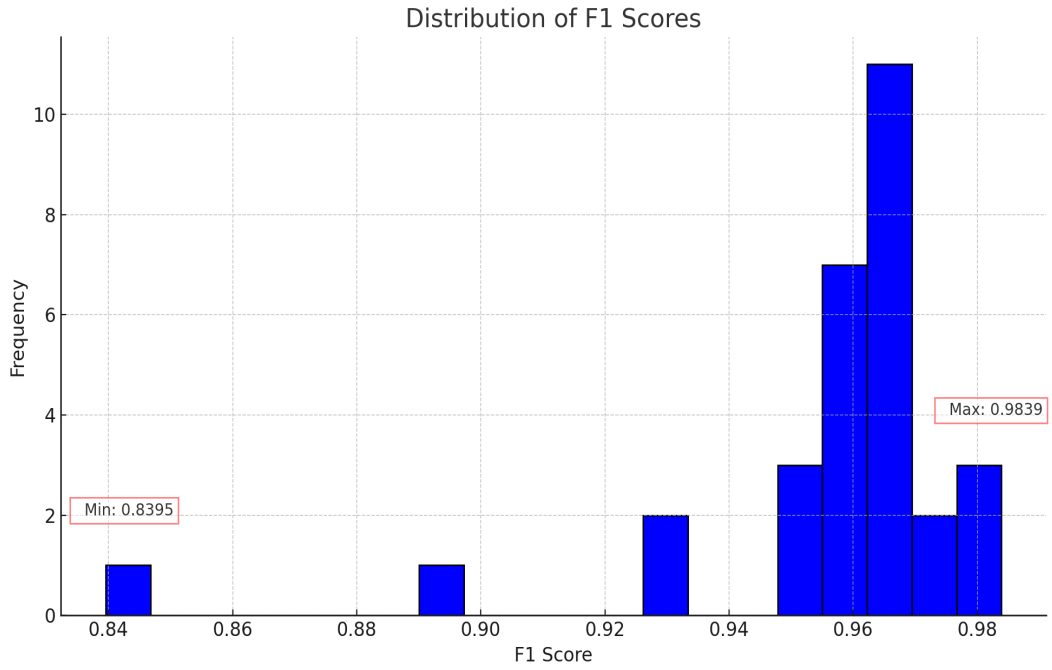


Figure 4.1: Distribution of F1 Scores from Model Evaluation, with Annotations for Minimum and Maximum Observed Values

The diagram illustrates the distribution of F1 scores across various runs of our predictive model, providing a harmonized assessment of precision and recall. Annotations within the visualization delineate the range of performance, marking the extremities of F1 scores achieved under different model configurations.

Highest F1 score

Dropout_rate	Learning_rate	num_layer	Hidden_size
0.1-0.6	1E-3 - 1E-1	1-3	64-128

Table 4.1: Parameter Range and Configuration for 100 LSTM Trials Conducted Using OPTUNA Framework

best_learning_rate	best_dropout_rate	hidden_size	num_layer
0.0011670	0.597456	93	1

Table 4.2: Optimal Hyperparameter Values for LSTM Model Determined by OPTUNA Optimization

Ave_precision	Ave_recall	Ave_F1	best_trial_value
0.982758	0.986206	0.983908	Best is trial 72 with value: -0.980

Table 4.3: Performance Metrics for LSTM Model Optimization

Tables (4.1, 4.2, and 4.3) display the fine-tuning steps and performance metrics for an LSTM model optimized through the OPTUNA framework, pinpointing hyperparameters that culminated in an optimal F1 score. This measure of precision and recall provides a nuanced view of the model's validation dataset performance.

However, the model's prowess on unseen data is a different story. An F1 score near 0.983908 suggests a model that balances precision and recall adeptly, but a stark contrast is observed with an accuracy of around 30% on new data. This significant drop raises questions about the model's capability to use what it has learned in different datasets. Insight into this performance gap comes from the loss plot in Figure 4.2, which tracks average training and validation losses throughout the LOOCV iterations. The descending trend and eventual convergence of these losses hint at a model that is not overfitting, yet the reality of unseen data challenges this notion. The factors contributing to this disparity are multifaceted:

- A Data Distribution Shift could imply the model's inability to adapt to new data patterns despite its success with similar validation data.
- The Model's Generalization Capability is called into question; the loss convergence does not equate to a comprehensive grasp of the data's underlying structure.
- A high F1 score only sometimes aligns with high Accuracy, a more direct performance measure on unseen data and can expose overlooked weaknesses.

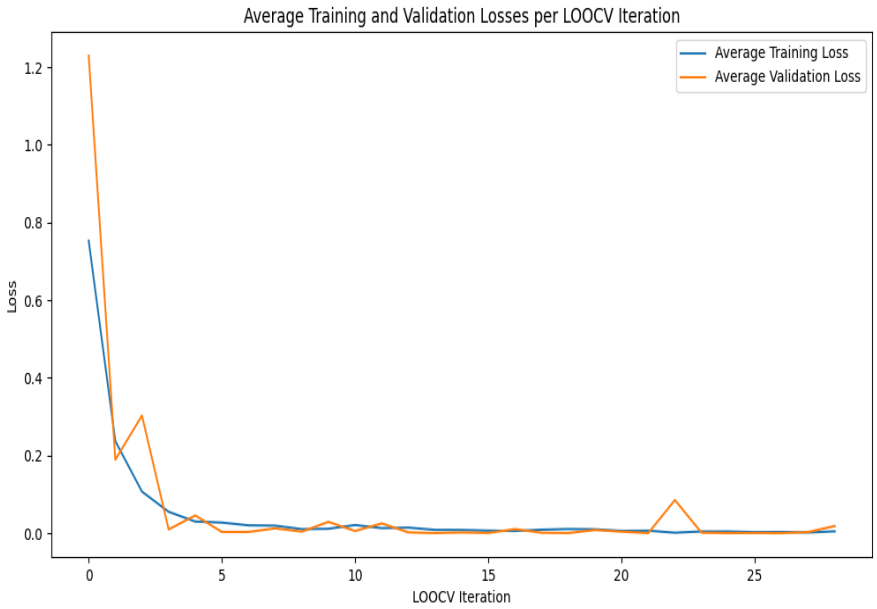


Figure 4.2: Trend of Average Training and Validation Losses Across Leave-One-Out Cross-Validation (LOOCV) Iterations Highest F1 configuration

- The Quality of Unseen Data may differ significantly from the training set, leading to mispredictions if new, noisy, or outlier features are introduced.

Although the model demonstrates promising performance within the training and validation scope, as indicated by the F1 score and loss trajectories, the diminished Accuracy of unseen data must be noticed. It signals the necessity for a deeper dive into the unseen data's idiosyncrasies, reevaluation of the model's generalization, and consideration of a broader array of architectural or preprocessing strategies to enhance the model's performance in predicting new data.

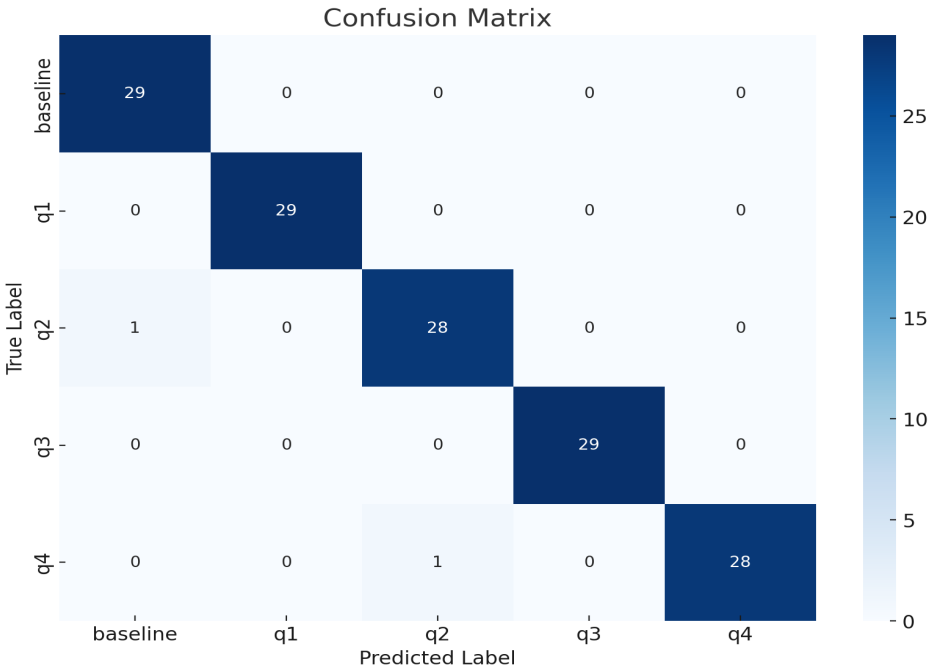


Figure 4.3: Confusion Matrix of Video Quality Predictions by LSTM

Highest Accuracy

Dropout_rate	Learning_rate	num_layer	Hidden_size
0.1-0.4	1E-5 - 1E-2	1-3	16-64

Table 4.4: Parameter Range and Configuration for 100 LSTM Trials Conducted Using OPTUNA Framework

best_learning_rate	best_dropout_rate	hidden_size	num_layer
0.00167924	0.1109262	55	1

Table 4.5: Optimal Hyperparameter Values for LSTM Model Determined by OPTUNA Optimization

Ave_precision	Ave_recall	Ave_F1	best_trial_value
0.96781	0.972413	0.968965	Best is trial 45 with value: 0.981379

Table 4.6: Performance Metrics for LSTM Model Optimization

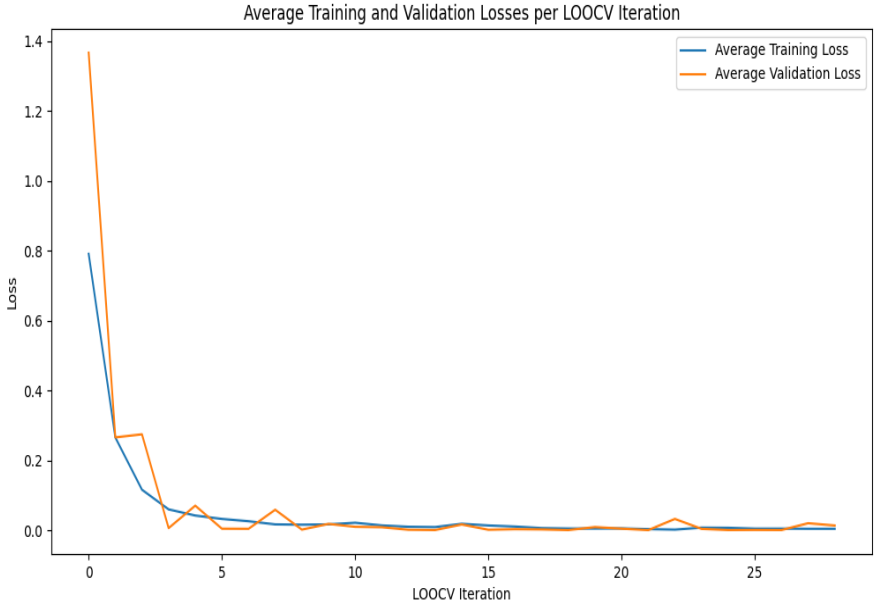


Figure 4.4: Trend of Average Training and Validation Losses Across Leave-One-Out Cross-Validation (LOOCV) Iterations

In the previous section, we thoroughly examined the configuration of the LSTM model, which achieved an outstanding F1 score, as shown in Table 4.3. The F1 score is crucial for imbalanced datasets as it balances precision and recall, comprehensively measuring a model's predictive accuracy. Now, we focus on a broader and equally important metric: accuracy.

Table 4.4 introduces this exploration by detailing the range of hyperparameters tested in the OPTUNA framework. This thorough testing included various dropout rates, learning rates, layer counts, and sizes of hidden units, leading to the findings presented in Table 4.5. The optimal hyperparameter set identified includes a learning rate of 0.0016, a dropout rate of 0.1109, a single LSTM layer, and a hidden size 55. These parameters were chosen for their exceptional performance potential.

Figure 4.5 features a confusion matrix visually demonstrating the model's generalization capabilities. With an accuracy rate of 0.70, the matrix indicates

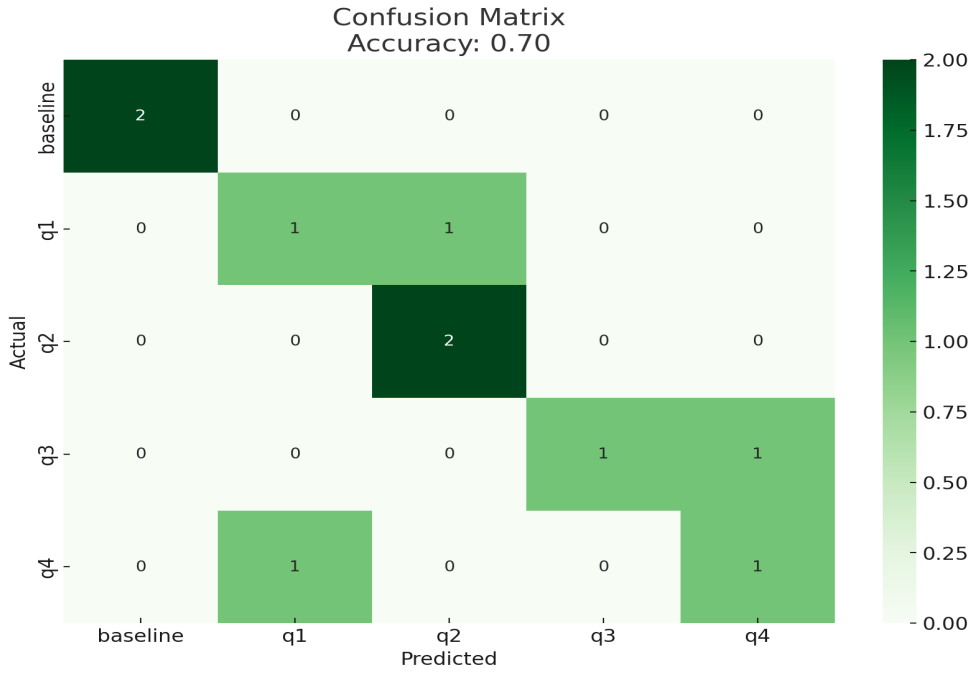


Figure 4.5: Confusion Matrix of Video Quality Predictions by LSTM on Unseen Data

the model's proficiency in handling unseen data and its reliability in classifying video quality. The matrix's diagonal, filled with correct predictions, highlights the model's strengths, while the sparsity of errors in the off-diagonal elements shows its resilience.

Accuracy is especially crucial in situations where errors in classification have equal consequences across all classes. In such cases, a model's high accuracy is synonymous with its trustworthiness. The LSTM model configuration stands out not only for its balanced F1 score but also for its consistent accuracy with new data.

The development of this model represents a journey of methodological rigour and empirical validation. From determining the best hyperparameters to validating the model through Leave-One-Out Cross-Validation (LOOCV) and its performance on new data, this process narrates a story of comprehensive learning. It transitions from theoretical optimization to practical application, ensuring the final model is well-suited for real-world use.

4.2 Multilayer Perceptron (MLP)

This section shows the outcomes produced using MLP models. In particular, the optimal performance was attained by varying the hyperparameter tuning of Optuna.

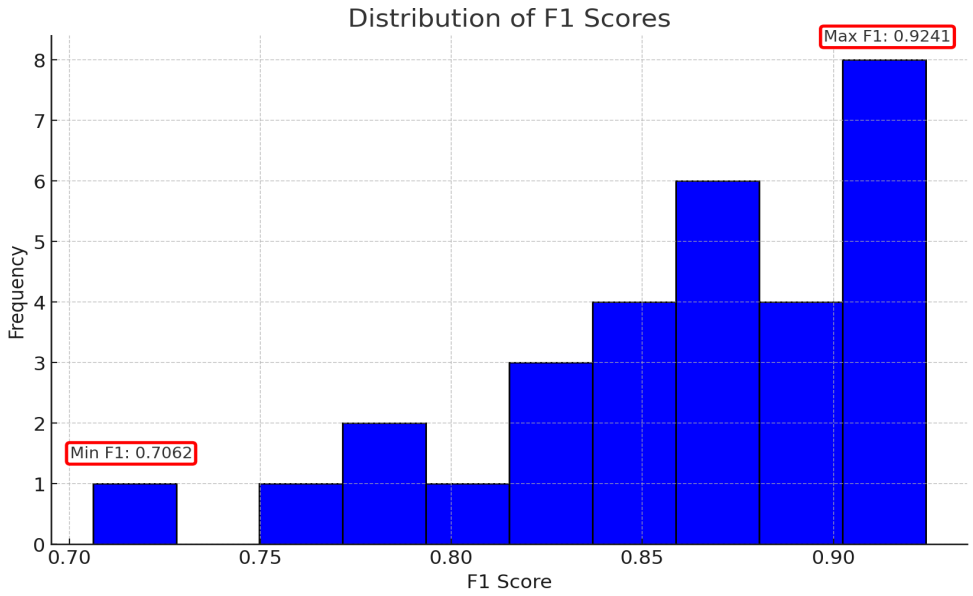


Figure 4.6: Distribution of F1 Scores from Model Evaluation, with Annotations for Minimum and Maximum Observed Values

Figure 4.6 presents the distribution of F1 scores throughout several iterations of our prediction model. Annotations in the visualization indicate the performance range, indicating the upper and lower bounds of F1 scores obtained with various model setups.

Highest F1

Dropout_rate	Learning_rate	Weight_decay	num_layer
0.1-0.7	1E-4 - 1E-1	1e-4	128-32

Table 4.7: Parameter Range and Configuration for 100 MLP Trials Conducted Using OPTUNA Framework

best_learning_rate	best_dropout_rate
0.0001223	0.1127674

Table 4.8: Optimal Hyperparameter Values for MLP Model Determined by OPTUNA Optimization

Ave_precision	Ave_recall	Ave_F1
0.918390	0.937931	0.9241379

Table 4.9: Performance Metrics for MLP Model Optimization

To develop a robust MLP model for cognitive load classification, an exhaustive evaluation of hyperparameters was undertaken, as chronicled in Table 4.7. This meticulous exploration, orchestrated via the OPTUNA framework, involved scrutinizing dropout rates, learning rates, weight decay parameters, and the depth of neural layers to determine the most effective model configuration. The academic inquiry was not confined to theoretical parameterization but extended to empirical performance, as evidenced by the loss patterns depicted in Figure 4.7.

In addition, the results of this thorough parameter search are summarised in Table 4.8, where the ideal hyperparameters are determined to have a learning rate of 0.00012 and a dropout rate of 0.11276. Carefully choosing these parameters demonstrates a delicate equilibrium between the complexity of the model and its generalizability, which is essential to prevent overfitting. The F1 score in Table 4.9 provides additional evidence for this delicate equilibrium. The model achieves an exceptional average score of 0.924, which encapsulates its precision and recall and highlights its balanced performance on the training data.

The learning process of the model is demonstrated in Figure 4.7 by the convergence and oscillations of validation and training losses over LOOCV iterations. The graph is noteworthy because it shows abrupt rises and drops in validation loss, indicating that the model is sensitive to the peculiarities of various data subsets

Figure 4.7: Trend of Average Training and Validation Losses Across Leave-One-Out Cross-Validation (LOOCV) Iterations

met during validation. The model's flexibility is shown in these undulations; every peak and trough signifies a re-calibration to account for the unique features found in every validation set.

The sensitivity to the data's structure, as portrayed in the loss plot, is revealing. It indicates the model's finesse in fine-tuning the nuanced features within the data. However, it also highlights the inherent challenge: maintaining consistent performance across varying data landscapes. The model's accuracy on unseen data, a modest 40%, juxtaposed with the highest F1 score, prompts a critical evaluation of the disparity between theoretical optimization and practical application.

This dichotomy between the model's theoretical understanding, as indicated by the high F1 score, and its practical applicability, as reflected by accuracy, emphasizes a fundamental aspect of machine learning models. While the F1 score implies a robust internal consistency, the accurate measure of the model's utility is its performance on new, unseen data.

Integrating parameter optimization results with the performance evaluation on unseen data unveils the intricate interplay between a model's learning capabilities and its capacity to generalize. The oscillations in the loss plot metaphorically echo

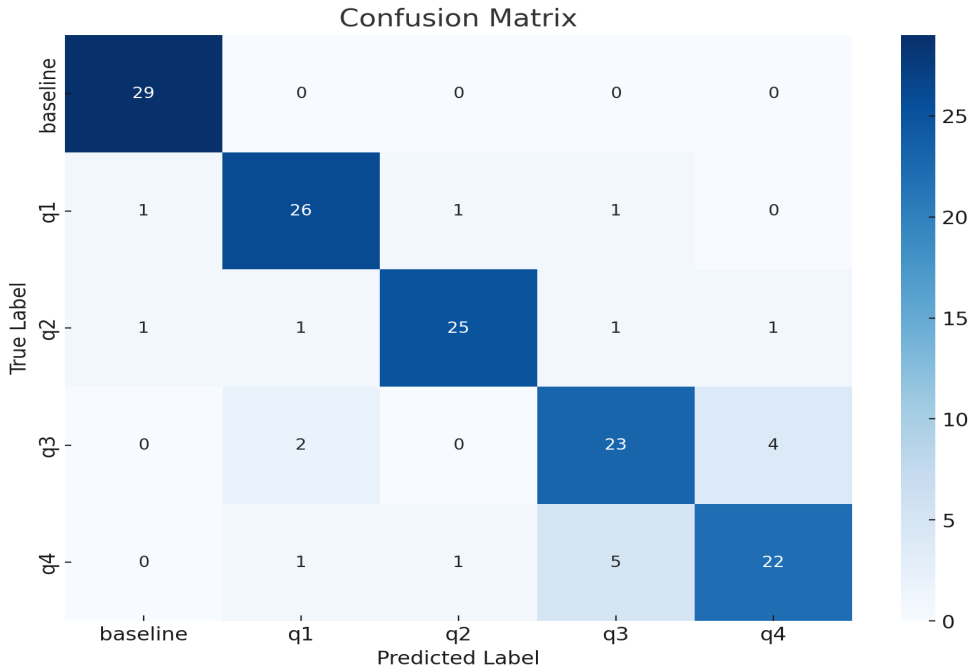


Figure 4.8: Confusion Matrix of Video Quality Predictions by MLP

the model's navigation through complex data terrains. As we traverse from the theoretical confines of parameter tuning to the practical realms of unseen data, the model's readiness for real-world deployment is rigorously assessed. Through this comprehensive lens of practical performance, the actual value of the MLP model is discerned, substantiating its potential for academic and practical applications in cognitive load assessment.

Highest Accuracy

Dropout_rate	Learning_rate	Weight_decay	num_layer
0.5-0.9	1E-5 - 1E-1	1e-4	128-32

Table 4.10: Parameter Range and Configuration for 100 MLP Trials Conducted Using OPTUNA Framework

best_learning_rate	best_dropout_rate
0.00006942	0.5169154

Table 4.11: Optimal Hyperparameter Values for MLP Model Determined by OPTUNA Optimization

Ave_precision	Ave_recall	Ave_F1
0.7459770	0.8206896	0.7687356

Table 4.12: Performance Metrics for MLP Model Optimization

Figure 4.9: Trend of Average Training and Validation Losses Across Leave-One-Out Cross-Validation (LOOCV) Iterations

The thorough analysis of the Multilayer Perceptron (MLP) model, shown in Tables 4.11 and 4.12 and illustrated in Figures 4.9 and 4.11, leads to a sophisticated comprehension of the model's capacity to decipher cognitive load from ECG data.

The ideal hyperparameters obtained by the methodical OPTUNA optimization framework are presented in Table 4.11, highlighting the optimized dropout and learning rates. These parameters are essential because they guarantee the model's accuracy in spotting patterns and guard against overfitting, guaranteeing strong generalization in the face of fresh data.

The performance metrics provided in Table 4.12 reflect the model's precision, recall, and F1 score, which collectively signal the model's balanced accuracy in classification. These scores, however, are theoretical until validated against unseen data. Figure 4.9 presents the model's learning curve through LOOCV iterations, showcasing a downward trend in training and validation losses, indicative of the

model's learning efficacy and ability to generalize.

The model's theoretical understanding is put to the test with the introduction of data from two unseen participants, as depicted in Figure 4.11. This confusion matrix provides an accuracy measure of 0.60, revealing the model's practical capabilities in a real-world scenario. The matrix demonstrates the model's classification performance across different video quality levels, offering insights into its applicability and adaptability to individual variations.

The amalgamation of these findings—the rigorous hyperparameter optimization and the model's validated performance on unseen data—illustrates the MLP model's robustness in cognitive load classification. This theoretical robustness is now evidenced through the model's application to new participant data, validating its potential for academic and practical applications. The ability to achieve a 60% accuracy rate on unseen data highlights the model's readiness for deployment in diverse settings, bridging the gap between controlled training environments and the unpredictability of real-world applications.

The journey from parameter optimization to real-world data evaluation encapsulates the intricate balance between a model's learning abilities and its generalization capabilities. The results collectively affirm the MLP model's position as a valuable tool for cognitive load assessment, spotlighting its significance for research and practical applications in understanding human behaviour.

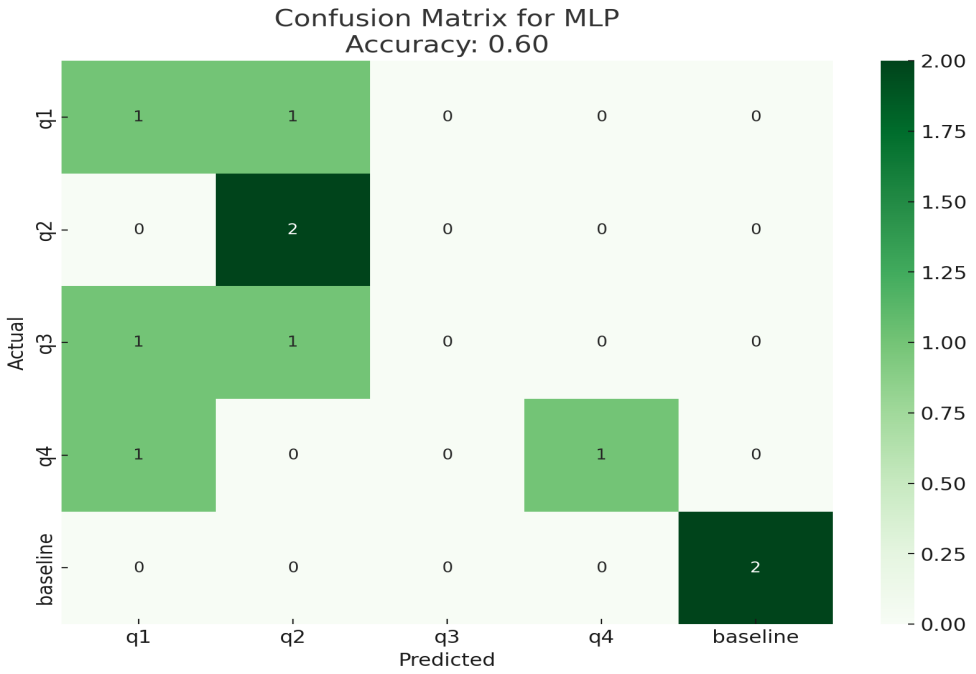


Figure 4.10: Confusion Matrix of Video Quality Predictions by MLP on Unseen Data

5 Conclusion

This chapter is dedicated to a comprehensive discussion and comparison of the results outlined in Chapter 4. It explores the project's accomplishments as well as the difficulties that emerged during implementation. Moreover, the chapter delves into possible directions for future research, extending the framework established by this thesis.

5.1 Achievement

This Master's thesis significantly advances the classification of video quality from EEG data by initially leveraging machine learning (ML) techniques. The research focuses on the impact of video quality degradation on the viewer's quality of experience (QoE) in virtual reality (VR) environments, aiming to uncover the nuanced relationship between video quality and the user's cognitive state in these settings. By integrating ML, this study lays the groundwork for understanding and enhancing how video quality influences cognitive load in VR, a concept further explored and supported by the findings of Zheleva et al. [29], which associated high-quality videos with cognitive load.

A pivotal methodological advancement in this research was the optimized division of data between training and validation sets, which significantly improved hyperparameter optimization accuracy. This optimization facilitated a deeper investigation into various architectures, hyperparameters, and preprocessing methods, focusing on the Long Short-Term Memory (LSTM) and Multilayer Perceptron (MLP) models. Although the LSTM model showed promise in training and valida-

tion performance, its generalization to unseen data was hampered by the limited number of new participants. Conversely, the MLP model exhibited balanced accuracy across different datasets, albeit with challenges in applying theoretical models to new, real-world data.

The study meticulously integrates parameter optimization with performance evaluation, shedding light on the intricate balance between a model's learning capabilities and its potential for generalization. This balanced approach has proven the MLP model to be a practical tool for assessing cognitive load, effectively bridging the gap between controlled training environments and real-world applications. A significant insight from this research is the critical role of dataset size in determining model performance, highlighting that deep learning models require substantial data to be truly effective and suggesting that traditional statistical methods may sometimes offer a more feasible alternative.

Preliminary results from LSTM-based classification systems provide promising directions for future EEG-based video quality assessment research, indicating potential areas for further exploration. In conclusion, this thesis contributes substantially to the field by methodically evaluating and validating the performance of LSTM and MLP models, thus providing a robust foundation for future investigations and practical applications in EEG-based video quality assessment.

5.2 Future work

Future work following this Master's thesis could focus on several avenues. Firstly, expanding the EEG data set would be pivotal for enhancing the robustness and generalization of the machine learning models. Delving into additional deep learning architectures like Convolutional Neural Networks could offer new perspectives in EEG-based video quality classification. Further development of advanced data preprocessing methods is also crucial to accurately capture the nuances of EEG signals. Exploring real-time classification of video quality in VR could significantly enhance the user experience by allowing immediate adjustments. Another promising area is examining different types of VR content and their impact on EEG responses and perceived quality. Moreover, conducting user-centric studies to understand individual differences in perception and response to VR video quality could lead to more personalized VR experiences. Finally, interdisciplinary research, especially in collaboration with cognitive psychology, could deepen the understanding of cognitive processes in VR experiences as reflected in EEG data.

These areas offer a comprehensive roadmap for advancing the understanding and application of EEG data in VR technology.

6

References

[1] Onyesolu, M. O., & Eze, U. F. (2011). Understanding Virtual Reality Technology: Advances and Applications. In Advances in Computer Science and Engineering. DOI:10.5772/15529.

[2] Kanev, K., & Sugiyama, T. (1998). Design and simulation of interactive 3D computer games. Computers & Graphics, 22(2–3), 281-300. https://doi.org/10.1016/S0097-8493(98)00038-7

[3] S. Kumari and N. Polke, Implementation Issues of Augmented Reality and Virtual Reality: A Survey, vol. 26. Springer International Publishing, 2019.

[4] Sutherland, I. (1965). "The ultimate display," in Proceedings of the International Federation of Information Processing (IFIP) (New York, NY), 506–508.

[5] Stengel, M., Grogorick, S., Eisemann, M., Eisemann, E., & Magnor, M. A. (2015). An affordable solution for binocular eye tracking and calibration in head-mounted displays. Proceedings of the 23rd ACM international conference on Multimedia (pp. 15-24). https://doi.org/10.1145/2733373.2806265.

[6] Byagowi, A., Singhal, S., Lambeta, M., Aldaba, C., & Moussavi, Z. (2014). Design of a naturalistic navigational virtual reality using oculus rift. Journal of Medical Devices, 8(3), 030946. doi: 10.1115/1.4027114

[7] Barnard, D. (2023, June 27). Degrees of Freedom (DoF): 3-DoF vs 6-DoF for VR Headset Selection. VirtualSpeech. https://virtualspeech.com/blog/degrees-of-freedom-vr

[8] Callet, P.L., Moller, S., Perkis, A.: Qualinet white paper on definitions of quality of experience. Eu- ropean Network on Quality of Experience in 1(March) (2012)

[9] Installations, T., Line, L., Systems, D.: Itu-t definition of quality of experience (qoe) - p.10/g.100. Itu 100(2006), 2 (2007).

[10] Reiter, U., Brunnstro¨m, K., De Moor, K., Larabi,M.C., Pereira, M., Pinheiro, A., You, J., Zgank, A.: Factors Influencing Quality of Experience. In: T-Labs Series in Telecommunication Services, pp. 55–72. Springer Science and Business Media B.V.

[11] Ventura, S., Brivio, E., Riva, G., & Baños, R. M. (2019). Immersive versus non-immersive experience: Exploring the feasibility of memory assessment through 360° technology. Frontiers in Psychology, 10.
https://doi.org/10.3389/fpsyg.2019.02509

[12] Baddeley, A. Working memory. 1986, New York: Oxford University Press.

[13] Baddeley, A. Working memory: Looking back and looking forward. Neuroscience, 2003. 4: 829-839.

[14] Sweller, J. Cognitive load during problem solving: Effects on learning. Cognitive Science, (1988). 12: 257-285.

[15] van Merrienboer, J. and Sweller, J. Cognitive load theory and complex learning: Recent developments and future directions. Educational Psychology Review, 2005. 17(2): 147-177.

[16] Paas, F., Tuovinen, J., Tabbers, H., and Van Gerven, P. Cognitive load measurement as a means to advance cognitive load theory. Educational Psychologist, 2003. 38(1): 63-71.

[17] Ho, C. and Spence, C. Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. Journal of Experimental Psychology: Applied, 2005. 11: 157-174.

[18] John, M., Kobus, D., and Morrison, J. Augmented cognition technical integration experiment (TIE). Technical Report 1905, December 2003, DARPA .

[19] Oviatt, S.L., Arthur, A.M., and Cohen, J. Quiet interfaces that help students think. In Proceedings of the Nineteenth Annual ACM Symposium on User Interface Software and Technology (UIST '06). 2006. New York, NY: ACM Press: in press.

[20] Wickens, C. Multiple resources and performance prediction. Theoretical Issues in Ergonomic Science, 2002. 3(2): 159- 177.

[21] Wickens, C., Sandry, D., and Vidulich, M. Compatibility and resource competition between modalities of input, central processing, and output. Human Factors, 1983. 25(2): 227-248.

[22] Tindall-Ford, S., Chandler, P., and Sweller, P. When two sensory modes

are better than one. Journal of Experimental Psychology: Applied, 1997. 3(3): 257-287.

[23] Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63, 81–97.

[24] Peterson, L., & Peterson, M. (1959). Short-term retention of individual verbal items. Journal of Experimental Psychology, 58, 193–198.

[25] Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12, 257–285.

[26] Klimesch, W., Schack, B., & Sauseng, P. (2005). The functional significance of theta and upper alpha oscillations for working memory: A review. Experimental Psychology, 52, 99–108.

[27] Basar, E. (1980). EEG brain dynamics: Relation between EEG and brain evoked potentials. Amsterdam: Elsevier.

[28] Basar, E. (1999). Brain function and oscillations. Integrative brain functions. Neurophysiology and Cognitive Processes, Vol. 2. Berlin: Springer.

[29] Zheleva, A., Durnez, W., Bombeke, K., van Wallendael, G., & de Marez, L. (2020). Seeing is Believing: The Effect of Video Quality on Quality of Experience in Virtual Reality. 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), 1–4.
https://doi.org/10.1109/QoMEX48832.2020.9123075.

[30] Desai, P. R., Desai, P. N., Ajmera, K. D., & Mehta, K. (2014). A Review Paper on Oculus Rift-A Virtual Reality Headset. arXiv:1408.1173 [cs.HC]. https://doi.org/10.48550/arXiv.1408.1173.

[31] Brooks Jr, F. P. (1999). What's real about virtual reality?. Computer Graphics and Applications, IEEE, 19(6), 16-27. doi: 10.1109/38.799723

[32] W. Piekarski, B. Gunther, and B. Thomas,Integrating virtual and augmented realities in an outdoor application, Proc. - 2nd IEEE ACM Int.Work. Augment. Reality, IWAR 1999, no. February 1999, pp. 45–54, 1999, doi: 10.1109/IWAR.1999.803805

[33] M. K. Bekele, R. Pierdicca, E. Frontoni, E. S. Malinverni, and J. Gain, A survey of augmented, virtual, and mixed reality for cultural heritage, J. Comput. Cult. Herit., vol. 11, no. 2, 2018, doi: 10.1145/3145534.

[34] Sanchez-Vives, Maria V., and Mel Slater. "From Presence to Consciousness Through Virtual Reality." Nature Reviews Neuroscience, vol. 6, 2005, pp. 332-339. https://doi.org/10.1038/nrn1651.

[35] Gaggioli, Andrea, Marta Bassi, and A. D. Fave. "Quality of experience in

virtual environments." Emerging Communication 5 (2003): 121-136.

[36] Perkis,A.,Timmerer,C.,Barakovi´c,S.,Husi´c, J.B., Bech, S., Bosse, S., Botev, J., Brunnstro¨m, K., Cruz, L., De Moor, K., Saibanti, A.d.P., Durnez, W., Egger-Lampl, S., Engelke, U., Falk, T.H., Hameed, A., Hines, A., Kojic, T., Kukolj, D., Liotou, E., Milovanovic, D., Mo¨ller, S., Mur- ray, N., Naderi, B., Pereira, M., Perry, S., Pin- heiro, A., Pinilla, A., Raake, A., Agrawal, S.R., Reiter, U., Rodrigues, R., Schatz, R., Schelkens, P., Schmidt, S., Sabet, S.S., Singla, A., Skorin- Kapov, L., Suznjevic, M., Uhrig, S., Vlahovi´c, S., Voigt-Antons, J.N., Zadtootaghaj, S.: Qualinet white paper on definitions of immersive media ex- perience. European Network on Quality of Experience in Multimedia Systems and Services (2020).

[37] Sweller, J., Van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. Educational Psychology Review, 10(3), 251-296. https://doi.org/10.1023/A:1022193728205

[38] Mostyn, G. R. (2012). Cognitive load theory: What it is, why it's important for accounting instruction and research. Issues in Accounting Education, 27(1), 227-245. https://doi.org/10.2308/iace-50099

[39] Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences, 24, 87–114.

[40] Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 1–66.

[41] Sweller, J. (2010). Element interactivity and intrinsic, extraneous and germane cognitive load. Accepted for publication in Educational Psychology Review. doi:10.1007/s10648-010-9128-5.

[42] Krashen, S. (1982). Principles and practice in second language acquisition. New York: Prentice-Hall.

[43] Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. British Journal of Educational Psychology, 62, 233–246.

[45] Haas, L. F. (2003). Hans Berger (1873-1941), Richard Caton (1842-1926), and electroencephalography. Journal of Neurology, Neurosurgery, and Psychiatry, 74(1), 9. https://doi.org/10.1136/jnnp.74.1.9

[46] Anderson, C.W., Bratman, J.A. 2008. Translating thoughts into actions by finding patterns in brainwave. In Proceedings of the Fourteenth Yale Workshop on Adaptive and Learning Systems, Yale University, New Haven, CT, June 2008, pp. 1–6.

[47] Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. Brain Research Reviews, 29, 169–195

[48] Berka, C., Levendowski, D. J., Olmstead, R. E., Popovic, M. V., Cvetinovic, M., & Petrovic, M. M. (2004). Real-time analysis of EEG indices of alertness, cognition, and memory with a wireless EEG headset. International Journal of Human-Computer Interaction, 17, 151–170.

[49] Stevens, R., Galloway, T., & Berka, C. (2007). EEG-related changes in cognitive workload, engagement and distraction as students acquire problem solving skills. In C. Conati, K. McCoy, & G. Paliouras (Eds.) User Modeling 2007. Lecture Notes in Artificial Intelligence, 4511, 197–206.

[50] Pfurtscheller, G., & Aranibar, A. (1977). Event-related cortical desynchronization detected by power measurements of scalp EEG. Electroencephalography and Clinical Neurophysiology, 42, 817–826.

[51] C. E. Shannon. "A Mathematical Theory of Communication." Reprinted with corrections from The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, July, October 1948. Introduction.

[52] Hall, M. A. (1998). Correlation-based feature selection for machine learning. PhD thesis. University of Waikato, Department of Computer Science.

[53] P. Zarjam, J. Epps, and N. H. Lovell, "Beyond Subjective Self-Rating: EEG Signal Classification of Cognitive Workload," IEEE Transactions on Autonomous Mental Development, vol. 7, no. 4, pp. 301-310, 2015.

[54] Huang, S. H. (2015). "Supervised feature selection: a tutorial". Artifical Intelligence Research 4:2, pp. 22–32.

[55] Penny, W. D., K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols (2011). Statistical parametric mapping: the analysis of functional brain images. Aca- demic press, Great Britain.

[56] Tremmel, C., Herff, C., Sato, T., Rechowicz, K., Yamani, Y., & Krusienski, D. J. (2019). Estimating Cognitive Workload in an Interactive Virtual Reality Environment Using EEG. Frontiers in Human Neuroscience, 13, Article 401. https://doi.org/10.3389/fnhum.2019.00401

[57] https://jasp-stats.org/wp-content/uploads/2020/11/
Statistical-Analysis-in-JASP-A-Students-Guide-v14-Nov2020.pdf

[58] Chang, E., Kim, H.-T., & Yoo, B. (2020). Virtual Reality Sickness: A Review of Causes and Measurements. International Journal of Human-Computer Interaction, 36(11), 1-25. https://doi.org/10.1080/10447318.2020.1778351

[59] Tian, F., Wang, X., Cheng, W., Lee, M., & Jin, Y. (2022). A Comparative Study on the Temporal Effects of 2D and VR Emotional Arousal. Sensors, 22(21), 8491. https://doi.org/10.3390/s22218491

[60] F. Dehais, A. Duprès, S. Blum, N. Drougard, S. Scannella, R. Roy, and F. Lotte, "Monitoring Pilot's Mental Workload Using ERPs and Spectral Power with a Six-Dry-Electrode EEG System in Real Flight Conditions," Sensors, vol. 19, no. 6, 2019.

[61] Van Emmerik, M.L., De Vries, S.C., Bos, J.E.: Internal and external fields of view affect cybersickness.
Displays 32(4), 169–174 (2011). DOI 10.1016/j.displa.2010.11.003.

[62] Weech, S., Kenny, S., Barnett-Cowan, M.: Presence and cybersickness in virtual reality are negatively related: A review (2019). DOI 10.3389/ fpsyg.2019.00158.

[63] Sas, C.: Individual differences in virtual environments. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial In- telligence and Lecture Notes in Bioinformatics) 3038, 1017–1024 (2004).

[64] Hughes JR. EEG in clinical practice. Boston: Butterworth-Heinemann; 1994.

[65] Atkinson, R.C. and Shiffrin, R.M. (1968). 'Human memory: A Proposed System and its Control Processes'. In Spence, K.W. and Spence, J.T. The psychology oflearning and motivation, (Volume 2). New York: Academic Press. pp. 89–195.

[66] Burle, B., Spieser, L., Roger, C., Casini, L., Hasbroucq, T., & Vidal, F. (2015). Spatial and temporal resolutions of EEG: Is it really black and white? A scalp current density view. https://doi.org/10.1016/j.ijpsycho.2015.05.004

[67] De Letter, J., Zheleva, A., Maes, M., All, A., De Marez, L., Durnez, W.: What did you expect? Quality and User Experience 6(1), 1–13 (2021). DOI 10.1007/s41233-021-00045-6.

[68] Handy, T. C. (2005). Event-related potentials: A methods handbook. MIT press.

[69] Niedermeyer, E., & da Silva, F. L. (Eds.). (2005). Electroencephalography: basic principles, clinical applications, and related fields. Lippincott Williams & Wilkins.

[70] Pfurtscheller, G., & Da Silva, F. L. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. Clinical neurophysiology, 110(11), 1842-1857. doi: 10.1016/s1388-2457(99)00141-8

[71] Palva, S., & Palva, J. M. (2007). New vistas for -frequency band oscillations. Trends in neurosciences, 30(4), 150-158. doi: 10.1016/j.tins.2007.02.001.

[72] Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. Biological psychology, 41(2), 103-146.
doi: 10.1016/0301-0511(95)05130-9

[73] R. Jenke, A. Peer, and M. Buss, "Feature Extraction and Selection for Emotion Recognition from EEG," IEEE Transactions on Affective Computing, vol. 5, no. 3, pp. 327-339, 2014.

[74] L. Zhang, J. Wade, D. Bian, J. Fan, A. Swanson, A. Weitlauf, Z. Warren, and N. Sarkar, "Cognitive Load Measurement in a Virtual Reality-Based Driving System for Autism Intervention," IEEE Transactions on Affective Computing, vol. 8, no. 2, pp. 176-189, 2017.

[75] A. Kok, "On the utility of P3 amplitude as a measure of processing capacity," Psychophysiology, vol. 38, no. 3, pp. 557-577, 2010.

[76] A. M. Brouwer, M. A. Hogervorst, J. B. F. Van Erp, T. Heffelaar, P. H. Zimmerman, and R. Oostenveld, "Estimating workload using EEG spectral power and ERPs in the n-back task," Journal of Neural Engineering, vol. 9, no. 4, pp. 045008, 2012.

[77] A. Subasi, "EEG signal classification using wavelet feature extraction and a mixture of expert model," Expert Systems with Applications, vol. 32, no. 4, pp. 1084-1093, 2007.

[78] A. Kabbara, "Brain network estimation from dense EEG signals: application to neurological disorders," Neurons and Cognition, 2018.

[79] G. F. Wilson, and C. A. Russell, "Performance Enhancement in an Uninhabited Air Vehicle Task Using Psychophysiologically Determined Adaptive Aiding," Human Factors, vol. 49, no. 6, pp.1005-1018, 2007.

[80] C. Mühl, C. Jeunet, and F. Lotte, "EEG-based workload estimation across affective contexts," Frontiers Neurosci., vol. 8, p. 114, Jun. 2014.

[81] P. Zarjam, J. Epps, and F. Chen, "Characterizing working memory load using EEG delta activity," in Proc. 19th Eur. Signal Process.Conf., Aug. pp. 1554–1558, 2011.

[82] C. Dijksterhuis, D. de Waard, K. Brookhuis, B. Mulder, and R. de Jong, "Classifying visuomotor workload in a driving simulator using subject specific spatial brain patterns," Frontiers Neurosci., vol. 7, p. 149, Aug. 2013.

[83] P. Zarjam, J. Epps, F. Chen, and N. H. Lovell, "Estimating cognitive workload using wavelet entropy-based features during an arithmetic task," Computers

in Biology and Medicine, vol. 43, no. 12, pp. 2186-2195, 2013.

[84] G. Borghini et al., "Assessment of mental fatigue during car driving by using high resolution EEG activity and neurophysiologic indices," in Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Aug. 2012, pp. 6442–6445.

[85] C. Walter, S. Schmidt, W. Rosenstiel, P. Gerjets, and M. Bogdan, "Using Cross-Task Classification for Classifying Workload Levels in Complex Learning Tasks," 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, pp. 876-881, 2013.

[86] Z. Wang, R. M. Hope, Z. Wang, Q. Ji, and W. D. Gray, "Cross-subject workload classification with a hierarchical Bayes model," Neuroimage, vol. 59, no. 1, pp. 64-69, 2012.

[87] M. K. Kıymık, İ. Güler, A. Dizibüyük, and M. Akın, "Comparison of STFT and wavelet transform methods in determining epileptic seizure activity in EEG signals for real-time application," Computers in Biology and Medicine, vol. 35, no. 7, pp. 603-616, 2005.

[88] D. P. Subha, P. K. Joseph, A. U. Rajendra, and C. M. Lim, "EEG signal analysis: a survey," Journal of Medical Systems, vol. 34, no. 2, pp. 195-212, 2010.

[89] Y. Ma, W. Shi, C. K. Peng, and A. C. Yang, "Nonlinear dynamical analysis of sleep electroencephalography using fractal and entropy approaches," Sleep Medicine Reviews, pp. S1087079217300187, 2017.

[90] Y. Bai, X. Li, and Z. Liang, "Nonlinear Neural Dynamics," EEG Signal Processing and Feature Extraction, L. Hu and Z. Zhang, eds., pp. 215-240, Singapore: Springer Singapore, 2019.

[91] Y. Tian, H. Zhang, Y. Jiang, P. Li, and Y. Li, "A Fusion Feature for Enhancing the Performance of Classification in Working Memory Load With Single-Trial Detection," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 27, no. 10, pp. 1985-1993, 2019.

[92] M. I. Jordan, and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," Science, vol. 349, no. 6245, pp. 255-260, 2015.

[93] X. Yang, Y. Wang, R. Byrne, G. Schneider, and S. Yang, "Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery," Chemical Reviews, vol. 119, no. 18, pp. 10520-10594, 2019.

[94] M. F. Ansari, D. R. Edla, S. Dodia, and V. Kuppili, "Brain-Computer Interface for wheelchair control operations: An approach based on Fast Fourier Transform and On-Line Sequential Extreme Learning Machine," Clin. Epidemiol. Glob. Heal., vol. 7, no. 3, pp. 274–278, 2019, doi: 10.1016/j.cegh.2018.10.007.

[95] D. R. Edla, M. F. Ansari, N. Chaudhary, and S. Dodia, "Classification of Facial Expressions from EEG signals using Wavelet Packet Transform and SVM for Wheelchair Control Operations," Procedia Comput. Sci., vol. 132, no. Iccids, pp. 1467–1476, 2018, doi: 10.1016/j.procs.2018.05.081.

[96] Witmer, B. G., & Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. In Presence (Vol. 7).

[97] International Telecommunication Union. (1996). Video Quality Assessment Methods for Multimedia Applications. Retrieved from
https://www.itu.int/rec/T-REC-P.910/en.

[98] Tcha-Tokey,K.,Loup-Escande,E.,Christmann,O.,&Richir,S.(2016). A questionnaire to measure the user eXperience in immersive virtual environments. ACM International Conference Proceeding Series.

[99] Usoh, M., Arthur, K., Whitton, M. C., Bastos, R., Steed, A., Slater, M.,& Brooks, F. P. (1999). Walking walking-in-place flying, in virtual environments. Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1999, 359–364.
https://doi.org/10.1145/311535.311589.

[100] Shahid, A., Wilkinson, K., Marcu, S., & Shapiro, C. M. (2012). Karolinska Sleepiness Scale (KSS). In STOP, THAT and One Hundred Other Sleep Scales (pp. 209–210).

[101] B. Mehlig, "Machine learning with neural networks," Machine Learning with Neural Networks, Jan. 2019. doi: 10.1017/9781108860604. [Online]. Available:
http://arxiv.org/abs/1901.05639%20http://dx.doi.org/10.1017 /9781108860604.

[102] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: A review," Data Mining and Knowledge Discovery, vol. 33, pp. 917–963, 4 Jul. 2019, issn: 1384-5810. doi: 10.1007/s 10618-019-00619-1. [Online]. Available: http://link.springer.com/10.1007/s10618-019-00619-1.

[103] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," Dec. 2014. [Online]. Avail- able: https://arxiv.org/abs/1412.3555v1.

[104] A. Gu, K. Goel, and C. Ré, "Efficiently modeling long sequences with struc- tured state spaces," Oct. 2021. [Online]. Available: https://arxiv.org/abs/2111.00396v1.

[105] A. Vaswani, N. Shazeer, N. Parmar, et al., "Attention is all you need,"

CoRR, vol. abs/1706.03762, 2017. arXiv: 1706.03762. [Online]. Available: http://a rxiv.org/abs/1706.03762.

[106] Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Netw 61:85–117.

[107] Li Y, Ma W (2010) Applications of artificial neural networks in financial economics: a survey. In: 2010 International symposium on computational intelligence and design, vol 1. IEEE, pp 211–214

[108] Alto V (2019) Neural networks: parameters, hyperparameters and optimization strategies. Towards Data Science.
https://towardsdatascience.com/neural-networks- parameters-hyperparameters-and-optimization-strategies-3f0842fac0a5.

[109] Sharma S, Sharma S, Athaiya A (2017) Activation functions in neural networks. Towards Data Sci 6:310–316.

[110] Konar J, Khandelwal P, Tripathi R (2020) Comparison of various learning rate scheduling techniques on convolutional neural network. In: 2020 IEEE international students' conference on electrical, electronics and computer science (SCEECS). IEEE, pp 1–5.

[111] Brownlee J (2018) What is the difference between a batch and an epoch in a neural network? Mach Learn Mastery 20:1–5.

[112] Halgamuge MN, Daminda E, Nirmalathas A (2020) Best optimizer selection for predicting bushfire occurrences using deep learning. Nat Hazards 103:845–860.

[113] Sepp Hochreiter and Ju¨rgen Schmidhuber. LSTM can solve hard long time lag problems. Neural Information Processing Systems, pages 473–479, 1997.

[114] Sepp Hochreiter, Martin Heusel, and Klaus Obermayer. Fast model-based protein homology detection without alignment.
Bioinformatics, 23(14):1728–1736, jul 2007.

[115] Felix A. Gers, Ju¨rgen Schmidhuber, and Fred Cummins. Learning to For- get: Continual Prediction with LSTM. Neural Computation, 12(10):2451–2471, oct 2000.

[116] Felix A. Gers, Nicol N. Schraudolph, and Ju¨rgen Schmidhuber. Learning precise timing with LSTM recurrent networks. Journal of Machine Learning Research (JMLR), 3(1):115–143, 2002.

[117] Rumelhart, D., MacClelland, J.L., 1986. Learning internal representations by error backpropagation. Parallel Distributed Processing, vol. 1. Foundations: The MIT Press (chapter 8).

[118] K.S. Narendra, K. Parthasarathy Identification and control of dynamical systems using neural networks IEEE Transactions on Neural Networks, 1 (1) (1990), pp. 4-27.

[119] K.S. Narendra, K. Parthasarathy Gradient methods for the optimisation of dynamical systems containing neural networks IEEE Transactions on Neural Networks, 2 (2) (1991), pp. 252-262.

[120] Weigend, A.S., Rumelhart, D., Huberman, B.A., 1990. Back-propagation, weight-elimination and time series prediction. In: Proceedings of the 1990 Connectionist models Summer School. Morgan Kaufman.

[121] P. Zufiria,An overview of dynamic system control using neural networks A.B. Bulsari (Ed.), Neural Networks for Chemical Engineers, 0-444-82097-3, Elsevier (1995), pp. 385-408.

[122] S. Kotte and J. R. Dabbakuti, "Methods for removal of artifacts from eeg signal: A review", Journal of Physics: Conference Series, vol. 1706, pp. 012093, 2020.

[123] S.J. Luck, An introduction to the event-related potential technique, Cambridge:MIT press, 2005.

[124] C. Dora and P. K. Biswal, "Automated detection of nonphysiological artifacts in polysomnographic eeg using conventional signal processing techniques", TENCON IEEE Region 10 Conference, pp. 1568-1572, 2017.

[125] M. M. Mannan, M. A. KAMRAN and M. Y. Jeong, "Identification and removal of physiological artifacts from electroencephalogram signals: A review", IEEE Access, vol. 6, pp. 30630-30652, 2018.

[126] Islam MK, Rastegarnia A, Yang Z. Methods for artifact detection and removal from scalp EEG: a review. Neurophysiol Clin. 2016;46(4–5):287–305.

[127] Muthukumaraswamy SD. High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. Front Hum Neurosci. 2013;7:138.

[128] Barlow J. Clinical applications of computer analysis of EEG and other neurophysiological signals. Handbook of EEG. Amsterdam: Elsevier; 1986.

[129]Gramfort, A.; Luessi, M.; Larson, E.; Engemann, D.A.; Strohmeier, D.; Brodbeck, C.; Hämäläinen, M. MEG and EEG data analysis with MNE-Python. Front. Neurosci. 2013, 7, 267. [CrossRef] [PubMed]

[130] http://research.ics.aalto.fi/ica/icademo/

[131] welch, P. D. (1967). The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified

Periodograms. IEEE Transactions on Audio and Electroacoustics, 15(2), 70–73. https://doi.org/10.1109/TAU.1967.1161901

[132] Herwig U, Satrapi P, Schonfeldt-Lecuona C. Using the international 10-20 EEG system for positioning of transcranial magnetic stimulation. Brain Topogr. 2003;16(2):95–9

[133] Hu S, Lai YX, Valdes-Sosa PA, Bringas-Vega ML, Yao DZ. How do reference montage and electrodes %20http://dx.doi.org/10.1017 /9781108860604.

[134] B. Mehlig, "Machine learning with neural networks," Machine Learning with Neural Networks, Jan. 2019. doi: 10.1017/9781108860604. [Online]. Available:
http://arxiv.org/abs/1901.05639%20http://dx.doi.org/10.1017 /9781108860604.

[135] A. Nicolae, "Plu: The piecewise linear unit activation function," Sep. 2018. doi: 10.48550/arxiv.1809.09534. [Online]. Available: http://arxiv.org /abs/1809.09534.

[136] Y. Zhou, Z. Zhu, and Z. Zhong, "Learning specialized activation functions with the piecewise linear unit," Apr. 2021. [Online]. Available: http://arxiv .org/abs/2104.03693.

[137] Tu Y, et al. Alpha and gamma oscillation amplitudes synergistically predict the perception of forthcoming nociceptive stimuli. Hum Brain Map. 2016;37:501–14.

[138] Yu, T., & Zhu, H. (2020). HyperParameter Optimization: A Review of Algorithms and Applications. arXiv preprint arXiv:2003.05689.
https://doi.org/10.48550/arXiv.2003.05689

[139] Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '19, August 4–8, 2019, Anchorage, AK, USA. Preferred Networks, Inc., Tokyo, Japan.

[140] Mwangi B, Tian TS, Soares JC. A review of feature reduction techniques in Neuroimaging. Neuroinformatics. 2014;12:229–44.

[141] Tu Y, Hung YS, Hu L, Zhang Z. PCA-SIR: a new nonlinear supervised dimension reduction method with application to pain prediction from EEG. In: 7th International IEEE/EMBS Conference on Neural Engineering (NER). 2015. pp. 1004–1007.

[142] Sadrawi, M.; Sun, W.Z.; Ma, M.M.; Yeh, Y.T.; Abbod, M.; Shieh, J.S. Ensemble Genetic Fuzzy Neuro Model Applied for the Emergency Medical Service

via Unbalanced Data Evaluation. Symmetry 2018, 10, 71.

[143] Guang-Hui, F.; Feng, X.; Bing-Yang, Z.; Lun-Zhao, Y. Stable variable selection of class-imbalanced data with precision-recall criterion. Chemom. Intell. Lab. Syst. 2017, 171, 241–250.

[144] Castro, C.; Vargas-Viveros, E.; Sánchez, A.; Gutiérrez-López, E.; Flores, D. Parkinson's Disease Classification Using Artificial Neural Networks. In Proceedings of the VIII Latin American Conference on Biomedical Engineering and XLII National Conference on Biomedical Engineering, CLAIB 2019, Cancun, Mexico, 2–5 October 2019; Volume 75.

[145] Metrics for Multi-Class Classification:
An Overview. arXiv 2020, arXiv:2008.05756.

[146] Al-Fahoum, A.S.; Al-Fraihat, A.A. Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains. Int. Sch. Res. Not. 2014, 2014, 730218.

[147] Jahidin, A.H.; Ali, M.M.; Taib, M.N.; Tahir, N.M.; Yassin, I.M.; Lias, S. Classification of intelligence quotient via brainwave sub-band power ratio features and artificial neural network. Comput. Methods Programs Biomed. 2014, 114, 50–59.

[148] Singh, P.; Joshi, S.D.; Patney, R.K.; Saha, K. The Fourier decomposition method for nonlinear and non-stationary time series analysis. Proc. R. Soc. A Math. Phys. Eng. Sci. 2017, 473, 20160871.

[149] Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. IEEE Trans. Signal Process. 2013, 62, 531–544.

[150] Peng, C.J.; Chen, Y.C.; Chen, C.C.; Chen, S.J.; Cagneau, B.; Chassagne, L. An EEG-based attentiveness recognition system using Hilbert–Huang transform and support vector machine. J. Med. Biol. Eng. 2020, 40, 230–238.

[151] Zwillinger, D.; Kokoska, S. CRC Standard Probability and Statistics Tables and Formulae; CRC Press: Boca Raton, FL, USA, 1999; Google-Books-ID: TB3RVEZ0UIMC.

[152] Hjorth, Bo; Elema-Schönander, AB (1970). "EEG analysis based on time domain properties". Electroencephalography and Clinical Neurophysiology. 29 (3): 306–310. doi:10.1016/0013-4694(70)90143-4. PMID 4195653.

[153] Boonyakitanont, P., Lek-uthai, A., Chomtho, K., & Songsiri, J. (2019). A review of feature extraction and performance evaluation in epileptic seizure detection using EEG. Biomedical Signal Processing and Control, 101702.
https://doi.org/10.1016/j.bspc.2019.101702

[154] P. Stoica & R. Moses (2005). "Spectral Analysis of Signals"

[155] Welch, P. D. (1967). "The use of Fast Fourier Transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms". IEEE Transactions on Audio and Electroacoustics, AU-15 (2), 70–73. Bibcode:1967ITAE...15...70W. doi:10.1109/TAU.1967.1161901.

[156] Abeele, V.V., Schraepen, B., Huygelier, H., Gillebert, C., Gerling, K., Van Ee, R.: Immersive Virtual Reality for Older Adults. ACM Trans- actions on Accessible Computing 14(3), 1–30 (2021). DOI 10.1145/3470743. URL https://dl.acm.org/doi/abs/10.1145/3470743https: //dl.acm.org/
doi:10.1145/3470743

[157] Barsasella, D., Malwade, S., Chang, C.C., Liu, M., Srikanth, S., Panja, A., Li, Y.C., Syed-Abdul, S.: Opinions regarding Virtual Reality among Older People in Taiwan. In: Proceedings of the 6th International Conference on Information and Communication Technologies for Ageing Well and e-Health, pp. 165–171. SCITEPRESS - Science and Technology Publications (2020). DOI 10.5220/0009425801650171. URL
http://www.scitepress.org/DigitalLibrary/Link.aspx
doi=10.5220/0009425801650171

[158] Weech, S., Kenny, S., Lenizky, M., Barnett- Cowan, M.: Narrative and gaming experience interact to a ect presence and cybersickness in virtual reality. International Journal of Human-Computer Studies 138, 102398 (2020). DOI 10.1016/j.ijhcs.2020.102398. URL https://linkinghub.elsevier.com/retrieve/pii/S1071581920300033.

[159] Harris, D.J., Buckingham, G., Wilson, M.R., Brookes, J., Mushtaq, F., Mon-Williams, M., Vine, S.J.: Exploring sensorimotor performance and user experience within a virtual reality golf putting simulator. Virtual Reality 25(3), 647–654 (2021). DOI 10.1007/s10055-020-00480-4.

[160] Ling, Y., Nefs, H.T., Brinkman, W.P., Qu, C., Heynderickx, I.: The relationship between indi- vidual characteristics and experienced presence. Computers in Human Behavior 29(4), 1519–1530(2013). DOI 10.1016/j.chb.2012.12.010.

[161] Reiter, U., Brunnstro¨m, K., De Moor, K., Larabi, M.C., Pereira, M., Pinheiro, A., You, J., Zgank, A.: Factors Influencing Quality of Experience. In: T-Labs Series in Telecommunication Services, pp. 55–72. Springer Science and Business Media B.V. (2014). DOI 10.1007/978-3-319-02681-7-4

[162] Ruan, J., Xie, D.: Networked vr: State of the art, solutions, and challenges (2021). DOI 10.3390/ electronics10020166.

[163] Santos Silva, R., Mol, A.M., Ishitani, L.: Virtual reality for older users: a systematic literature review. International Journal of Virtual Reality 19(1), 11–25 (2019). DOI 10.20870/ijvr.2019.19. 1.2908. URL https://ijvr.eu/article/view/2908

[164] Saleme, E.B., Covaci, A., Assres, G., Comsa, I.S., Trestian, R., Santos, C.A., Ghinea, G.: The influence of human factors on 360 mulsemedia QoE. International Journal of Human Computer Studies 146, 102550 (2021). DOI 10.1016/j.ijhcs.2020. 102550

[165] Baker, S., Waycott, J., Robertson, E., Carrasco, R., Neves, B.B., Hampson, R., Vetere, F.: Evaluating the use of interactive virtual reality technology with older adults living in residential aged care. Information Processing and Management 57(3), 102105 (2020). DOI 10.1016/j.ipm.2019.102105.

[166] Kent, M., Ellis, K.: Live music in a virtual world: Exuberant flourishing and disability at wheelies nightclub in second life. In: The Digital Evolution of Live Music, pp. 85–98. Elsevier Inc. (2015). DOI 10.1016/B978-0-08-100067-0.00007-5.

[167] Samana, R., Wallach, H.S., Safir, M.P.: The Impact of Personality Traits on the Experience of Presence. In: 2009 Virtual Rehabilitation International Conference, pp. 1–7. IEEE (2009). DOI 10.1109/ICVR.2009.5174197. URL http://ieeexplore.ieee.org/document/5174197/

[168] Kober, S.E., Neuper, C.: Personality and Presence in Virtual Reality: Does Their Relationship Depend on the Used Presence Measure? International Journal of Human-Computer Interaction 29(1), 13–25 (2013).
DOI 10.1080/10447318.2012.668131. URLhttps://www.tandfonline.com/doi/abs/10.1080/10447318.2012.668131.

[169] Rosa, P.J., Morais, D., Gamito, P., Oliveira, J., Saraiva, T.: The Immersive Virtual Reality Experience: A Typology of Users Revealed Through Multiple Correspondence Analysis Combined with Cluster Analysis Technique. Cyberpsychology, Behavior, and Social Networking 19(3), 209–216 (2016). DOI 10.1089/cyber.2015. 0130. URL
https://www.liebertpub.com/doi/ abs/10.1089/cyber.2015.0130.

[170] Bulu, S.T.: Place presence, social presence, copresence, and satisfaction in virtual worlds. Computers and Education 58(1), 154–161 (2012). DOI 10.1016/j.compedu.2011.08.024.

[171] Cho, Y.H., Yim, S.Y., Paik, S.: Physical and social presence in 3D virtual role-play for pre service teachers. Internet and Higher Education 25, 70– 77 (2015).

DOI 10.1016/j.iheduc.2015.01.002.

[172] Galloso, I., Palacios, J.F., Feij´oo, C., Santa- mar´ıa, A.: On the influence of individual characteristics and personality traits on the user experience with multisensorial media: an experimental insight. Multimedia Tools and Applications 75(20), 12365–12408 (2016). DOI 10.1007/s11042-016-3360-z. URL http://link.springer.com/10.1007/s11042-016-3360-z

[173] Whalen, T., Noel, S., Stewart, J.: Measuring the human side of virtual reality. In: IEEE International Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2003. VECIMS '03. 2003, pp. 8– 12. IEEE (2003). DOI 10.1109/VECIMS.2003. 1227022.
URL http://ieeexplore.ieee.org/document/1227022/

[174] Melo, M., Vasconcelos-Raposo, J., Bessa, M.: Presence and cybersickness in immersive content: Effects of content type, exposure time and gender. Computers and Graphics (Pergamon) 71, 159–165 (2018).
DOI 10.1016/j.cag.2017.11.007.

[.175] Slater, M., Usoh, M., Steed, A.: Depth of Presence in Virtual Environments. Presence: Teleoperators and Virtual Environments 3(2), 130– 144 (1994). DOI 10.1162/pres.1994.3.2.130. URL
https://direct.mit.edu/pvar/article/ 3/2/130-144/58820.

[176] Witmer, B.G., Singer, M.J.: Measuring presence in virtual environments: A presence questionnaire. Presence: Teleoperators and Virtual Environments 7(3), 225–240 (1998). DOI 10.1162/105474698565686. URL /record/2016-00465-001.

[177] Frolich,L.,&Dowding,I.(2018).RemovalofmuscularartifactsinEEG signals: a comparison of linear decomposition methods. Brain Informatics, 5(1), 13–22. https://doi.org/10.1007/s40708-017-0074-6

[178] C. R. Harris, K. J. Millman, S. J. van der Walt, et al., "Array programming with NumPy," Nature, vol. 585, no. 7825, pp. 357–362, Sep. 2020. doi: 10.10 38/s41586-020-2649-2. [Online]. Available: https://doi.org/10.1038/s41586-020-2649-2.

[179] W. McKinney, "Data Structures for Statistical Computing in Python," in Proceedings of the 9th Python in Science Conference, S. van der Walt and J. Millman, Eds., 2010, pp. 56–61. doi: 10.25080/Majora-92bf1922-00a.

[180] T. pandas development team, Pandas-dev/pandas: Pandas, version latest, Feb. 2020. doi: 10.5281/zenodo.3509134. [Online].
Available: https://doi.org /10.5281/zenodo.3509134.

[181] A. Paszke, S. Gross, F. Massa, et al., "Pytorch: An imperative style, high-

performance deep learning library," in Advances in Neural Information Processing Systems 32, Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperati ve-style-high-performance-deep-learning-library.pdf.

[182] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

[183] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internalcovariate shift. In International Conference on Machine Learning, pages 448–456, 2015.

[184] Anders Krogh and John A Hertz. A simple weight decay can improve generalization. In Advances in neural information processing systems, pages 950–957, 1992.

[185] Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. Journal of machine learning research, 15(1):1929–1958, 2014.

[186] Tromp, J., Peeters, D., Meyer, A.S. et al. The combined use of virtual reality and EEG to study language processing in naturalistic environments. Behav Res 50, 862–869 (2018). https://doi.org/10.3758/s13428-017-0911-9

[187] https://campaign.ant-neuro.com/unleash-the-power-of-high-density-eeg?source=google&gad_source=1&gclid=CjwKCAiA1-6sBhAoEiwArqlGPgGXDegcqYjqvtEgiDCCwKwh6Y_FHEdSt7dSZXgZJaGbGcw-m1YeohoCxLcQAvD_BwE

7 Appendix

The theory of artificial neural network ideas, including neurons, activation functions, loss functions, and stochastic gradient descent, is thoroughly covered in this chapter.

7.1 Activation Functions

In the realm of artificial neural networks (ANNs), activation functions play a critical role in determining the network's ability to solve complex problems. These functions can be broadly categorized into three types:

- linear
- non-linear
- piecewise linear

The choice of activation function significantly influences the problems an ANN can effectively tackle.

Linear activation functions are suitable for problems that are linearly separable, meaning they can be resolved using a linear approach. However, the scope of such functions is limited as they are inadequate for handling more complex, non-linearly separable issues. This limitation underscores the importance of non-linear activation functions in ANNs.

Non-linear activation functions, such as Tanh (tangent hyperbolic), sigmoid, and softmax, are essential for addressing problems beyond the scope of linear separability. Tanh, for instance, is particularly useful due to its non-linear nature.

These functions enable the network to capture and learn from the non-linear relationships in the data, which is crucial for tasks like image recognition, natural language processing, and more.

Despite their advantages, non-linear activation functions are not without challenges. A significant issue associated with them is the vanishing gradient problem. This problem becomes more pronounced as ANNs increase in depth, i.e., when they have more layers. In deep networks, if the gradient of the activation function in several layers is less than 1, the cumulative effect during backpropagation leads to exceedingly small gradient values. Consequently, the network's ability to learn diminishes, as indicated by the product of these gradients approaching zero. This phenomenon is evident in references [134-135], highlighting the complexities and challenges involved in the design and implementation of effective ANNs. Careful consideration and selection of activation functions are therefore pivotal to the successful operation of these networks.

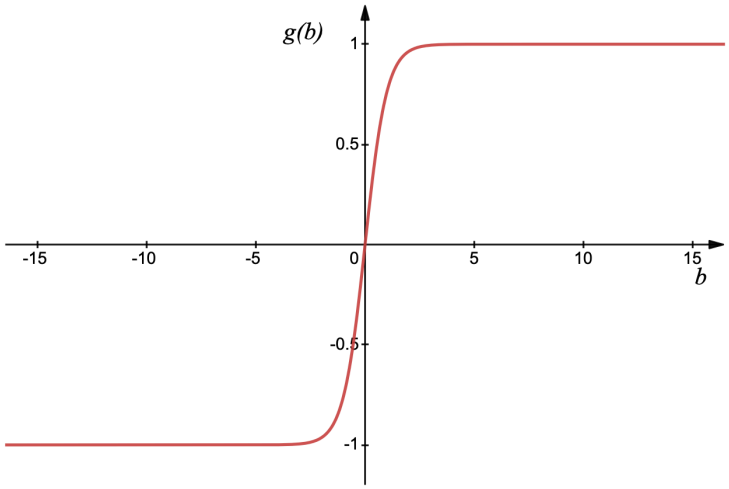


Figure 7.1: A continuous, non-linear activation function. Hyperbolic tangent(g(b) = tanh(b)).

Piecewise linear activation functions are increasingly used to mitigate the vanishing gradient problem in deep neural networks, as discussed in references [135-136]. A prime example is the Rectified Linear Unit (ReLU), defined as g(b)=max(0,b). The key advantage of ReLU is its derivative, which is consistently 1 for inputs greater than zero, preventing the gradient from diminishing to zero during backpropagation in deep networks.

Further advancements include more sophisticated piecewise linear functions like the Piecewise Linear Unit (PLU), which mimics the Tanh function. Developed by

Nicolae at the University of Washington, the PLU offers similar functionality to Tanh but without leading to the vanishing gradient problem, making it a superior option in certain scenarios, as highlighted in reference [.135]. These developments demonstrate significant progress in addressing the challenges posed by deep network training.

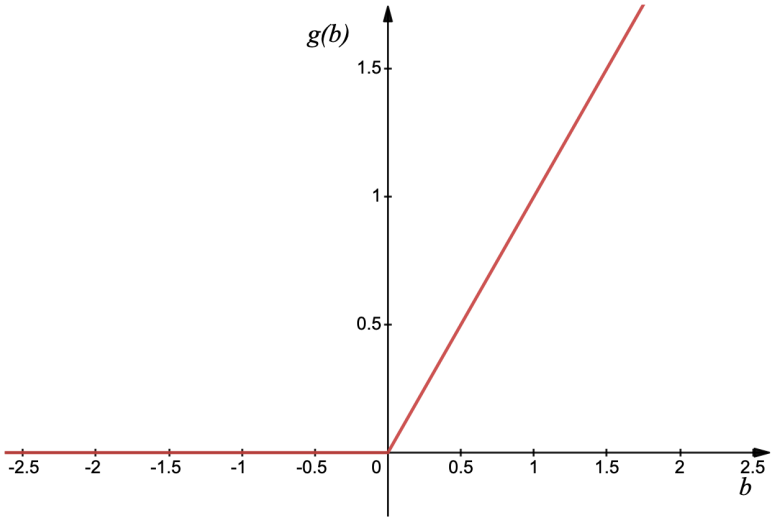


Figure 7.2: An activation function with piecewise linearity: ReLU(max(0, b) = g(b)).

7.2 Loss Functions

For an artificial neural network (ANN) to approximate a function effectively, it must learn from its inaccuracies in classifying inputs. This learning process is fundamentally guided by loss functions, which play a pivotal role in quantifying the deviation of the ANN's predictions from the actual ground truth, as detailed in references [22][58]. The choice of an appropriate loss function is crucial and heavily depends on the nature of the problem being addressed.

When considering the problem type, one must ascertain whether it is a linear regression problem, a binary classification problem, or a multi-class classification problem. This distinction is vital because the nature of the output being predicted differs substantially between regression and classification problems. Regression problems focus on predicting continuous quantities, whereas classification problems involve predicting discrete class labels, a distinction underscored in reference [59].

The formula for a commonly used loss function, Mean Squared Error (MSE), which is often applied in regression problems, is:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

where n is the number of samples, y_i is the actual value, and $\hat{y}_i$ is the predicted value.

In binary classification problems, Binary Cross-Entropy is a frequently used loss function, with the formula:

$$\text{Binary Cross-Entropy} = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right]$$

Here, y_i represents the actual label (0 or 1), and $\hat{y}_i$ is the predicted probability of the class label being 1.

For multi-class classification problems, Categorical Cross-Entropy is often chosen, and its formula is:

$$\text{Categorical Cross-Entropy} = -\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij}\log(\hat{y}_{ij})$$

where m is the number of classes, y_{ij} is the actual probability distribution over the classes, and $\hat{y}_{ij}$ is the predicted probability distribution.

7.3 Fast Fourier Transform

The Fast Fourier Transform (FFT) stands as a cornerstone in signal processing, particularly renowned for its efficiency with stationary signals. By converting signals from the time domain to the frequency domain, FFT facilitates spectral analysis, a critical operation in extracting features by computing the Power Spectral Density (PSD). Non-parametric techniques like Welch's method enhance the PSD estimation, reinforcing FFT's utility in signal analysis [146-147].

Despite FFT's prominence, its application is limited when confronting non-stationary and nonlinear data, such as electroencephalogram (EEG) signals. The inherent characteristics of such signals often render the FFT-derived results less than reliable, spotlighting a significant gap in signal processing methodologies for these types of data [146].

Addressing this gap, researchers have pioneered alternative methods tailored for non-stationary signal analysis. Among these, the Fourier Decomposition Method (FDM) emerges as a significant contribution [148]. Similarly, the Variational Mode Decomposition (VMD) method [.149] and the Hilbert-Huang Transform (HHT) [.150] have been developed. These innovative approaches are specifically designed to provide more reliable analysis of non-stationary signals, offering a substantial leap forward from traditional FFT applications.

This evolution of methodologies underscores a commitment to precision and adaptability in signal processing, ensuring that the analysis of The Fast Fourier Transform (FFT) is an established method for the analysis of stationary signals, adept at transforming signals from the time domain to the frequency domain to enable spectral analysis [146]. Feature extraction within FFT involves the calculation of the Power Spectral Density (PSD) using mathematical tools, with non-parametric approaches like Welch's method often employed for enhanced PSD estimation [146-147].

However, the efficacy of FFT is constrained when applied to nonlinear and nonstationary data, such as EEG signals, where the results may lack reliability. This limitation has prompted the development of alternative analytic techniques specifically tailored for such data. Notably, the Fourier Decomposition Method (FDM) 5.2 [148], Variational Mode Decomposition (VMD) [149], and Hilbert-Huang Transform (HHT) [150] are among the novel procedures that have been introduced.

These advanced methods signify a pivotal shift in signal analysis, particularly in handling the complexities of non-stationary signals. They offer a more accurate and reliable framework for the analysis of EEG signals, reflecting the dynamic progress in the domain of signal processing. signals like EEG is both accurate and trustworthy, guided by the latest advancements in the field.

$$F(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-j\omega t}\, dt \tag{7.2}$$

7.4 ANOVA

The Analysis of Variance (ANOVA) plays a pivotal role in experimental data analysis, particularly when assessing results from experiments that involve testing various parameters [55]. This statistical technique is instrumental in our work, employing one-way ANOVA to examine hypotheses concerning the means of different classes. One-way ANOVA tests explicitly the null hypothesis that posits no significant difference in means across the classes. The rejection of this hypothesis is contingent on a confidence level α, indicating a $1 - \alpha$ probability that the means are not identical. In other words, the risk of erroneously discarding this hypothesis is represented by the value of α.

A crucial aspect of ANOVA is its reliance on the assumption that the data is usually distributed. This assumption, however, is only sometimes valid, especially in cases involving EEG data. Despite this, the utility of ANOVA is still somewhat maintained. A high confidence level (low α value) in the ANOVA test still provides strong evidence against the null hypothesis, suggesting that the means of the groups are not all the same.

Moreover, ANOVA can be employed as an effective method for feature selection. In this context, it functions as a filter method, where features are individually assessed using the ANOVA test. Features where the means significantly differ (as indicated by the ANOVA test) are selected, as these are likely to be more discriminative.

When dealing with experiments involving only two classes, a T-test is generally sufficient to determine if there is a significant difference in the means of the groups. However, one-way ANOVA extends this concept, making it applicable to scenarios with more than two groups. It is a generalization of the T-test, and hence, understanding the principles of a T-test provides foundational knowledge for comprehending one-way ANOVA.

An additional component integral to this analysis is the p-value, a statistical measure that helps determine the significance of the results obtained from ANOVA. The p-value indicates the likelihood of keeping the results if the null hypothesis were true. A small p-value (typically less than 0.05, corresponding to a 95% confidence level) suggests strong evidence against the null hypothesis, implying a statistically significant difference in means across the groups. Thus, in the context of feature selection, features that yield a low p-value in the ANOVA test are considered significant and are more likely to be selected for further analysis.

In summary, with its ability to test the equality of means across multiple groups and its applicability as a filter method in feature selection, one-way ANOVA is a robust statistical tool. Its effectiveness, combined with the interpretative power of the p-value, makes it highly valuable in experimental analysis, particularly in areas where normal distribution cannot be assumed, such as EEG data analysis.

7.5 Deep Learning Software and Library:Tourch

Machine learning techniques are supported extensively by Torch, a scientific library, computational framework, and scripting language built on the Lua programming language. Torch gives priority to the GPU. Several research laboratories are using Torch, including those at NYU, IDIAP, Google, Twitter, and Purdue. Designed for embedded systems, Lua is a programming language. Neural network graphs may be efficiently created in Torch and then parallelized across CPUs and GPUs. It makes it possible to build a deep network with several layer stacks in a sequential fashion. For research on reinforcement learning, Torch is the primary suggested platform. The latest machine-learning methods are widely used in Torch.