From Traditional BCIs to Real-World Applications: Toward Noise-Resilient Brain-Computer Interfaces

Muhammad Ahsan Awais, B.Eng., M.Sc.

A dissertation submitted in fulfilment of the requirements for the award of Doctor of Philosophy (Ph.D.)

to the

Dublin City University

School of Computing

Supervised by

Asst. Prof. Graham Healy

Prof. Tomas Ward

January 2026

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy (Ph.D.) is entirely my own work, and that I have exercised reasonable care to ensure that the work is original and have conformed to the regulations on the use and declaration of Generative AI, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work. I hereby certify that no Generative Artificial Intelligence (Gen AI) tools have been used in the creation of the thesis.

Signed: *Muhammad Ahsan Awais*

ID No.: 22263261

Date: 13-January-2026

Dedication

This thesis is dedicated to:

The loving memory of my father, who would have been proud to see me achieve this milestone. His absence is deeply felt, but his values and dreams continue to guide and inspire me.

The cherished memory of my mother-in-law, who was always there to guide and support me. She took great joy in my growth and success, and her encouragement and kindness continue to inspire me.

My mother, whose endless prayers, unconditional love, and unwavering support have been the foundation of all my achievements. Her belief in me gave me strength when I needed it most.

My wife, for her boundless patience, encouragement, and love. Her quiet sacrifices and constant support have been my anchor throughout this journey.

My twin baby boys, whose smiles bring light to my darkest days and joy to every moment. They are my greatest motivation and the reason I strive to be better every single day.

Acknowledgements

"In the name of ALLAH, the Most Gracious, the Most Merciful."

Alhamdulillah, all praises be to ALLAH for His infinite mercy, strength, and blessings, which have guided me throughout this journey. It is through His will that I have been able to acquire knowledge, overcome challenges, and complete this PhD.

I would like to express my deepest gratitude to all those who have supported me throughout this PhD journey. I am immensely thankful to my supervisors, Dr. Graham Healy and Prof. Tomas Ward, for their unwavering support, patience, and kindness. Their valuable time, prompt feedback, and expert guidance have been instrumental in shaping this research. I am truly fortunate to have worked under their mentorship during this challenging yet rewarding endeavor.

I also extend my sincere thanks to the Insight Research Ireland Centre for Data Analytics for providing a supportive research environment. I gratefully acknowledge the financial support received from Science Foundation Ireland (now Research Ireland) and CHIST-ERA, under grant numbers SFI/12/RC/2289_P2 and CHISTERA IV 2020 - PCI2021-122058-2A.

My appreciation also goes to the Faculty of Engineering and Computing at Dublin City University for giving me the opportunity to work as a tutor and demonstrator across various undergraduate courses. These teaching experiences enriched my journey and strengthened my academic skills.

I am thankful to everyone at DCU who helped me, whether directly or indirectly, especially my friends and colleagues who stood by me in moments of need. I am particularly grateful to those with whom I shared countless tea and coffee breaks,

those moments of laughter and conversation helped ease the stress and made daily life more enjoyable.

Finally, I wish to express my heartfelt appreciation to my beloved family. To my mother, whose constant prayers have always been a source of strength, and to the cherished memory of my late father and late mother-in-law, who would have been immensely proud and happy to see this milestone achieved. To my wife, whose unwavering support, patience, and love have carried me through every challenge; and to my two lovely sons, whose smiles and presence keep me motivated, ambitious, and full of life. Without their love, sacrifices, and belief in me, this thesis would not have been possible.

List of Publications

Journal Publications

- Published

    1. **Muhammad Ahsan Awais**, Tomas Emmanuel Ward, Peter Redmond, and Graham Healy. "From lab to life: assessing the impact of real-world interactions on the operation of rapid serial visual presentation-based brain-computer interfaces." Journal of Neural Engineering, 21, no. 4, 046011, 2024.
    2. **Muhammad Ahsan Awais**, Peter Redmond, Tomas Emmanuel Ward, and Graham Healy. "AMBER: advancing multimodal brain-computer interfaces for enhanced robustness—A dataset for naturalistic settings." Frontiers in Neuroergonomics, 4, no. 1, 1216440, 2023.

- In preparation

    1. **Muhammad Ahsan Awais**, Tomas Emmanuel Ward, and Graham Healy. "Deploying EEG-Based BCIs Beyond the Lab: Fine-Tuned Subject-Independent Classification Under Real-World Conditions." IEEE Transactions on Biomedical Engineering, 2025.
    2. **Muhammad Ahsan Awais**, Tomas Emmanuel Ward, and Graham Healy. "BCIs Beyond the Lab: Practical Deployment Factors and Performance Trade-offs." Engineering Applications of Artificial Intelligence, 2025.

3. Deirdre Byrne, **Muhammad Ahsan Awais**, Ankit Vijayvargiya, Sonal Santosh Baberwal, Ronan Langan, and Tomas Emmanuel Ward. "An EEG dataset of upper and lower limb motor Imagery in those with Spinal Cord Injury." Scientific Data, 2025.

Conference Publications

- Papers (in press)

    1. **Muhammad Ahsan Awais**, Tomas Emmanuel Ward, and Graham Healy. "Enhancing Subject-Independent P300 Classification in RSVP-Based BCIs with Deep Learning". 36th Irish Signals and Systems Conference, Donegal, Ireland, 2025.
    2. **Muhammad Ahsan Awais**, Tomas Emmanuel Ward, and Graham Healy. "AMBER 2.0: A Dataset for Naturalistic Settings with HMD-Based RSVP Tasks". 7th International Conference on Smart Computing and Informatics, Kuala Lumpur, Malaysia, 2025.

- Abstracts (published)

    1. **Muhammad Ahsan Awais**, Tomas Emmanuel Ward, and Graham Healy. "Comparative Analysis of P300 Detection Accuracy in Traditional and Head Mounted Display Environments". 30th Annual Conference of the RAMI Section of Bioengineering, Athlone, Ireland, 2025.
    2. Marc Welter, **Muhammad Ahsan Awais**, Tomas Emmanuel Ward, and Fabien Lotte. "Challenges towards automated art presentation brain-computer interfaces". 10th Visual Science of Art Conference, Aberdeen, Scotland, 2024.
    3. Peter Redmond, **Muhammad Ahsan Awais**, and Tomas Emmanuel Ward. "Ethical and GDPR-Compliant Sharing of Multimodal Stress Data: Contributions to Environmental and Health Research". Joint Conference of ISEH ICEPH & ISEG, Galway, Ireland, 2024.

4. **Muhammad Ahsan Awais**, Tomas Emmanuel Ward, and Graham Healy. "Investigating the Impact of Ecologically Valid Interactions on Rapid Serial Visual Presentation-based Brain-Computer Interface Performance". 10th International BCI Meeting, Brussels, Belgium, 2023.

5. Peter Redmond, **Muhammad Ahsan Awais**, and Tomas Emmanuel Ward. "Automatic Tagging of BCI Artefacts Using Computer Vision". 10th International BCI Meeting, Brussels, Belgium, 2023.

Contents

List of Figures

List of Tables

List of Abbreviations

AD	Alzheimer's Disease
ADHD	Attention Deficit Hyperactivity Disorder
AEP	Auditory Evoked Potentials
ALS	Amyotrophic Lateral Sclerosis
AR	Augmented Reality
Avg	Average
BCI	Brain-computer interface
BLR	Bayesian Linear Regression
BMI	Brain-Machine Interface
CAR	Common Average Reference
CM	Classification Model
CNN	Convolutional Neural Networks
CSP	Common spatial pattern
ECoG	Electrocorticography
EDF	European Data Format
EEG	Electroencephalography
ERD	Event-Related Desynchronisation
ERP	Event-Related Potentials
fMRI	Functional Magnetic Resonance Imaging
fNIRS	Functional Near-Infrared Spectroscopy
FT	Fine Tuning
HMD	Head Mounted Display
ICA	Independent Component Analysis
IC-RSVP	Intentionally Contaminated - Rapid Serial Visual Presentation

IIR	Infinite Impulse Response
LDA	Linear Discriminant Analysis
LOSO	Leave-One-Subject-Out
MB-MoE	Multiband Mixture-of-Experts
MEG	Magnetoencephalography
ML	Machine Learning
MoE	Mixture-of-Experts
PCA	Principal Component Analysis
RNNs	Recurrent Neural Network
ROC-AUC	Receiver Operating Characteristic-Area Under the Curve
RQ	Research Question
RSVP	Rapid Serial Visual Presentation
SD	Standard Deviation
SNR	Signal-to-Noise Ratio
SPS	Samples per Second
Ss1	Session 1
Ss12	Session 1 and Session 2
Ss123	Session 1, Session 2 and Session 3
SSAEP	Steady-State Auditory Evoked Potentials
SSEP	Steady-State Evoked Potentials
SSSEP	Steady-State Somatosensory Evoked Potentials
SSVEP	Steady-State Visual Evoked Potentials
SVM	Support Vector Machines
Trials Dr	Trials Dropped
Trials Rt	Trials Retained
VEP	Visual Evoked Potentials
VR	Virtual Reality

Abstract

Muhammad Ahsan Awais

From Traditional BCIs to Real-World Applications: Toward Noise-Resilient Brain-Computer Interfaces

Brain–Computer Interfaces (BCIs) offer promising communication pathways between the human brain and external devices, yet their deployment in real-world settings remains limited by noise susceptibility, subject variability, and practical constraints. This thesis addresses these challenges using Rapid Serial Visual Presentation (RSVP)-based P300 paradigms, with a focus on enhancing robustness, generalisability, and real-world applicability.

To begin, I systematically investigated the impact of real-world behavioral artifacts, body movement, head movement, and talking, on EEG signal quality and classification performance. Results reveal that such noise substantially degrades system accuracy and often leads to significant data loss during artifact rejection, underscoring the limitations of traditional denoising methods and motivating adaptive solutions.

Next, I explored subject-independent classification using various deep learning models. Transformer-based architectures, especially when combined with EEG-specific convolutional neural networks (i.e., EEGNet), demonstrate superior generalisation under the Leave-One-Subject-Out (LOSO) framework. In particular, the hybrid Transformer with a multiband input strategy consistently outperformed other architectures, highlighting the effectiveness of integrating temporal attention mechanisms with frequency-specific spatial filtering. This underscores robust generalisation across subjects, a crucial step toward the practical and scalable deployment of BCIs.

While subject-independent models offer strong baseline performance, their real-world adaptability is further enhanced through lightweight subject-specific calibra-

tion, where incorporating a small amount of personalised data yields noticeable performance gains. To reduce system complexity without compromising accuracy, I identified a 16-channel EEG configuration that retains over 95% of baseline performance, with 10–12 channel setups also viable for certain applications. However, channel selection remains inherently subject- and task-dependent, necessitating application-specific validation. Additionally, I evaluated the impact of display modality, finding that both head-mounted displays (HMDs) and traditional monitors support comparable classification performance. With their added portability, HMDs represent a promising avenue for mobile and ecologically valid BCI applications.

Practical recommendations are offered on model choices, data volume, channel optimisation, and calibration, aiming to bridge the gap between lab-based BCI research and real-world deployment. While the work primarily focuses on RSVP-based P300 detection, the insights contribute broadly to the development of more noise-tolerant and effective BCI systems.

Chapter 1

Introduction

Brain-computer interfaces (BCIs) are systems that enable direct communication between the brain and external devices by translating neural activity into actionable commands without the need for peripheral muscle control [1], [2]. BCIs hold immense promise across a wide range of applications, particularly for individuals with motor impairments who cannot rely on traditional forms of communication or interaction. By leveraging electrical brain signals, BCIs allow users to control assistive technologies, communicate through spellers, manipulate robotic arms, or interact with virtual environments.

Among various neuroimaging methods used for BCI development, Electroencephalography (EEG) [3] has emerged as one of the most practical and widely used modalities. EEG measures the brain's electrical activity through electrodes placed on the scalp and offers several advantages: it is non-invasive, cost-effective, portable, and provides high temporal resolution, making it well-suited for real-time applications. As a result, EEG-based BCIs have found increasing utility in domains such as communication aids for locked-in patients, neurorehabilitation, gaming, cognitive state monitoring, and more.

Despite their potential, EEG-based BCI systems face critical challenges, especially when deployed outside controlled laboratory environments. EEG signals are inherently low in amplitude and highly susceptible to noise and artifacts. Common sources of contamination include ocular movements (e.g., blinking, saccades),

muscular activity (e.g., jaw clenching, speaking), head and body movements, and external electrical interference. These noise sources significantly degrade signal quality, compromise model accuracy, and reduce the overall usability of BCI systems [4].

This challenge is especially relevant in applications such as virtual art galleries [5], [6], where BCIs are used to assess user engagement or attentional interest during natural, unconstrained exploration in immersive VR environments [7], [8]. In such contexts, artifacts arising from free head movement, eye movement behavior, and spontaneous reactions are not only expected but inevitable, underscoring the need for noise-resilient BCI solutions.

In traditional lab-based BCI studies, participants are typically asked to remain still, and the environment is carefully controlled to minimise interference. However, such constraints are unrealistic in real-world settings where users are expected to move, speak, or interact with their environment naturally. This discrepancy highlights a major limitation of many BCI systems: their lack of robustness to non-ideal conditions. Developing BCIs that can operate reliably under these conditions is essential for transitioning BCI technologies from research labs into everyday use.

To explore these challenges, this thesis adopts the Rapid Serial Visual Presentation (RSVP) P300 paradigm, a widely used task in BCI research that is both scientifically well-characterised and practical for controlled and naturalistic conditions. In RSVP, a rapid sequence of visual stimuli (e.g., images) is presented on the screen, among which infrequent target items are embedded, eliciting the P300 event-related potential when detected by the participant [9]. By utilising the RSVP-based P300 task across both clean and noisy conditions, the aim is to differentiate between the EEG responses captured in ideal lab settings from those affected by real-world artifacts. This controlled variation enables a systematic investigation of how different types of noise, such as head movement, speech, or body movement, impact P300 detection and classification accuracy.

The RSVP-P300 paradigm is particularly well-suited for this investigation because it offers a reliable ground truth and provides accuracy as a straightforward

performance metric [10]. This allows us to assess not only the robustness of various classification models but also the effectiveness of signal denoising techniques. Through this framework, I gain insights into the relationship between behavioral artifacts and EEG signal degradation, which is critical for building noise-resilient BCI systems.

A general diagram illustrating how the RSVP experiment was employed for EEG recordings under both noisy and noise-free conditions is provided below.

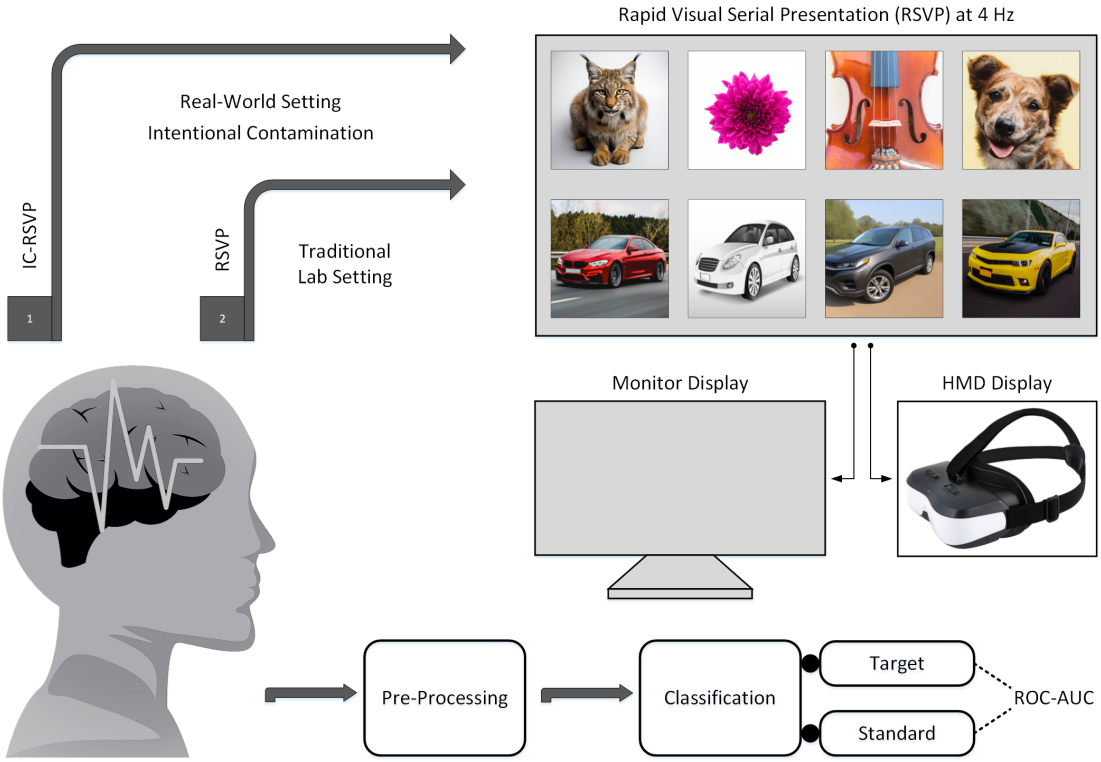


Figure 1.1: Schematic diagram of the experimental process.

This thesis is motivated by the need to bridge the gap between high-performing BCI systems developed in ideal lab conditions and their effective deployment in real-world environments. Specifically, it investigates noise-resilient EEG-based BCIs through the lens of single-trial classification, subject-independent generalisation, minimal calibration requirements, and robustness to behavioral and environmental noise. The findings have implications for improving the usability and reliability of BCI systems in naturalistic contexts, where artifacts are unavoidable but performance must remain reliable.

1.1 Hypothesis

This thesis hypothesises that real-world behavioural artifacts, such as body movement, head movement, and talking, introduce systematic disruptions in EEG signals that substantially reduce the performance of P300-based RSVP BCI systems. Models trained solely on clean laboratory data are expected to show limited generalisability under such noisy conditions, whereas incorporating noisy data and advanced deep learning architectures should improve robustness in subject-independent scenarios. The work further assumes that addressing practical considerations for real-world deployment will be essential to achieving resilient and efficient BCI systems.

1.2 Research Questions

(**RQ.1**) How do specific types of intentional behavioral artifacts (body movement, head movement, talking) impact EEG signal measurement and the performance of a traditional RSVP-P300 BCI system?

(**RQ.2**) How well do BCI models trained on clean versus artifact-contaminated EEG data generalise to real-world conditions, and can training on noisy data yield comparable performance when tested on similarly noisy inputs?

(**RQ.3**) Which advanced machine learning models are most effective in achieving subject-independent P300 detection robust to real-world noise?

(**RQ.4**) What are the key practical refinements and design considerations needed to transition BCI systems from lab-controlled environments to effective real-world applications?

- (**RQ.4a**) What is the impact of the amount of training data from different subjects on the performance of subject-independent BCI models?
- (**RQ.4b**) How much subject-specific calibration data is required to effectively adapt subject-independent models for improved performance in the presence of noise?

(**RQ.4c)** What is the optimal number and combination of EEG channels required to maintain high classification accuracy while minimising complexity in real-world BCI applications?

(**RQ.4d)** How does the use of a head-mounted display (HMD) for RSVP presentation, compared to a traditional monitor, affect the EEG signals and BCI performance in both clean and noisy conditions?

1.3 Objectives

1. To systematically investigate the impact of common real-world behavioral artifacts on RSVP-P300 BCI performance.

2. To evaluate the effectiveness of different machine learning approaches for subject-independent and noise-resilient P300 detection.

3. To explore the potential of using head-mounted displays to create more ecologically valid and mobile BCI paradigms.

4. To determine the feasibility and data requirements for subject-specific adaptation (fine-tuning) of generalisable BCI models in noisy environments.

5. To identify the minimum number and optimal combinations of EEG channels that maintain robust classification performance in RSVP-based P300 BCIs, particularly under subject-independent settings and real-world noise conditions.

6. To provide insights into the design and development of BCIs that are more suitable for real-world applications.

1.4 Contributions

1. A systematic protocol was developed to collect EEG data under varied environmental conditions and display modalities (AMBER and AMBER 2.0),

enabling controlled comparisons of laboratory and real-world settings.

2. Noise was found to significantly degrade EEG signal quality, with head movement causing the strongest disruptions among the tested artifact conditions.

3. Across a wide range of models, a hybrid Multiband EEGNet-Transformer achieved the best performance in both clean and noisy conditions, outperforming traditional and baseline deep learning approaches.

4. Session-based fine-tuning demonstrated that using as little as two sessions of subject-specific data notably improved performance, enhancing real-world adaptability of subject-independent models.

5. A direct relationship between training data volume and performance was revealed, with larger datasets and greater subject diversity improving generalisation across individuals.

6. A reduced EEG channel set of 16 electrodes preserved about 95% of performance compared to the full 32-channel configuration, supporting the development of lightweight, wearable BCI systems.

7. The use of head-mounted displays showed no significant differences from monitor-based presentation, indicating their potential as a practical and portable solution for real-world BCI applications.

1.5 Thesis Structure

The remainder of this thesis is organised as follows:

Chapter 2 – Background and Related Work: This chapter introduces the foundational concepts of Brain-Computer Interfaces, including the P300 paradigm and the Rapid Serial Visual Presentation protocol. It then reviews related work in the areas of P300 detection, machine learning approaches for BCI, and the

transition from controlled to real-world BCI systems, highlighting the need for noise-resilient and generalisable models.

Chapter 3 – Systematic EEG Data Collection for Real-World BCIs: Protocol Design and Acquisition: This chapter presents the design of the data acquisition protocol and the development of two datasets, AMBER and AMBER 2.0, capturing EEG signals under both clean and intentionally contaminated conditions using different display modalities.

Chapter 4 – Impact of Noise on P300 Detection: This chapter investigates the influence of specific behavioral artifacts (e.g., body movement, head movement, and speech) on EEG signal quality and classification performance. It also evaluates model generalisability across clean and noisy conditions.

Chapter 5 – Subject-Independent Benchmarking: This chapter explores subject-independent classification approaches, benchmarking a range of models including traditional machine learning algorithms, CNNs, transformers, and hybrid architectures under noisy conditions.

Chapter 7 – Conclusion and Future Work: This final chapter summarises the research contributions, reflects on the findings, and outlines potential directions for future work toward developing robust, adaptive, and deployable BCI systems.

Chapter 2

Background and Related Work

This chapter provides an overview of the existing literature and key research developments in the field of brain-computer interfaces (BCIs), with a particular focus on the transition from traditional laboratory-based settings to real-world applications. It covers relevant topics such as the P300 event-related potential, RSVP paradigms, the impact of artifacts and noise in EEG signals, traditional and deep learning classification methods, and recent advances in robust and adaptive models for BCI applications. Special emphasis is placed on subject-independent classification, data volume requirements, and calibration techniques, all of which are pivotal to enabling BCIs that operate reliably in non-ideal settings.

The remainder of this chapter is organized as follows: first, I provide an overview of BCIs, their applications, and the main paradigms, highlighting current trends and challenges. I then discuss the P300 event-related potential, followed by an examination of RSVP paradigms. Next, I review machine learning and deep learning approaches for RSVP-P300 classification. Finally, I address the challenges of moving beyond laboratory-based evaluations and emphasize the need for developing BCIs that can perform reliably in real-world environments.

2.1 Brain-Computer Interfaces: An Overview

Brain-Computer Interfaces, often also referred to as Brain-Machine Interfaces (BMIs), represent an emerging technology that establishes a direct communication link between the brain and an external device, bypassing the body's typical efferent motor pathways [1], [2]. This allows for the translation of neural activity directly into commands for controlling computers, prosthetic limbs, or other technologies [11], [12].

The significance of BCIs is far-reaching, impacting several key scientific and technological domains. In neuroscience, BCIs serve as invaluable tools for probing the complexities of brain function, investigating neural encoding mechanisms, and studying principles of neuroplasticity. Beyond fundamental research, BCIs play a significant role in cognitive and neurological diagnostics. For instance, clinicians are exploring BCI applications to aid in the diagnosis and monitoring of conditions such as Parkinson's disease, ADHD (Attention deficit hyperactivity disorder), Alzheimer's disease, and traumatic brain injuries by analyzing distinct neural signatures associated with these disorders [13]–[17]. These applications also extend to monitoring cognitive load and fatigue, offering objective measures that can inform therapy planning and rehabilitation strategies [18]–[21]. For assistive technology, BCIs offer transformative possibilities for individuals with severe neuromuscular disabilities, such as those resulting from amyotrophic lateral sclerosis (ALS), spinal cord injury, stroke, or cerebral palsy [1], [22]. By enabling alternative means of communication and control over assistive devices, BCIs can substantially improve autonomy and enhance the quality of life for these individuals [23]–[26].

2.1.1 Background and History of BCIs

The conceptual roots of BCIs can be traced back to early investigations into the electrical activity of the brain. A seminal moment was the invention of electroencephalography by Hans Berger in the 1920s, which demonstrated the ability to record electrical oscillations from the human scalp [27], [28]. This breakthrough provided

the idea that brain signals could potentially be used for purposes beyond clinical diagnosis.

While the notion of a direct brain-machine link had been considered, the term "Brain-Computer Interface" was formally introduced by Jacques Vidal in the 1970s [29]. Vidal's pioneering work at UCLA (University of California, Los Angeles) explored the modulation of slow cortical potentials and visually evoked potentials as a direct means for users to interact with a computer [30].

The 1980s represented a pivotal period in the advancement of BCI research, characterised by foundational experiments conducted primarily in animal models. During this time, researchers demonstrated that neural signals recorded from single neurons or small populations within the motor cortex could be harnessed to control external devices. Notably, several studies successfully trained monkeys to control the movement of a computer cursor through the intentional modulation of their neural activity [31], [32]. These findings offered strong empirical support for the feasibility of translating brain signals into functional control commands.

Initially, BCI research primarily focused on invasive techniques that required surgical procedures to record neural activity directly from the brain. However, during the 1980s and 1990s, there was a significant shift toward non-invasive approaches, particularly EEG, to capture brain signals from the scalp. This transition was driven by the desire to develop safer, more accessible, and user-friendly BCI systems suitable for broader clinical and non-clinical applications [33].

The late 20th and early 21st centuries saw the translation of BCI research to human participants, particularly in the development of assistive communication and control systems for individuals with severe motor impairments, showcasing the practical utility of both invasive and non-invasive approaches [33], [34]. The continuous progress in signal acquisition hardware, processing algorithms, and machine learning techniques continues to drive the capabilities of both invasive and non-invasive BCI technologies [35]–[37]

Today, BCI research spans a diverse array of applications, ranging from clinical

rehabilitation and assistive technologies for individuals with motor impairments [38], [39] to cutting-edge developments in gaming [40], virtual and augmented reality [41], mental workload monitoring [42], and smart home control [43]. As the field evolves, researchers continue to focus on enhancing the accuracy, usability, and real-time responsiveness of BCI systems, while also addressing challenges related to noise robustness, portability, and user adaptability, key factors for successful deployment in everyday environments.

2.1.2 Applications of BCIs

Clinical Applications

Over the past two decades, BCI research and development has seen vigorous growth in clinical applications, primarily focusing on restoring lost motor and communication functions, and aiding in the diagnosis and monitoring of neurological conditions. BCIs offer significant hope to individuals with severe movement disorders, including those with Amyotrophic Lateral Sclerosis (ALS), spinal cord injury, stroke, cerebral palsy, and other significant neuromuscular conditions [1], [13].

A major focus within clinical BCIs is the restoration of motor control. Researchers are actively developing systems that enable individuals with paralysis to control external devices such as robotic arms for reaching and grasping tasks or manipulate computer cursors using decoded neural signals [38], [39], [44]. BCIs are being explored for controlling prosthetic limbs, including lower limb prosthetics, allowing users to regain mobility through thought control [45], [46]. Restoring locomotion is another critical area, with significant effort directed towards developing BCI-actuated wheelchairs and exoskeletons controlled by various BCI paradigms such as MI, P300 event-related potentials, Steady-State Visual Evoked Potentials (SSVEP), and hybrid approaches combining multiple signal types [46]–[50].

Beyond motor restoration, EEG plays an increasingly important role in the diagnosis and monitoring of neurological and cognitive disorders. By analyzing specific brain signal patterns, EEG can provide objective measures related to cognitive

function and neurological state. For example, the P300 event-related potential and serves as a biomarker that reflects cognitive processes like attention and decision-making [51]. Deviations in P300 characteristics can offer insights into conditions affecting cognitive processing. Furthermore, analyses of EEG and other neurophysiological signals are being investigated as diagnostic and monitoring tools for a range of neurological and psychiatric conditions, including Parkinson's disease, ADHD, Alzheimer's disease, epilepsy, and the assessment of consciousness in patients with traumatic brain injuries and disorders of consciousness [15], [52]–[54]. Functional neuroimaging techniques like functional Magnetic Resonance Imaging (fMRI) and functional Near-Infrared Spectroscopy (fNIRS), which measure changes in blood oxygenation as an indicator of neural activity, are also employed in BCI research for diagnosis and rehabilitation monitoring, providing insights into brain activation patterns associated with different states and tasks [55], [56].

BCIs are also being integrated into neurorehabilitation protocols to promote neural plasticity and functional recovery after stroke or spinal cord injury. By providing real-time feedback on brain activity related to attempted movements, BCIs can help patients retrain brain circuits and improve motor function [57]. Neurofeedback, a form of BCI, has also shown promise in alleviating symptoms of anxiety and depression by training individuals to self-regulate specific brain activity patterns [56], [58]–[60].

Non-Clinical Applications

While clinical applications remain a primary driver for BCI development, the technology is increasingly extending into non-clinical domains, offering novel ways to interact with technology and enhance human capabilities for the general population. This diversification is driven by advancements in non-invasive sensing technologies and more sophisticated signal processing.

One of the most pertinent non-clinical application areas is entertainment and gaming. The gaming industry offers a fertile ground for BCI experimentation, en-

abling immersive interaction through direct neural control. Studies have shown that games can act not only as engaging environments for experimentation but also as effective tools for neurofeedback training, cognitive rehabilitation, and stress reduction [40], [61]. Researchers have successfully integrated BCIs with popular game genres, such as virtual shooters, puzzle games, and sports simulations, to create experiences where users control game play through motor imagery or attention modulation [62]–[64].

BCIs are also being integrated into Virtual Reality (VR) and Augmented Reality (AR) environments to enhance user interaction and presence, enabling control or adapting the virtual experience based on neural signals [41], [65].

Smart home integration is another promising frontier. BCI-driven home automation allows users, particularly individuals with motor impairments, to control lights, appliances, or other connected devices using thought alone. These interfaces provide an intuitive and accessible alternative to traditional input methods and could play a pivotal role in the design of more inclusive smart environments [43], [66].

Beyond consumer applications, BCIs are being investigated in neuromarketing, where researchers analyze neural responses to advertisements and branding elements to better understand consumer attention, engagement, and emotional responses [67], [68]. Although this application area raises ethical concerns, it demonstrates the broader potential of EEG signals in interpreting internal cognitive and emotional states in naturalistic contexts.

Security and authentication represent another emerging use case. EEG-based biometric systems explore brainwave patterns as unique identifiers, offering an alternative to traditional authentication mechanisms such as fingerprints or facial recognition. Brainprint authentication methods may offer higher resistance to spoofing and improved privacy, though challenges around consistency and environmental variability remain [69], [70].

Current trends in non-clinical BCI research include exploring applications in demanding operational environments such as driving and aviation. Monitoring a

driver's cognitive state, such as drowsiness or distraction, using passive BCIs can enhance safety systems [42], [71]. In aviation, BCIs are being investigated for pilot state monitoring, workload assessment, and potentially as an additional control modality or communication channel in complex cockpit environments [72], [73]. These applications highlight the growing interest in using BCIs to augment human performance and improve safety in critical tasks.

As non-clinical applications continue to grow, they expand the role of BCIs beyond assistive technologies into tools for augmenting human interaction, decision-making, and everyday convenience. This shift reflects a broader trend in neurotechnology: the transition from laboratory experiments to real-world, user-centered systems that enhance both capability and quality of life.

2.1.3 Types of BCIs

Brain-Computer Interfaces can be categorized based on several criteria, reflecting different aspects of their design, operation, and how they interact with the user and the brain. Understanding these classifications is crucial for comprehending the diverse landscape of BCI research and applications.

Based on User Intention

This categorisation focuses on how the user engages with the BCI system, whether through deliberate effort or spontaneous brain activity.

1. **Active BCIs:** Active BCIs are systems where the user consciously and intentionally generates brain activity to control an external device. Examples include motor imagery (MI) paradigms, where users imagine limb movements to navigate a cursor or operate assistive devices [74].

2. **Reactive BCIs:** Reactive BCIs rely on brain responses that are elicited by external stimuli. The system detects event-related potentials (ERPs) such as the P300 or Steady-State Visual Evoked Potentials (SSVEPs) in response to

specific sensory inputs. Users do not need to initiate control voluntarily but instead react to presented stimuli by intentionally attending to them, making reactive BCIs highly suitable for spelling systems or stimulus-driven interfaces [75].

3. **Passive BCIs:** Unlike active and reactive systems, passive BCIs monitor spontaneous brain activity without requiring any intentional effort from the user. These systems are often used for assessing mental states such as fatigue, emotional engagement, or workload [74].

Based on System Interaction

This categorisation relates to the timing and control of the interaction between the user and the BCI system.

1. **Synchronous BCIs:** Synchronous BCIs operate in a time-locked fashion, where the system prompts the user to issue commands at predefined intervals. The system prompts the user with a cue, and the user is expected to perform a specific mental task or attend to a stimulus within a designated time frame [76].

2. **Asynchronous BCIs:** Asynchronous or self-paced BCIs allow users to issue commands at any time, providing a more natural and flexible interaction paradigm. These systems pose greater challenges for signal detection due to the need to distinguish intentional commands from resting brain activity but are essential for real-world BCI deployment [76].

Based on Invasiveness

The final categorisation is based on how the neural signals are acquired from the brain, distinguishing BCIs by their level of invasiveness.

1. **Invasive BCIs:** At the inception of BCI research, systems were primarily invasive in nature, involving direct implantation of electrodes into the brain

tissue, typically in the cortex. These systems provide high spatial and temporal resolution and have been used in clinical trials to restore motor functions in individuals with severe paralysis. However, they require neurosurgery and carry significant medical risks [77].

2. **Semi-invasive BCIs:** As technology advanced, a middle ground was established in the form of semi-invasive BCIs. These systems, such as electrocorticography (ECoG), place electrodes on the surface of the brain without penetrating the cortex. These provide a balance between signal quality and surgical risk, offering better resolution than non-invasive methods while avoiding penetration into brain tissue [77].

3. **Non-invasive BCIs:** With growing interest in safety, accessibility, and real-world applications, the field has predominantly shifted toward non-invasive BCIs. Unlike invasive methods that marked the early beginnings of BCI research, the non-invasive story began much later, gaining momentum in the 1990s as technological advances enabled practical and reliable brain signal acquisition without surgical intervention [78]–[80]. These systems record brain activity without penetrating the skull, using modalities such as electroencephalography. Although they offer lower signal quality compared to invasive methods, they are more accessible and suitable for a broad range of applications.

2.1.4 Neuroimaging Modalities Used in BCIs

The performance and potential applications of a Brain-Computer Interface are fundamentally dependent on the neuroimaging modality used to acquire brain signals. Each modality offers different trade-offs in terms of spatial resolution, temporal resolution, invasiveness, and portability. A comparative overview of these modalities is presented in Table 2.1, followed by a detailed description of each technique below.

1. **Electroencephalography (EEG):** EEG is a widely used technique for record-

Table 2.1: Comparison of different neuroimaging modalities [81]–[83]

Method	Invasiveness	Spatial Resolution	Temporal Resolution	Portability	Signal Type
EEG	Non-invasive	10 mm	0.05 s	Portable	Electrical
MEG	Non-invasive	5 mm	0.05 s	Non-portable	Magnetic
fMRI	Non-invasive	1 mm	1 s	Non-portable	Metabolic
fNIRS	Non-invasive	5 mm	1 s	Portable	Metabolic
ECoG	Semi-Invasive	1 mm	0.003 s	Portable	Electrical
Intracortical	Invasive	0.5 mm	0.003 s	Portable	Electrical

ing the brain's electrical activity through the placement of electrodes on the scalp surface [3]. These signals arise from the electrical interactions among neurons within the brain. EEG-based BCIs detect specific frequency patterns by measuring small voltage fluctuations generated during various mental activities. Typically, EEG signals occupy a frequency range of up to 50 Hz and are characterised by high temporal resolution, making them particularly suitable for real-time applications [82].

EEG is a non-invasive, relatively low-cost, and portable neuroimaging technique that offers high temporal resolution and is easy to set up for basic applications. However, it suffers from poor spatial resolution and is highly susceptible to artifacts from muscle movements and environmental noise, with signals often attenuated and distorted by the skull and scalp [33].

2. **Magnetoencephalography (MEG):** MEG measures the magnetic fields generated by neuronal activity in the brain. It offers excellent temporal resolution comparable to EEG, but with improved spatial localisation due to reduced distortion from the skull and scalp. MEG is often used in clinical research and neuroscience but is less common in everyday BCI systems due to its high cost, immobility, and the need for magnetically shielded rooms.

   MEG-based BCIs have shown promise in detecting cognitive processes and have been used to decode motor intentions and language processing. However, its utility in practical BCI applications remains limited due to logistical

constraints [84].

3. **Functional Near-Infrared Spectroscopy (fNIRS):** fNIRS is a non-invasive optical imaging technique that measures changes in oxygenated and deoxygenated hemoglobin concentrations in the cortex, reflecting underlying neural activity. Although fNIRS has lower temporal resolution compared to EEG or MEG, it offers decent spatial resolution and is relatively tolerant to motion artifacts.

   fNIRS-based BCIs are gaining attention for portable and wearable applications, especially in mental workload assessment and affective computing, and are often combined with EEG in hybrid systems to enhance performance [85].

4. **Functional Magnetic Resonance Imaging (fMRI):** fMRI detects brain activity by measuring changes in blood oxygenation (BOLD signals) associated with neural activation. fMRI offers exceptional spatial resolution, making it valuable for identifying specific brain regions involved in various cognitive tasks.

   However, fMRI's extremely low temporal resolution, high operational cost, and large, immobile equipment significantly limit its use in real-time or portable BCIs. Most fMRI-based BCI studies remain confined to research environments focused on understanding brain function rather than deploying practical systems [86].

5. **Electrocorticography (ECoG):** ECoG involves placing electrodes directly on the cortical surface, offering a balance between the high spatial resolution of fMRI and the high temporal resolution of EEG. Being semi-invasive, ECoG provides clearer signals with less contamination from skull or scalp artifacts.

   ECoG is predominantly used in clinical contexts, especially for patients undergoing epilepsy surgery, where electrodes are already implanted for monitoring purposes. ECoG-based BCIs have demonstrated impressive accuracy in decoding motor and language intentions, and are an active area of research in

both invasive and hybrid BCI systems [87].

6. **Intracortical:** Intracortical BCIs represent the most invasive neuroimaging approach currently used in BCI systems. These systems involve implanting microelectrode arrays directly into the cortical tissue, allowing for the recording of action potentials from individual neurons or small populations of neurons. This method offers the highest spatial and temporal resolution among all BCI modalities, making it highly effective for decoding fine motor intentions [88].

   Despite their precision, intracortical BCIs require invasive neurosurgery and carry risks such as infection, inflammation, and long-term signal degradation, which limits their widespread clinical use.

2.1.5 EEG-Based BCIs

The discovery of EEG laid the groundwork for the eventual development of brain-computer interface systems [28], [34], [57]. The origins of EEG can be traced back to the 19th century, when English physician Richard Caton first recorded electrical activity from the brains of animals and published his findings in 1875 in the British Medical Journal [89]. Around the same period, Polish scientists Napoleon Nikodem Cybulski and Adolf Beck were also conducting pioneering work on bioelectrical brain signal recordings, contributing significantly to the early foundations of neurophysiology [90], [91].

One of the primary strengths of EEG is its non-invasive nature, allowing safe and repeatable use across diverse populations. Its high temporal resolution makes it ideal for applications requiring real-time feedback, such as neurofeedback, assistive communication, and gaming. EEG systems are often lightweight and portable, making them suitable for out-of-laboratory environments and practical for long-term or mobile use.

Despite these advantages, EEG suffers from several limitations. The most critical among them is the low signal-to-noise ratio (SNR), as EEG signals are often contaminated by physiological artifacts (e.g., eye blinks, muscle movements) and en-

Table 2.2: EEG Frequency bands and their description [33], [82], [93]

Band	**Frequency Range**	**Brain State**	**Functional Significance**
Delta	0.5 – 4 Hz	Deep sleep, unconsciousness	Associated with deep, dreamless sleep and unconscious states. It reflects a lack of cognitive engagement, typically observed when the brain is at its lowest level of arousal.
Theta	4 – 8 Hz	Drowsiness, meditation	Linked to deep relaxation, meditative states, and inward-focused mental activity. It is often associated with creativity, intuition, daydreaming, and access to subconscious material such as imagery and fantasy.
Alpha	8 – 13 Hz	Relaxed, eyes closed	Represents a relaxed, wakeful state characterized by calmness and a lack of active cognitive processing. It is prominent during eyes-closed rest and is related to a balanced mood, enhanced self-awareness, and readiness to learn new information.
Beta	13 – 30 Hz	Active concentration	Associated with active mental engagement, attention, and problem-solving. It reflects states of alertness, concentration, and sometimes stress or agitation during high cognitive load.
Gamma	>30 Hz	High-level cognition	Linked to high-level cognitive functions such as memory, perception, sensory processing, and consciousness. It is thought to play a role in complex tasks involving learning, language, and information integration

vironmental electrical noise [92]. EEG also has limited spatial resolution due to the skull's filtering effect, making it difficult to localize neural sources precisely. These factors can reduce the reliability and accuracy of EEG-based BCIs, particularly in uncontrolled environments.

EEG signals can be decomposed into different frequency bands, each associated with specific cognitive or physiological states. The most commonly studied bands include delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (> 30 Hz) rhythms. These rhythms are implicated in various brain functions, such as attention, arousal, memory, etc [33], [82], [93].

Due to its millisecond-level time resolution, EEG is particularly well-suited for real-time BCI applications. It has been successfully deployed in assistive technologies such as spellers, wheelchair control, and robotic arm manipulation. In addition, EEG is widely used in cognitive workload monitoring, gaming, VR/AR integration, and neurorehabilitation. Advances in wireless EEG systems and real-time signal processing have significantly improved its practicality for continuous, user-friendly

BCI systems beyond the lab.

2.1.6 BCI Paradigms

The design and functionality of Brain-Computer Interfaces depend heavily on the paradigm used to induce meaningful patterns that can be extracted from brain activity. These paradigms entail the mental strategies that users employ to generate distinguishable neural signals. Some of the most widely adopted paradigms in EEG-based BCIs include evoked potentials, event-related potentials, motor imagery, and hybrid BCIs [94].

Evoked potentials (EPs)

Evoked Potentials are neural responses elicited by external stimuli and recorded as time-locked signals using neuroimaging tools like EEG. They are classified based on the sensory modality used to stimulate the brain:

1. **Visual Evoked Potentials (VEPs)** are generated in response to visual stimuli and reflect activity in the visual cortex (occipital lobe). VEP-based BCIs aim to detect the user's visual attention by analyzing brain responses to flashing or flickering stimuli, often used in gaze-based spellers or target detection tasks [95].

2. **Auditory Evoked Potentials (AEPs)** result from auditory inputs such as tones, speech, or clicks. These small electrical signals originate from the auditory cortex and are useful in scenarios where visual input is limited or inaccessible [96].

3. **Steady-State Evoked Potentials (SSEPs)** are elicited by repetitive, periodic stimuli at a fixed frequency [97]. The most commonly used form is the Steady-State Visual Evoked Potential (SSVEP), triggered by flickering visual stimuli. SSVEP responses appear at the stimulus frequency and its harmonics, typically in the occipital region [98]. Other types include Steady-State

Auditory Evoked Potentials (SSAEPs) [99] and Steady-State Somatosensory Evoked Potentials (SSSEPs) [100], though these are less common in practical BCI systems.

Event-Related Potentials (ERPs)

ERPs are brain responses time-locked to specific sensory, cognitive, or motor events. Unlike EPs, ERPs can reflect higher-order cognitive processing and are typically categorized by latency and polarity:

1. **Exogenous ERPs** (e.g., N100, P100) occur closer to 200ms of stimulus onset and reflect early sensory processing [101].

2. **Endogenous ERPs** (e.g., P300, N400) appear later and are associated with cognitive processes like decision-making or semantic evaluation [101].

Figure 2.1 illustrates the typical morphology of event-related potentials, including prominent components such as N100, P200, P300, and N400, commonly observed in EEG recordings during cognitive tasks

Motor Imagery (MI)

Motor Imagery refers to the mental simulation of movement without actual execution. MI-based BCIs detect changes in brain activity (particularly over the sensorimotor cortex) when a user imagines performing actions such as moving a limb [102].

The main neural signatures are Event-Related Desynchronisation (ERD) and Event-Related Synchronisation (ERS), which correspond to decreases and increases in power within the 8–30 Hz frequency band, respectively [103]. These patterns mirror those produced during real movements and are typically lateralized: imagined movement of the right hand activates the left hemisphere motor cortex, and vice versa. MI is considered a robust and active mental strategy for continuous control in BCIs.

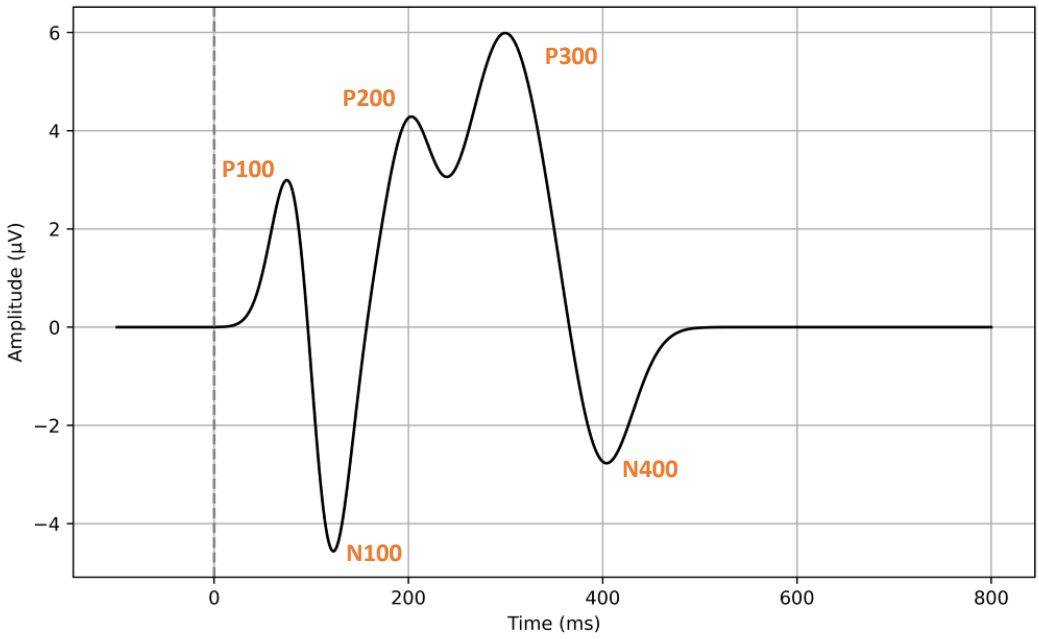


Figure 2.1: ERP components in an EEG signal

Hybrid

Hybrid BCIs integrate two or more types of neural signals or combine different BCI paradigms [104]. These systems aim to improve classification accuracy, information transfer rate, and robustness by leveraging the strengths of multiple paradigms.

For instance, a hybrid system might use SSVEP for target selection and P300 for confirmation, or combine motor imagery with a P300-based speller to enhance communication [105]. Hybrid systems can operate in parallel or sequentially, with one component acting as a trigger or filter for the other, such as a "brain switch" that activates only when a certain brain pattern is detected.

2.1.7 Current Trends and Challenges in BCI Research

A long-standing challenge in BCI research is the issue of subject variability. Traditional BCI models often rely on subject-specific training data, which hinders their scalability and generalisation across users. Subject-independent BCIs aim to overcome this limitation by creating models that generalise well across different indi-

viduals without requiring extensive per-user calibration. However, variability in neurophysiological signals between individuals, due to anatomy, cognitive states, and recording conditions, continues to pose a challenge for universal models [106].

Recent advances in deep learning have enabled the development of sophisticated neural decoding models capable of extracting complex patterns from brain signals. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformer-based architectures have been applied to EEG, ECoG, and fMRI data to decode motor intentions, cognitive states, and even visual imagery [107], [108]. These approaches outperform traditional machine learning in many tasks, yet they require large datasets, are computationally intensive, and often lack interpretability.

Translating BCI systems from controlled lab conditions to real-world applications introduces new challenges due to the presence of various artifacts (e.g., muscle movements, eye blinks, and environmental noise) [109].

Extensive user-specific calibration has traditionally been required before deploying BCI systems. This process is time-consuming and can be frustrating for end-users. So, creating plug-and-play BCIs that require little or no calibration without compromising accuracy is essential for mainstream adoption [110].

As BCIs become more integrated into healthcare, entertainment, and communication technologies, ethical concerns surrounding privacy, user consent, cognitive liberty, and neurosecurity are gaining prominence. The ability to record, interpret, and potentially manipulate neural activity raises profound questions about autonomy and data protection [111].

2.2 The P300 Event-Related Potentials

One of the most well-known and extensively studied ERPs is the P300 ERP, named for its positive polarity and typical latency of around 300 milliseconds post-stimulus [112]. The P300 component is associated with a wide range of cognitive processes, where its distinct characteristics have made it a crucial response for studying attention, memory, and decision-making. The P300 ERP is typically elicited when an

individual detects a rare, unexpected, or salient event within a sequence of standard events [113]. It is most prominent over the parietal cortex and is commonly used in speller systems and target detection tasks [51].

Applications of P300 systems

P300-based brain-computer interfaces offer significant opportunities in both non-clinical applications [114], such as gaming, adaptive learning systems, and user authentication, and clinical applications [115], [116], particularly in assistive communication for individuals with severe motor disabilities.

For instance, P300 spellers allow individuals with locked-in syndrome, such as those affected by amyotrophic lateral sclerosis (ALS), to communicate by selecting letters or words through brain activity alone [117]. This non-invasive technology offers a critical bridge to independence and enhances the quality of life for patients who have lost motor function.

Deceptive information processing in the brain elicits a P300 response, which can be detected using the Concealed Information Test (CIT) [118], [119]. More recent work has shown that analyzing brain connectivity between brain regions improves classification accuracy, distinguishing liars from the innocents [120], [121].

P300-based BCIs have shown potential for smart home control, enabling users to interact with devices through brain signals alone. Several recent studies have demonstrated successful implementations, such as controlling lights, appliances, and other home systems using P300 responses in response to visual stimuli [122].

P300-based BCIs can assist individuals with amyotrophic lateral sclerosis in browsing the internet by enabling them to select web links using brain signals. This allows users to navigate webpages and access content such as news articles without any physical movement [123].

P300-based BCIs provide a creative and entertaining medium, enabling users to paint using their brain signals, enhancing engagement and playful interaction [124]. Similar work has been reported under the application known as “Brain Painting”

[125].

P300 BCIs have been used to develop simple games like chess [126], MindGame [127], Bacteria Hunt [128], etc. The games use brain responses, such as stronger P300 ERPs, to control actions like movement, image manipulation, or targeting. Since they require no prior training, such games can help users become familiar with BCI systems and can also be used to study attention and cognitive engagement [129].

P300-based brain-computer interfaces have become a valuable tool in cognitive and neurological diagnostics due to their ability to capture neural correlates of higher-order brain functions. Clinicians leverage P300 responses to assess and monitor various neurological and psychiatric disorders, providing objective measures that complement behavioral assessments. For example, Altered P300 responses are biomarkers for neurological disorders such as Alzheimer's [130] and schizophrenia [131], making them valuable for early diagnosis and monitoring cognitive states.

Post-stroke patients with aphasia often struggle to communicate due to impaired motor pathways. A P300 BCI paradigm has been used to provide an alternative communication channel, helping activate language circuits and potentially accelerating post-stroke recovery [132], [133].

Abnormalities in P300 amplitude and latency have been linked to Parkinson's disease, where diminished P300 responses may reflect dopaminergic dysfunction and attentional deficits [15]. In attention-deficit/hyperactivity disorder (ADHD), altered P300 characteristics help differentiate subtypes and evaluate treatment efficacy [16]. Similarly, in Alzheimer's disease, P300 alterations correlate with cognitive decline, offering a potential biomarker for early detection and disease progression [17]. Furthermore, traumatic brain injury patients often exhibit disrupted P300 patterns, aiding in the assessment of cognitive impairment severity and recovery trajectories.

Beyond diagnostics and assistive communication, P300 BCIs are increasingly used in cognitive rehabilitation and neurofeedback therapy [134]. By monitoring real-time P300 responses, clinicians can evaluate cognitive load, mental fatigue, and attentional engagement, enabling personalized therapeutic interventions. These ap-

plications are particularly relevant in stroke rehabilitation, where P300-based feedback can enhance neural plasticity, and in occupational settings, where mental workload assessment improves safety and performance optimisation [18]. As research advances, P300 BCIs continue to expand their role in both clinical and research settings, offering non-invasive, objective insights into brain function and dysfunction.

2.3 Rapid Serial Visual Presentation (RSVP)

The Rapid Serial Visual Presentation is an approach to BCI in which a series of images is displayed at high speed. Participants are asked to differentiate between a set of target images and a set of standard images, where P300 ERP is evoked by a target image but not by standard images [9], [10]. BCI signal processing algorithms are then used to recognize spatio-temporal electrophysiological responses and link them to target image identification, ideally on a single-trial basis [135].

RSVP-based BCIs are typically deployed using two modes of visual stimulus presentation: static and moving [136], [137]. Both modes can be used with or without a button press. In some studies, users manually respond to detected targets using a button press to assess baseline performance, measure reaction times, or enhance classification accuracy.

Static Mode

In static RSVP mode, images are briefly displayed, typically for 100–500 ms, before disappearing. Each image enters and exits the screen at the same location, occupying most of the display area, which enhances visibility and recognition. Several task types have been explored in the literature. In one common approach, participants are shown a specific target image prior to the RSVP stream and are instructed to detect its appearance during the sequence. High recognition rates have been reported even at rapid presentation rates of up to 10 images per second [138]. In another variation, participants are asked to identify broader categories of targets (e.g., animals) among non-target images. This type of task generally requires slower

presentation rates, around 4 images per second, to allow sufficient processing time for accurate detection [139].

Moving Mode

In moving RSVP mode, participants view short video clips and are tasked with identifying one or more target events within each clip. Since targets span over time, it is important to ensure sufficient temporal separation between them to avoid ERP suppression effects such as diminished P300 amplitude. The literature presents various task formats. In simpler scenarios, participants are asked to detect predefined visual targets, such as a "person" or "vehicle", within dynamic scenes [140]. More complex tasks involve recognizing behavioral patterns that unfold over time, such as identifying a person placing a suspicious object in a public area. These tasks require the integration of both motion and form cues across spatial and temporal dimensions [141].

RSVP based BCIs

The RSVP paradigm is widely adopted in BCI systems due to several key advantages. One of its primary strengths is the high information transfer rate, made possible by the rapid presentation of visual stimuli, which facilitates quick decision-making. Additionally, RSVP-based BCIs benefit from robust ERP signatures, particularly the P300 component, which can be consistently detected with minimal subject-specific calibration. Unlike motor imagery paradigms that require extensive training and user adaptation, RSVP-based BCIs demand little to no prior training, making them especially suitable for users with limited experience or cognitive impairments [142].

RSVP-based BCIs have been widely employed to detect and recognize specific targets such as objects, scenes, people, and relevant information within static images and videos. This paradigm holds significant potential for domains like counterintelligence, law enforcement, and healthcare, where professionals are required to process

and analyze vast amounts of visual information on a regular basis. While human observers are superior to computers in understanding complex imagery, manual analysis is often time-consuming and inefficient [143], [144].

A growing body of research [142] demonstrates the effectiveness of RSVP-based BCIs in identifying target items across various image types. These systems have shown promise in supporting visual search tasks and, in some cases, leveraging learned visual recognition skills. Furthermore, researchers such as Huang et al. [145] have explored whether combining RSVP-based BCIs with behavioral responses could lead to greater efficiencies, suggesting a possible direction for enhancing performance in real-world applications.

Mental fatigue is a critical factor influencing performance in RSVP-based target detection tasks. Zhou et al. [146] examined the impact of mental fatigue on RSVP-based small target detection using a 75-minute multi-stage experiment. Their findings showed a significant decline in P300 amplitude at the Pz electrode across sessions, suggesting it as a reliable indicator of fatigue. This highlights the importance of considering cognitive fatigue in sustained RSVP-BCI applications. Blanco-Díaz et al. [147] investigated how cognitive and physical factors, specifically concentration level, eye fatigue, and coffee consumption, affect P300-based BCI performance. Using a public dataset and comparing multiple detection methods, they found that P300 detection accuracy improved by 3–6% under high concentration and after coffee intake, highlighting the need to account for user state in optimising BCI performance.

A study by Ricardo et al. [148] explored non-gaze dependent RSVP-based BCI spellers, comparing three types of visual stimuli: white letters (WL), famous faces (FF), and neutral pictures (NP). Their results demonstrated that FF and NP stimuli yielded better classification performance and user preference than traditional letter-based stimuli, suggesting that such visually rich alternatives may enhance communication efficiency, particularly for users with impaired eye-motor control.

Riccio et al. [149] examined how attentional and memory processes affect P300-

based BCI control in individuals with ALS. Participants completed RSVP and change detection tasks to assess temporal and spatial filtering, along with a P300 BCI spelling task. The results showed that only temporal filtering capacity significantly predicted BCI accuracy and P300 amplitude, highlighting the critical role of sustained attentional filtering in successful BCI use.

Cognitive biometrics is a developing research topic that enables user authentication and identification by exploiting the mental states of an individual. In this context, Gupta et al. [150] explored the potential of RSVP-based P300 BCIs for cognitive biometrics, introducing the concept of a "BrainWord" that could serve as a secure authentication method. They investigated how variations in stimulus relevance and spatial presentation affect performance, comparing RSVP and a Spatially Varying Paradigm across eight participants. Their findings indicate that RSVP offers reliable classification accuracy and bit rates, making it a promising approach for EEG-based biometric identification.

Matthew et al. [151] developed a novel unified deep learning framework for hybrid RSVP-SSVEP BCIs, specifically optimised for gaming applications. Implemented in a Bejeweled-inspired interface, the system demonstrates how single-model processing can maintain classification accuracy while significantly improving user engagement and system simplicity. The approach marks an important advancement in making responsive, intuitive BCI controls for extended gaming sessions, bridging the gap between laboratory prototypes and consumer-ready brain-controlled gaming technology.

BCI systems struggle with real-world robustness due to insufficient data on individual neural differences and intra-individual variability. Rexwinkle et al. [152] explored using a game with a purpose (GWAP) as an engaging platform for large-scale, longitudinal BCI data collection. Pilot results demonstrate that the game-embedded tasks elicit comparable neural responses to traditional BCI paradigms while maintaining participant engagement through reward mechanisms. This gamified approach shows promise for addressing critical gaps in BCI development by

enabling the collection of more comprehensive datasets that capture both inter- and intra-individual neural variability, potentially improving system robustness and reducing BCI illiteracy.

Recent research has explored the application of RSVP in optimising memory retention and cognitive processing, particularly among young adults. Studies emphasize the critical role of interval timing and word count in maximizing memory effectiveness, with findings suggesting that an optimal balance between presentation speed and information load enhances recall performance [153]. These findings support tailored RSVP protocols for memory training and adaptive learning systems, offering practical applications in education and cognitive enhancement.

Galvin et al. [154] explored RSVP-BCI-based neurofeedback for cognitive training in mild Alzheimer's disease (AD). Results indicated that participants with mild AD could consistently complete multiple weekly assessments, showed reliable baseline performance, and successfully learned to use a BCI spelling system with training. The findings support BCI's potential as a cognitive intervention tool for AD and highlight important considerations for future research. The work expands BCI applications into neurodegenerative disease management.

Cognitive fatigue during prolonged RSVP tasks reduces target detection accuracy. Recent work by Zhou et al. [155] uses fNIRS-measured prefrontal HbO levels and reaction times to objectively monitor fatigue, showing increased HbO and slower, lognormally-distributed responses as fatigue progresses. These biomarkers could enable adaptive RSVP systems that adjust presentation rates dynamically, particularly for demanding applications like medical image review.

The growing demand for efficient breast cancer screening has driven interest in RSVP-based BCIs, which leverage rapid image presentation and EEG-recorded neural responses to accelerate detection. Hope et al. [156] demonstrates the feasibility of applying cortically-coupled computer vision to mammography, with studies exploring both single-electrode and multi-electrode configurations to optimise detection sensitivity. These developments highlight RSVP's potential as a high-throughput

adjunct to traditional screening methods, particularly for processing large medical imaging datasets.

Recent advances in RSVP-based BCIs have expanded communication options for people with locked-in syndrome, and Fried-Oken et al. [157] developed the first standardized protocol to assess sensory, cognitive, and motor prerequisites for RSVP Keyboard use, combining adapted standardized tests (e.g., auditory comprehension, working memory) with novel visual attention tasks. These brief ($<$ 1 hour), multidisciplinary evaluations test essential cognitive, sensory, and communication skills through adapted tasks using eye gaze responses. Initial validation shows high accuracy (92-100%) in identifying appropriate BCI candidates, establishing a crucial framework for clinical implementation, while highlighting the need for accessible communication technologies in rehabilitation.

Despite the extensive research and diverse applications explored with RSVP-based BCIs, their evaluation in real-world, noisy environments remains limited. Most existing studies are confined to controlled laboratory settings, overlooking the practical challenges posed by everyday behavioral artifacts and environmental variability. This gap highlights a critical limitation in current literature and underscores the need for further research to understand, adapt, and optimise RSVP paradigms for deployment in naturalistic, real-world scenarios.

2.4 Machine Learning and Deep Learning for RSVP-P300 Classification

Classic classification algorithms such as Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), and Bayesian Ridge Regression have been extensively used in RSVP-P300-based BCI studies.

For instance, Sheng et al. [158] in their study on RSVP-based P300 spellers for patients with oculomotor dysfunction (e.g., ALS), employed Linear Discriminant Analysis to classify target symbols presented rapidly (88 ms on, 22 ms off)

without requiring eye movements. The paradigm achieved 90% offline accuracy in identifying targets among 26 symbols, though performance varied (up to 68%) with increasing participant sample size. The results demonstrated LDA's effectiveness in decoding RSVP-P300 responses, highlighting its potential for BCIs in eye-movement-impaired populations. Cui et al. [159] proposed an ERP feature enhancement framework combining latency detection and EEG reconstruction to address variability in RSVP-based P300 classification. Using LDA-based methods, their approach improved AUC, accuracy, and robustness with fewer training samples/channels, outperforming conventional RSVP classification frameworks.

Zhang et al. [160] introduced an RSVP-EEG benchmark dataset for target detection (human/non-human images at 10 Hz) and evaluated four different classifiers, including SVM, SWFP (spatially weighted Fisher LDA-PCA), Discriminative Canonical Pattern Matching (DCPM), and Hierarchical Discriminant Component Analysis (HDCA), providing a standardized comparison for RSVP-BCI algorithms.

Mijani et al. [161] introduced Dual/Triple shifted RSVP paradigms, where 2-3 characters appeared simultaneously (displaying targets 2-3 times vs. once in single RSVP). Using LDA and SVM, they tested all three paradigms with three subjects, finding the Dual paradigm outperformed the single and Triple RSVP paradigms in balancing speed and reliability.

Xiao et al. [162] evaluated Bayesian LDA alongside other classifiers (standard LDA, stepwise LDA, shrinkage LDA, STDA, and DCPM) for single-trial ERP detection in RSVP-based BCIs using an N200/P300 dataset. The study demonstrated Bayesian LDA's utility as part of a broader comparison of advanced LDA variants for noisy, small-sample ERP classification. Rahman et al. [163] in another study employed Bayesian methods alongside other machine learning approaches to predict drivers' reaction times to road events using pre-event EEG spectral features (analyzing 2-second windows before events). The Bayesian framework helped classify drivers as slow/fast responders while enabling subject-independent reaction time prediction in a driving simulator environment, demonstrating its utility for EEG-

based driver safety systems.

Wang et al . [9] introduced a novel Multiple Time Window LDA Beamformer (MTWLB) spatial filtering method and compared nine classification pipelines for RSVP-based BCI. Among three classifiers tested (LDA, Bayesian Linear Regression (BLR), and Logistic Regression), BLR demonstrated robust performance when combined with effective spatial filters (MTWLB or xDAWN), though the proposed MTWLB method generally enhanced classification most significantly (as measured by AUC). The results showed that while spatial filtering improved detection, CSP underperformed compared to other methods, and Bayesian approaches remained competitive when paired with proper feature enhancement. Another study [164] investigated single-trial ERP detection in RSVP tasks using a Bayesian discriminant analysis approach combined with spatial filtering to maximize signal-to-noise ratio, demonstrating Bayesian effectiveness for robust classification in challenging RSVP conditions.

These algorithms typically require hand-crafted features such as signal amplitude, latency, or time-frequency features. Although these models are computationally lightweight, they often fall short when dealing with high inter-subject and inter-condition variability, especially in the presence of noise.

Deep learning has transformed the landscape of EEG classification. Convolutional Neural Networks, recurrent neural networks, and other CNN-based optimised algorithms such as the EEGNet [165] and CNN1 [166] architectures, have demonstrated the capacity to learn spatial and temporal features from raw EEG signals. These models outperform traditional methods when sufficient training data is available and are more resilient to moderate noise. Their ability to process raw signals without manual feature engineering makes them particularly valuable for real-world EEG applications.

For instance, Uma et al. [167] employed Convolutional Neural Networks to classify P300 signals in a gaze-independent RSVP speller system, achieving 97% accuracy for target character prediction. Their CNN approach outperformed tradi-

tional methods (LDA, SVM) by effectively extracting high-level features from raw EEG data, demonstrating particular utility for paralyzed patients lacking oculomotor control.

Zang et al. [168] proposed a phase-locked CNN architecture specifically designed for single-trial RSVP EEG classification, addressing a critical gap in existing models that inadequately exploit ERP's phase-locked characteristics. By integrating standard convolutional, permute, and depthwise convolutional layers, the model separately processes spatial features across distinct ERP time windows, achieving superior classification performance compared to conventional deep learning and traditional methods.

Yuan et al. [169] proposed PSAEEGNet, an optimised CNN integrating pyramid squeeze attention (PSA) modules with deep convolutional layers to address the low SNR and small sample sizes of single-trial EEG in RSVP tasks. By hierarchically refining spatiotemporal P300 features, the model achieved a superior classification score over existing methods, demonstrating its potential for robust EEG-based target recognition. In another study, Wang et al. [170] developed a cascade network combining EEGNet with xDAWN to create a unified P300 detection system for both speller and RSVP paradigms. This hybrid architecture effectively handled variable P300 responses across different stimulus types while reducing the required repetition rounds. The approach demonstrated superior performance in cross-paradigm applications, showcasing EEGNet's adaptability when integrated with xDAWN preprocessing for robust P300 classification in diverse BCI tasks.

Aljlaly et al. [171] explored LSTM networks as an alternative to CNNs for P300 signal detection in BCIs designed for motor-impaired patients. While both architectures achieved comparable accuracy in P300 speller applications, the study highlighted LSTM's potential for modeling temporal dynamics in brain signals, offering a viable deep learning approach for robust P300 recognition despite the signal's inherent noise and complexity.

To address severe class imbalance in RSVP-based BCI systems, Xu et al. [172]

developed BWGAN-GP, a novel data augmentation method combining Wasserstein GANs with gradient penalty and autoencoder initialisation. This approach effectively synthesized minority-class (target) EEG data by learning discriminative features from majority classes. When integrated with EEGNet, the augmented dataset improved the prediction by 3.7% as compared to the non-augmented data. The results demonstrate the ability of GAN-based models to mitigate class imbalance issues and enhance deep learning performance in RSVP paradigms, with potential applications across other EEG-based BCI systems.

While deep learning models have demonstrated strong performance in RSVP-P300 classification, their validation remains largely confined to controlled experimental settings. Real-world applications like engaging environments or their use in clinical settings introduce challenges like variable noise and non-stationary EEG signals.

2.5 Challenges and the Need for Real-World BCI Evaluation

Deploying BCI systems beyond controlled laboratory settings introduces several critical challenges. While BCI performance in ideal conditions has advanced significantly, transferring this performance to dynamic, real-life environments remains non-trivial. One major limitation is the lack of ecologically valid datasets that reflect realistic, noisy conditions, a gap that hinders the development and benchmarking of BCI systems for practical use. Additional challenges include inter-subject variability, the burden of subject-specific calibration, limited training data volume, and degraded signal quality due to real-world artifacts, all of which constrain the robustness and scalability of current BCI solutions.

2.5.1 From Traditional to Noisy Datasets

Prior studies have predominantly focused on EEG signals in isolation, yielding limited success in addressing these artefacts. EEG, known for its accessibility and safety, has been widely utilised in building and operating Brain-Computer Interface applications. The RSVP paradigm, used in a number of BCI applications, involves the rapid display of images, requiring participants to differentiate between target and standard images, to evoke the P300 Event-related Potential. Although the RSVP approach has demonstrated promising results in controlled laboratory settings, its translation into non-laboratory contexts necessitates a deeper understanding of performance in ecologically valid scenarios, including online worlds, metaverse environments, and gaming contexts. These application scenarios are characterised by less constrained user behavior, including talking, head movements, and hand movements, thus involving a greater number of EEG-contaminating artefacts.

Prior studies [135], [173], [174] have typically focused on using characteristics of EEG signals in isolation without contextual signal sources to identify and ameliorate such artefacts with limited success. Presently, however, no suitable dataset exists in order to train and evaluate such approaches.

Several research articles have discussed P300 RSVP datasets available to researchers [142]. Notably, Won et al. [175] collected P300 RSVP data from 50 participants using a 32-channel Biosemi ActiveTwo system. The presented dataset could potentially improve P300 speller applications and is particularly useful for evaluating feature extraction and classification algorithms in various contexts such as cross-subject, cross-dataset, and cross-paradigm BCI model development.

Another dataset was presented by Acqualagna et al. [176], acquired from 12 subjects, which explored non-gaze dependent RSVP paradigms, offering insights into gaze-independent BCI control.

An extensive dataset introduced by Zhang et al. [160], involved 64 subjects recorded with a 64-channel Synamps2 system. This benchmark dataset can support comparisons of RSVP-based target identification algorithms, offline simulations of

novel system architectures, and detailed analyses of ERP and SSVEP components in RSVP-based BCIs.

Furthermore, Xue et al. [177] collected a multi-subject, multi-session EEG dataset (MSS) for modeling human visual object recognition using the RSVP paradigm from 32 participants. Data were recorded using a 128-channel Quik-Cap (Compumedics Neuroscan) with SynAmps 2/RT and Neuvo amplifiers. The MSS dataset can be used to explore characteristics of visual EEG responses, compare visual paradigms, and develop robust cross-subject and cross-session machine learning models for BCIs.

These datasets share a common limitation in that they solely focus on data collected in traditional laboratory environments using standard monitor displays, without incorporating real-world contextual variability. While they provide valuable insights into RSVP-based P300 responses and serve as strong benchmarks for algorithm development, they do not reflect the conditions under which BCIs would be deployed in everyday scenarios. The absence of contextual signal sources, such as those arising from natural head movement, body motion, or conversational interaction, limits their ecological validity. To enable the practical deployment of EEG-based BCIs outside the lab, it is essential to bridge this gap by collecting and analyzing EEG data under realistic, noisy, and dynamic conditions. Only then can we develop systems that are robust and adaptive enough for real-world applications.

2.5.2 Subject-Independent Classification and Calibration

Despite recent advancements in machine learning and signal processing techniques, training models for robust P300 classification continue to be a substantial challenge. One crucial obstacle lies in the variability of EEG signals across individuals. Factors such as differences in brain anatomy, cognitive state, electrode placement, and external noise significantly influence the recorded signals [106].

Consequently, machine learning models trained on subject-specific data often fail to generalise effectively to new users, limiting the scalability and usability of BCIs

in real-world scenarios [178].

Traditional approaches to P300 classification have predominantly relied on preprocessing techniques such as filtering, trial averaging, and dimensionality reduction, followed by classical classifiers like Linear Discriminant Analysis [179], [180]. While effective in subject-specific scenarios, these methods often struggle with generalisation. LDA, in particular, performs poorly with high-dimensional input signals, which are typically required to capture cross-subject variability. Similarly, Support Vector Machines with linear kernels have been widely adopted [181], but they tend to require high computational resources and are usually applied to transformed inputs (e.g., via PCA), which limits their ability to learn transferable, subject-independent features.

To address these limitations, recent studies have explored deep learning techniques such as CNNs and RNNs [182]–[184]. CNNs can extract spatial features from EEG data and learn complex patterns that are difficult for linear models to capture. However, they often demand high-dimensional inputs and remain largely subject-specific, showing limited ability to generalise to unseen individuals. RNNs, particularly suited for sequential data like EEG, aim to capture temporal dynamics more effectively [184]. Yet, their performance on subject-independent tasks remains suboptimal, and they often require computationally intensive data transformations that are not well-optimised for P300 classification [185].

Hybrid architectures that combine spatial and temporal feature extraction have shown promise for single-trial P300 classification [186]. However, these models are still primarily trained in subject-specific contexts and struggle to learn generalised features applicable across subjects.

In addition to the scalability limitations of subject-independent approaches, which often struggle to generalise across users, the burden of user-specific calibration poses another challenge for practical BCI deployment. While subject-independent models aim to reduce or eliminate the need for per-user training, they frequently require some level of fine-tuning to achieve acceptable performance, especially un-

der noisy conditions. Existing solutions have explored transfer learning, such as fine-tuning pre-trained models with 5–10 user-specific trials [187], [188]. However, fine-tuning has primarily been validated on clean, lab-based datasets and tends to degrade in performance when applied to data collected under realistic conditions.

Moreover, an open question remains: how much fine-tuning is truly sufficient, especially under noisy conditions? While a few clean trials may suffice in laboratory settings, it is unclear whether the same applies when calibration data is contaminated by movement, speech, or environmental artifacts. There is currently a gap in the literature exploring the reliability and optimal amount of fine-tuning required on noisy trials. Addressing this issue is essential to designing BCI systems that can adapt quickly and reliably in real-world contexts, where obtaining clean calibration data may not be feasible. Developing models that can be fine-tuned effectively with limited, noisy data would represent a major step toward practical, plug-and-play BCIs.

2.5.3 The Role of Training Data Volume

The volume of training data is a critical factor in determining the effectiveness of EEG-based Brain-Computer Interfaces. In general, larger datasets lead to better generalisation, improved robustness, and enhanced decoding accuracy of brain signals. However, collecting high-quality EEG data at scale is often impractical due to challenges such as participant fatigue, inter-subject variability, and the time-intensive nature of EEG recording sessions.

Despite its importance, the impact of training data size remains relatively under-explored in BCI research. A few notable studies have addressed this issue. Śliwowski et al. [189] systematically investigated the effect of dataset size on the performance of deep learning models using ECoG data. They observed that while model accuracy generally improved with more training data, the gains diminished beyond a certain threshold, approximately 40 minutes of data, suggesting a saturation point where additional data offers limited performance benefit.

In a broader context, Roy et al. [190] conducted a systematic review of deep learning applications in EEG analysis. They pointed out that a significant number of studies are constrained by small dataset sizes, which not only hinders the training of deep architectures but also limits the generalisability of the models across subjects and tasks. The authors emphasized the need for more extensive and diverse EEG datasets to fully exploit the capabilities of deep learning in BCIs.

When access to large datasets is limited, data augmentation has emerged as a practical strategy to expand the training set and improve model performance. Rommel et al. [191] conducted a comprehensive comparison of EEG data augmentation techniques and demonstrated that, in low-data scenarios, appropriate augmentation could enhance classification accuracies. Their findings underscore the potential of augmentation to compensate for insufficient data, especially in real-world BCI applications.

Importantly, while these studies highlight the influence of data volume on model performance, the relationship is often non-linear and highly task-dependent. Factors such as the type of BCI paradigm, target brain signals (e.g., ERPs), and model architecture can influence how much data is "enough."

Despite these insights, a significant gap remains in the literature regarding the systematic exploration of training data volume in EEG-based BCIs. There is a clear need for studies that rigorously compare model performance across varying dataset sizes, particularly under noisy or real-world conditions. Understanding how data requirements scale with different levels of noise and complexity is essential to determine how much data is truly sufficient for reliable and practical BCI deployment outside controlled lab settings.

2.5.4 Optimal Channel Selection

Electrode channel selection plays a critical role in optimising the performance and practicality of P300-based BCI systems. Identifying a minimal yet effective subset of electrodes can reduce system complexity, setup time, and user discomfort.

Numerous studies have investigated the impact of electrode selection on P300 speller performance. Some works, such as [192]–[194], employ spatial filtering methods to identify subject-specific optimal electrode sets. Others, including Ludwig and Kong [195], determine the optimal electrodes empirically. Interestingly, Ludwig and Kong observed significant inter-subject variability: while some users achieved peak performance using just 3 electrodes, others required up to 10 out of a 14-electrode array. Despite this variability, parietal and occipital electrodes consistently emerged as the most informative regions.

However, tailoring the electrode configuration for each user typically requires data collection with a large electrode set before identifying a smaller optimal subset, a process that can be impractical in clinical or commercial deployments. To address this, Krusienski et al. [196] proposed a fixed 6-electrode configuration (Fz, Cz, Pz, PO7, PO8, Oz) that performed comparably or better than larger 19-channel setups across users. Similarly, Speier et al. [197] found that a reduced set of channels (PO7, PO8, POz, and CPz) was nearly as effective as a full 32-channel montage. McCann et al. [198] systematically evaluated electrode sets ranging from 1 to 16 channels and reported only marginal gains beyond 4 electrodes. Additionally, Noble et al. [199] examined configurations of 32, 16, 8, 6, and 4 channels, concluding that a 6-electrode set (P3, Pz, P4, POz, O1, and O2) without spatial filtering yielded the highest performance. Furthermore, Sahay et al. [200] identified 8-channels (P7, P3, Pz, Oz, P4, P8, Fz, and Cz) as an optimal configuration for P300-based classification, striking a balance between accuracy and channel reduction.

These findings underscore the importance of identifying an optimal electrode subset, as reducing channel count is crucial for real-world BCI use where portability, ease of use, and user comfort are essential.

While existing studies have made significant progress in optimising electrode configurations for P300-based BCIs, several critical limitations remain unaddressed. First, most research has focused predominantly on posterior and occipital regions, potentially overlooking the role of other brain areas that may become relevant in

real-world usage scenarios. Second, few studies have systematically evaluated progressive channel reduction (e.g., comparing 32, 16, 8, 6, and 4 channel configurations) to determine the precise trade-off between channel count and classification accuracy. Third, and perhaps most importantly, current approaches have been validated primarily under controlled laboratory conditions, leaving open questions about their performance in practical settings with motion artifacts, variable electrode contact, or when integrated with head-mounted displays. These gaps highlight the need for more comprehensive investigations that address both technical performance and real-world usability constraints.

2.5.5 Noise and Artifact Robustness

There is an extensive body of literature that has explored various approaches for noise mitigation in EEG signals within the domain of Brain-Computer Interfaces, where researchers have investigated challenges and advancements in efficiently filtering unwanted noise to improve the reliability and accuracy of EEG-based BCIs [201], [202]. Non-biological sources of noise, such as poor electrode settings, high impedances, and electrical interference, along with non-neural noise sources like eye movements, muscle activity, cardiac signals, and body movements, play a significant role in the contamination of EEG signals [4].

Recognizing the importance of artifact removal in EEG-based BCIs, researchers employ a range of techniques, including both fundamental and advanced methods to effectively eliminate noise from EEG signals [203].

Researchers have extensively utilised Independent Component Analysis (ICA) to eliminate noise sources from EEG signals recorded in traditional lab environments, with particular emphasis on ocular and muscular artifacts. Dimigen et al. [204] proposed an artifact removal method based on ICA optimisation by systematically varying a range of parameters, including high-pass and low-pass filters, and thresholding for component rejection. They reported a significant improvement in performance with minimal distortion of neural activity. Gao et al. [205] proposed

a combination of canonical correlation analysis (CCA) and ICA to adaptively remove ocular artifacts from the EEG data. Liu et al. [206] explored the effect of ocular artifacts on EEG recordings during both rapid eye movement (REM) and non-REM sleeping stages. Shi et al. [207] proposed a denosing method by using CCA and second-order blind identification (SOBI) followed by adaptive and strict fixed thresholds to localize and remove artifacts. In addition, some researchers [208], [209] have employed ICAs to eliminate various noise sources, including movement artifacts.

Moreover, the literature showcases other diverse approaches to noise mitigation in EEG signals. Specifically, Cimmino et al. [210] and Javed et al. [211] leveraged principal component analysis, a few researchers [212]–[214] applied regression methods and wavelet transforms, whereas brophy et al. [215] and others [216]–[218] employed GANs and CNNs for addressing various noise sources in EEG signals recorded in traditional lab settings.

However, a significant research gap remains regarding the management of EEG signals in real-world, out-of-the-lab environments. Most existing noise removal techniques are designed and evaluated under clean, controlled conditions and often fail to generalise well when applied to signals contaminated by real-life artifacts such as movement, speech, or environmental noise. Furthermore, many of these methods are computationally intensive, requiring considerable processing time and power, making them less feasible for real-time or portable BCI systems. Another key concern is that aggressive denoising may unintentionally discard informative neural data along with noise, thereby reducing the quality and quantity of usable EEG information. These limitations highlight the need for more adaptive, efficient, and artifact-resilient preprocessing methods tailored for practical, real-world BCI deployment.

2.6 Emerging Directions: Robust and Adaptive Models

To address the inherent variability in EEG signals and the limitations of conventional classifiers, recent research has begun to explore more flexible and expressive deep learning models. Among these, Transformer-based architectures [219], [220] and hybrid models [221]–[224] represent promising directions due to their ability to model long-range dependencies and complex spatiotemporal dynamics in neural data.

Transformers, initially developed for natural language processing tasks, have been successfully adapted to EEG decoding tasks thanks to their self-attention mechanism, which can dynamically weigh contributions from different channels and time points [225]. This allows them to capture both global and local patterns in EEG signals more effectively than traditional CNNs or RNNs [226].

Hybrid models that combine CNN-based architectures with Transformer layers have also gained attention. For example, EEGNet + Transformer frameworks aim to extract frequency-specific features via EEGNet and refine temporal or spatial relationships using Transformer encoders [223], [227]. Similarly, EEG-Conformer models integrate convolutional and self-attention modules with feed-forward layers to exploit both local dependencies and global contextual cues, offering a robust approach to modeling EEG signals [222]. In parallel, contrastive learning techniques have been introduced to learn subject-invariant and noise-resilient representations, especially in subject-independent settings, by encouraging the model to cluster similar neural patterns while distinguishing unrelated ones [228]–[230].

Despite recent advancements in deep learning for EEG decoding, the application of such architectures remains limited in the context of P300-based RSVP paradigms. Most existing Transformer-based and hybrid EEG models have predominantly focused on domains like motor imagery, emotion recognition, or cognitive workload assessment, with comparatively little emphasis on RSVP tasks. Moreover, their robustness in the presence of noise or under naturalistic conditions, key for real-world

BCI deployment, remains largely unexplored.

In particular, the integration of Transformers and their hybrid variants into diverse EEG datasets that include artifact-contaminated or ecologically valid scenarios is still an open research area. Exploring such models in noisy environments can unlock their true potential for adaptive and generalisable real-world BCI systems. Furthermore, advanced strategies like multiband processing [231]–[233] or Mixture-of-Experts (MoE) architectures [234]–[237], designed to capture frequency-specific and spatially diverse EEG features, have shown promise in other domains but are still underexplored with Transformers and in RSVP-based BCIs. Investigating these approaches could lead to more resilient models capable of handling the complex variability seen in real-life EEG applications.

2.7 Summary

Brain-Computer Interfaces enable direct communication between the brain and external devices, with applications ranging from clinical neurorehabilitation to non-clinical domains such as gaming and assistive technologies. Among various neuroimaging techniques, EEG remains the most widely adopted modality for BCI development due to its non-invasive nature, portability, and high temporal resolution. While laboratory-based studies have significantly advanced BCI research, translating these systems into real-world settings remains a major challenge.

One critical bottleneck is the lack of datasets recorded in real-world environments. Most existing EEG datasets are collected in highly controlled lab conditions, limiting their ecological validity. To bridge the gap between research and deployment, it is imperative to record and utilise datasets that reflect the complexities of naturalistic environments. Such data can facilitate the development of models robust to real-world variability and noise. This limitation is addressed in Chapter 3 through the introduction of a novel dataset that captures EEG data in both traditional laboratory and real-world environments.

This lack of dynamic data contributes to an additional limitation: the under-

explored nature of how real-world artifacts, such as movement, speech, and environmental interference, affect EEG signal quality and processing. This limitation has been addressed in Chapter 4, which systematically investigates the impact of various noise types on EEG signal characteristics and classification performance.

Moreover, most BCI systems have been designed and evaluated in a subject-dependent fashion, which restricts scalability and practical use. In real-world scenarios, subject-independent classification is essential to ensure that models generalise across users without extensive retraining. However, even with subject-independent models, a minimal level of fine-tuning or calibration is often needed. A key gap in the literature is understanding how much calibration data is sufficient, particularly when dealing with noisy EEG signals. These limitations are addressed in Chapter 5, where a range of models, from traditional to advanced adaptive architectures, are explored for subject-independent classification, and in Chapter 6, which investigates fine-tuning strategies and the amount of calibration data required.

Another underexplored factor is the role of training data volume. While it is well-acknowledged that larger datasets tend to yield more robust models, systematic investigations into how data volume affects subject-independent classification, especially in noisy or unconstrained settings, are still limited. This research gap is addressed in Chapter 6, where the effect of training data volume is systematically explored.

In addition, the number of electrodes used in traditional EEG-based BCIs is often impractical for real-world use. While identifying an optimal electrode subset that balances performance with practicality remains a priority, current approaches exhibit notable limitations. Most studies focus narrowly on posterior-occipital regions, potentially overlooking configurations better suited for real-world noise or motion artifacts. Few systematically evaluate progressive channel reduction to quantify accuracy trade-offs, and none address the unique constraints of HMD-integrated systems, where electrode placement must accommodate physical hardware obstructions. These gaps highlight the need for electrode optimisation frameworks that prioritize

both classification robustness and real-world usability, particularly for wearable applications where portability and ergonomics are paramount. This gap is addressed in Chapter 6, where progressively reduced channel configurations are systematically examined.

Noise and artifact removal remains a central concern in EEG processing. While various algorithms exist, many are computationally expensive and may result in the loss of important signal components. This highlights the need for adaptive and artifact-resilient models that can operate effectively without intensive preprocessing.

Emerging directions such as transformer-based architectures and hybrid models (e.g., EEGNet combined with transformers) offer promising avenues for developing robust BCIs. However, these methods are still largely explored in subject-dependent settings under controlled conditions, and their adaptation to real-world, subject-independent scenarios remains an open and vital research direction. This is addressed in Chapter 5 by exploring a number of advanced methods to investigate their performance in the presence of noise and assess their resilience to real-world artifacts.

Chapter 3

Systematic EEG Data Collection for Real-World BCIs: Protocol Design and Acquisition

Brain-Computer Interfaces have shown immense promise in enabling direct communication between the brain and external devices. As discussed in Section 2.5.1 of the literature review, one of the major limitations in current BCI research is the lack of datasets that reflect realistic, noisy, and dynamic environments. Most existing datasets supporting BCI research and development are limited to highly controlled laboratory environments. These datasets fail to account for the challenges of real-world applications, where various sources of noise and dynamic environments can significantly affect performance and reliability.

To bridge this gap, I introduced "AMBER: Advancing Multimodal Brain-Computer Interfaces for Enhanced Robustness – A Dataset for Naturalistic Settings". This dataset was designed to provide a comprehensive resource for researchers aiming to build robust BCIs capable of functioning outside traditional laboratory environments. Following the success and learnings from AMBER, its successor AMBER 2.0 was developed, which extends the experimental design to scenarios even closer to real-world applications.

This chapter addresses the contribution of systematically collecting EEG data under controlled and intentionally induced noisy conditions using different display modalities. Specifically, it outlines the development of two complementary datasets: the AMBER dataset, which captures EEG recordings during RSVP tasks under well-defined noise scenarios (e.g., body movement, head movement, and speech), and the AMBER 2.0 dataset, which extends this work by incorporating a more ecologically valid setup using a head-mounted display. Together, these datasets provide a structured foundation for investigating BCI robustness in both traditional and real-world-inspired environments.

Both the AMBER and AMBER 2.0 datasets have been used to support the analyses presented in Chapters 4, 5, and 6 of this thesis. These datasets, collected under controlled and real-world conditions, provide a comprehensive foundation for investigating noise impact, model robustness, fine-tuning strategies, and subject-independent classification performance.

3.1 Dataset 1 – AMBER

The AMBER dataset incorporates a multimodal and contextual approach by capturing video data of the participant alongside the EEG recording. The inclusion of video data enriches the dataset by providing contextual information and additional modalities for analysis. By recording video data simultaneously with EEG signals, researchers can leverage a broader range of contextual cues, such as facial expressions, body movements, and eye movements.

EEG is an accessible and safe method for researchers and users to build and operate BCI applications [238]. While the initial use of such techniques began in clinical/rehabilitative settings for the purposes of augmenting communication and control, a recent trend has been to use such signals and methods in new domains, such as the image annotation task, which relies on the identification of target brain events to trigger labelling [239]–[241].

The Rapid Serial Visual Presentation is an approach to BCI in which a series of

images is displayed at high speed. Participants are asked to differentiate between a set of target images and a set of standard images, where P300 ERP is evoked by a target image but not by standard images [9], [10].

While the RSVP-BCI paradigm can be demonstrated in controlled lab-based environments, translation of this paradigm into consumer contexts requires a better understanding of real-world EEG artefacts that impact EEG signal quality in ecologically valid settings e.g., in online worlds, metaverse, and gaming contexts. Such application scenarios are characterised by less constrained user behavior, some of which is entirely necessary for the normally expected interactions. Examples include talking, head and hand movements, all of which generate artefacts that impede the application of EEG in BCIs in real-world contexts.

EEG results obtained in the presence of noise and those obtained in noise-free conditions are differentiated using the P300/RSVP task in this dataset. This differentiation allows for a comprehensive analysis of the impact of noise on EEG signals and facilitates the evaluation of signal-denoising techniques. The P300/RSVP task is particularly relevant as it involves measuring accuracy, making it an effective dependent variable to assess the efficacy of noise cleaning and evaluate the influence of behavioral artefacts on the EEG signals. Furthermore, this can provide insights into the relationship between noise, behavioral artefacts, and the quality of EEG signals, ultimately enhancing our understanding of the robustness of EEG data in real-world settings.

For the purpose of creating this dataset, participants were instructed to produce particular artefacts at particular times via a carefully controlled protocol, e.g., moving head left to right versus up and down, eye movement, eye blinks, facial expressions, lip movement, body movement, etc. The specific artefacts that participants were instructed to produce during data recording reflect the most problematic artefacts encountered in real-world EEG recording [202], [242], [243].

Moreover, the AMBER dataset represents a significant advancement in the field of brain-computer interfaces by providing researchers with a resource to address

one of the key challenges in EEG data analysis, signal denoising. EEG recordings are often plagued by various artifacts and noise, which can obscure the underlying neural signals and hinder accurate analysis. The importance of signal-denoising lies in its potential to enhance the quality and reliability of EEG measurements, enabling more accurate identification and interpretation of neural responses. By effectively removing unwanted artifacts and noise, researchers can gain deeper insights into brain activity and cognitive processes, leading to a more comprehensive understanding of neural mechanisms. Moreover, denoising techniques are crucial for developing robust BCIs that can reliably detect and interpret brain signals in real-world scenarios. The dataset serves as a valuable resource for exploring and developing novel signal denoising techniques, ultimately paving the way for more effective and practical applications of BCIs in naturalistic settings.

The AMBER dataset aims to support researchers in developing signal-denoising techniques that mitigate the impact of noise sources such as eye blinks, eye gaze, talking and body movements, in order to ameliorate the signal-to-noise characteristics of Electroencephalography measurements. The emphasis of this research is to enable the robust performance of brain-computer interface systems in naturalistic real-world settings, i.e., outside of the lab.

3.1.1 Data Acquisition

Ten healthy participants aged between 20 and 35 years were recruited from Dublin City University to participate in the data collection. Among these participants, there were 6 males and 4 females, and each participant was assigned an alias ranging from "P1" to "P10". Data acquisition was performed with approval from the Dublin City University (DCU) Research Ethics Committee (DCUREC/2021/175).

Prior to the experiment, each participant was briefed about the study protocol, including its purpose, procedures, and potential risks. A plain language statement detailing this information was provided, followed by an informed consent form to ensure voluntary participation. Both documents, the plain language statement and

the informed consent form, are included in Appendix A.

The hardware used for data collection was the ANT-Neuro eego sports mobile EEG system. A 32-channel EEG cap positioned according to the 10-20 international electrode system was used for the data acquisition (See figure 3.1). CPz was used as the online reference channel, and the impedance of all electrodes was kept under 15 kOhm. Data was collected at a sampling rate of 1000 Hz (using a lowpass filter of 500 Hz) and saved in EDF format.

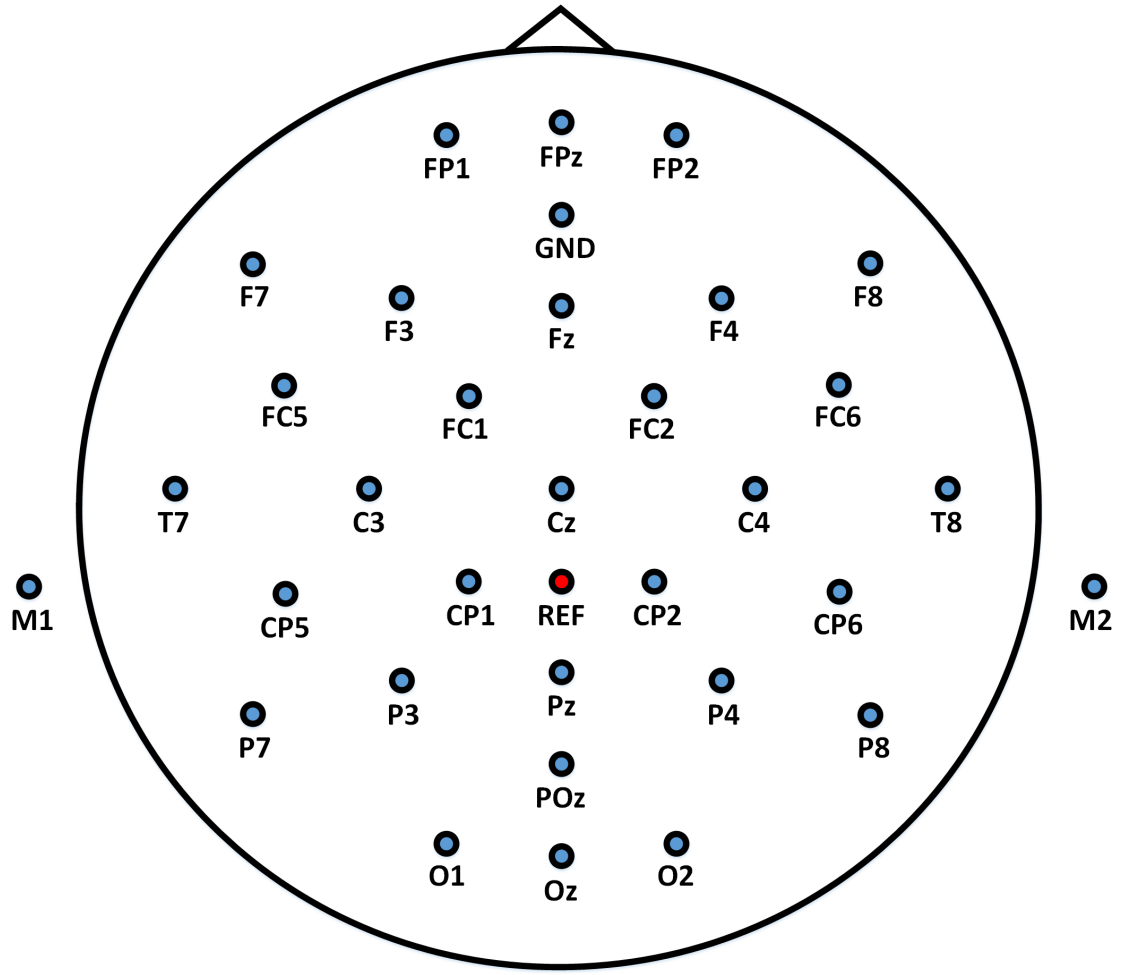


Figure 3.1: Standard 10–20 EEG electrode configuration

EEG was recorded from the ten participants while they followed a pre-defined protocol of tasks. Timestamp information for image presentation (via a photodiode and hardware trigger) was also captured to allow for precise epoching of the EEG signals for each trial [244].

3.1.2 Experimental Protocol

The AMBER dataset employs a multimodal approach that encompasses two primary signal sources: 1) EEG data collection and 2) video recording. Both EEG recordings and video are captured at the same time.

EEG Recording

The dataset contains EEG responses to 10,500 images, in total. Each participant completed 4 sessions, where each session contained 8 blocks, followed by 3 different baselines. Each session lasted approximately 30 minutes, including inter-trial intervals. Between sessions, participants were given a short break; the duration of the break was not fixed but typically lasted around 5 minutes. During these breaks, the EEG cap remained in place. Before the experiment, participants were instructed in detail on how to perform the required movements and actions, and were reminded to minimize unnecessary movements during the recordings. A description of each task in a single session is given in Table 3.1.

Table 3.1: AMBER: Description of tasks performed in a single session

No.	Task ID	Task Description	Duration
1.	B1	Baseline Eyes Open	60s
2.	B0	Baseline Eyes Movement	10s
3.	B2	Baseline Eyes Close	60s
4.	X1	RSVP Task	90s
5.	X2	RSVP Task	90s
6.	X3	Artefact 1: Body Movement	90s
7.	X4	RSVP + Body Movement	90s
8.	X5	Artefact 2: Talking	90s
9.	X6	RSVP + Talking	90s
10.	X7	Artefact 3: Head Movement	90s
11.	X8	RSVP + Head Movement	90s

As seen in Table 3.1, the data collection is split into three sections:

1. The standard RSVP image search paradigm, in which the subjects perform the target search task while sitting still in front of the monitor (X1 and X2);

2. The RSVP paradigm with participant-induced noise, in which the participants

through following a protocol, generate three distinct types of EEG artefactual noise through movements (i.e., head movement, body movement, and talking) while doing the RSVP task (X4, X6 and X8);

3. The noise paradigm, in which the participants generate three distinct types of noises throughout the trial without performing any RSVP image search task (X3, X5 and X7).

In the image search RSVP task, participants searched for a known type of target (e.g., a car) and were instructed to covertly count occurrences of target images in the RSVP sequence so as to maintain their attention on the task. Figure 3.2 presents examples of the target search images used. In each 90-second RSVP block, images were presented successively at a rate of 4 Hz with target images randomly interspersed among standard images with a percentage of 10% across all blocks. In each block, 360 images (36 targets/324 standards) were presented in rapid succession on screen. The images selected for presentation were chosen in accordance with standard RSVP paradigms reported in the literature [10], [245].

There were 288/2592 target/standard trials captured for the standard RSVP task per participant, whereas, in the case of each noisy RSVP task (1-body movement, 2-talking and 3-head movement), there were 144/1296 target/standard trials available.

In addition to recording using the pre-defined RSVP paradigm, intentional artefacts were also generated by the participants following a protocol. In the first scenario, participants generated noise in parallel with the RSVP task, and in the second scenario, they generated artefactual noise without performing the RSVP task.

These intentional artefacts were induced to simulate realistic scenarios and study the effects of body movement, talking, and head movement on EEG data. By performing these intentional artefact-generating tasks in parallel with the RSVP task, this dataset captures the effect of coincident noise on task performance and EEG data. The inclusion of intentionally generated artefacts also enables a more comprehensive analysis of the noise characteristics and their impact on EEG data, enabling researchers to develop more effective signal-denoising techniques.

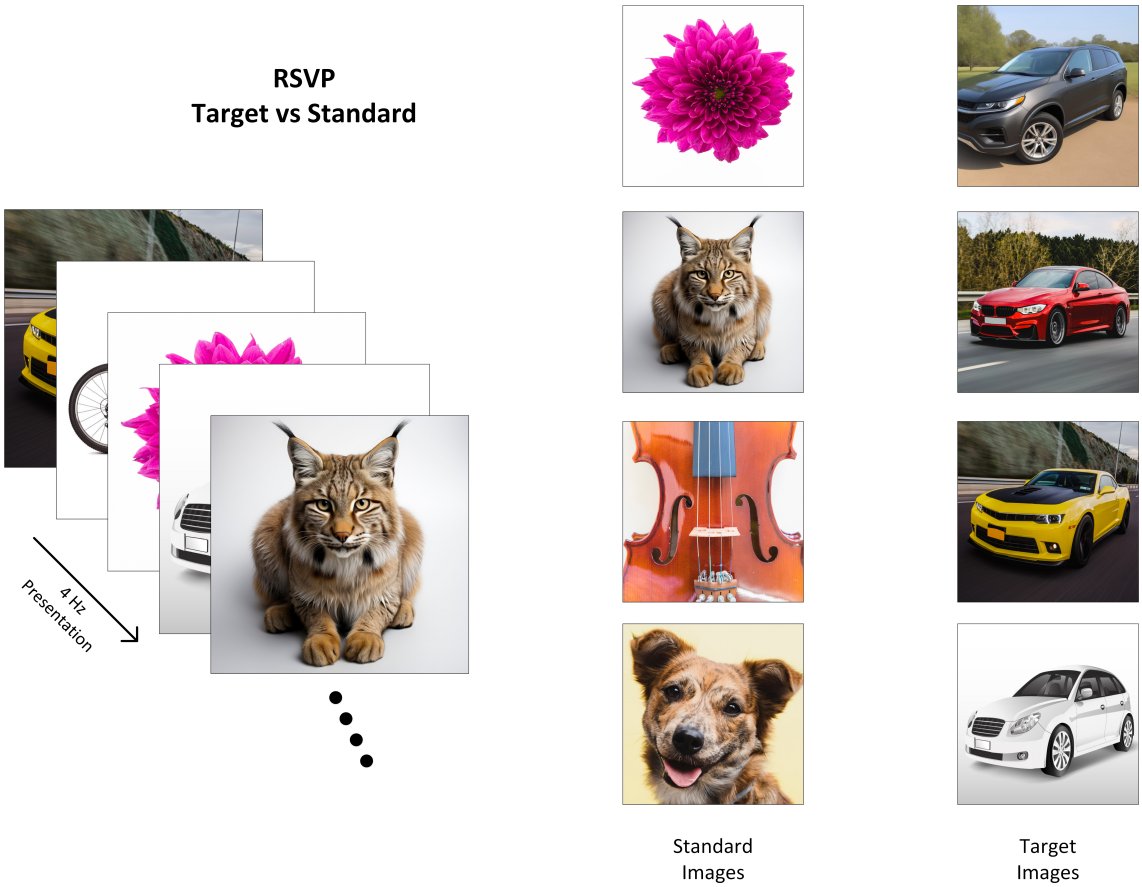


Figure 3.2: Illustration of the P300/RSVP task where images of cars have been used as the target class.

Artefact 1: Body Movement

To induce body movement artefacts, participants were instructed to repeatedly raise and wave their hands, followed by putting their hands down, and repeating this sequence of movements for the entire duration of the 90-second block. Participants were given the freedom to choose which hand(s) to raise and in what sequence, with the intention of inducing variability around artefact production. They were instructed to perform the task at a comfortable speed, neither too fast nor too slow, in order to maintain consistency across participants. This task was designed to simulate a real-world or metaverse-type environment where a person might be moving their hands and arms while using a virtual reality setup such as playing a game or engaging in any other activity that involves body movements.

Artefact 2: Talking

To generate talking artefacts, participants were instructed to count aloud, mixing numbers and letters in random sequences, for the entire 90-second block. They were given the freedom to choose the order and sequence of their counting, with the aim of inducing variability.

During the RSVP task, participants were instructed to count and continuously repeat the number of target images out loud to create talking artefactual noise in the EEG. The aim of this instruction was to simulate a real-life scenario where individuals may need to focus on a task while simultaneously communicating verbally in situations like playing a game or collaborating with others in a virtual space.

Artefact 3: Head Movement

To generate head movement artefacts, participants were instructed to move their heads in an up-down (nodding) or left-right motion, at a natural speed, throughout the entire 90-second block. Participants were specifically instructed to alternate between up-down and left-right head movements in a randomised manner. This task was designed to simulate real-world scenarios where a person might produce head movements in situations like nodding in agreement during a conversation or shaking their head in response to a question. Additionally, this task aimed to simulate VR environments, where a person might use a VR headset and move their head to explore different directions while playing games or visiting virtual spaces.

Figure 3.3 illustrates the organisation of the EEG recordings for each participant. It also provides an overview of the files (i.e., raw EEG data and labels) associated with each task.

3.1.3 Video Recording

A multimodal approach was employed that involved capturing video data of BCI participants while they produced specific artefacts.

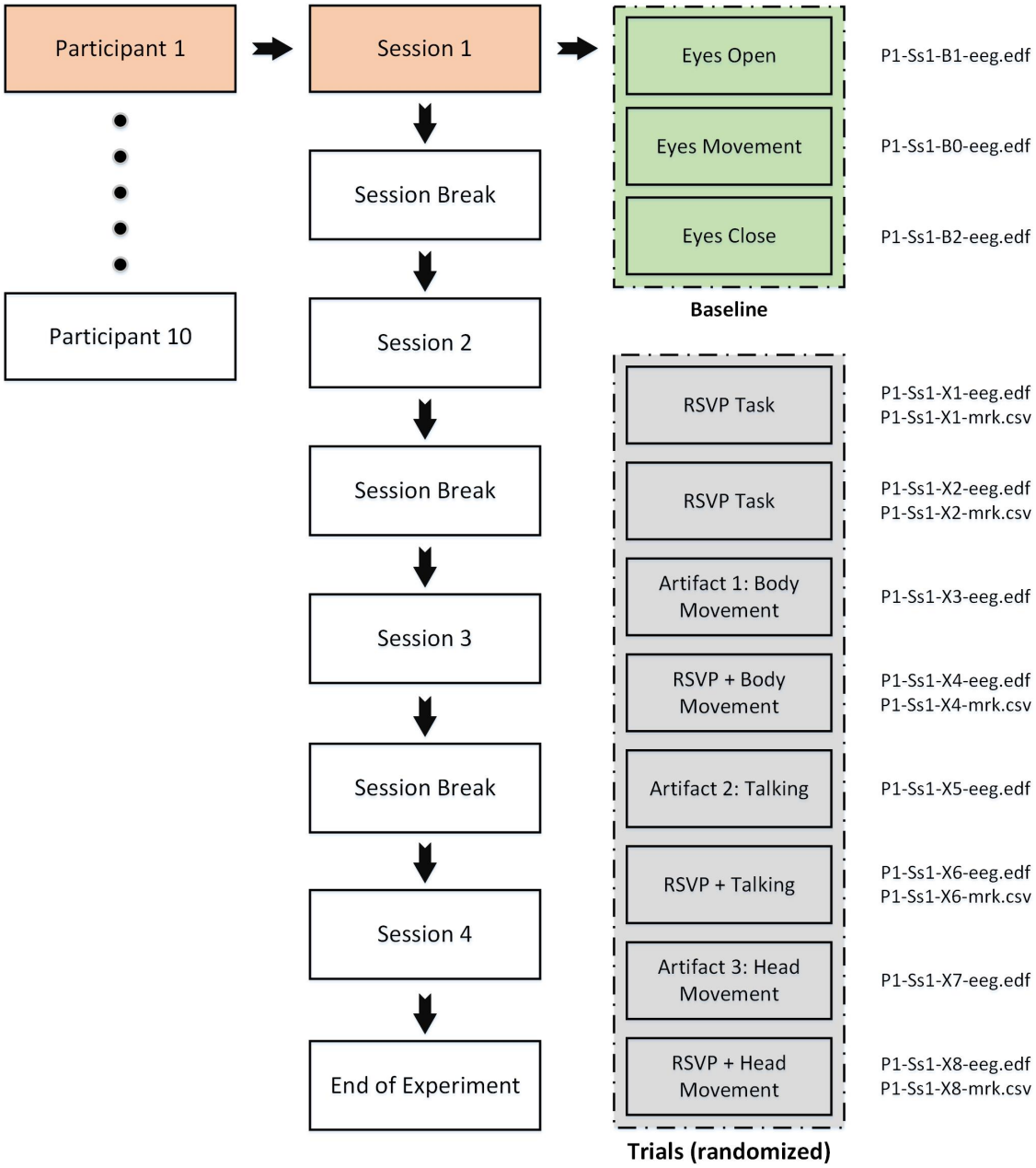


Figure 3.3: AMBER: Organisation of the EEG recordings for each participant

Video Data Collection and Framing

To accurately capture the various artefacts, video data was recorded using a Logitech C920 webcam, which has a frame rate of 20 fps. The camera was calibrated using OpenCV [246] to determine the camera matrix and distortion coefficients, ensuring the accuracy of the recorded video data.

During the recording of eye, head, and mouth-related artefacts, the participant's face was kept fully in the frame. This framing approach allowed for a clear and

unobstructed view of the participant's movements, including pupil dilation, eye-opening size, and movement. For arm movements, the upper torso was maintained in full frame, ensuring proper visibility of the body movements.

Video Data Annotation and Time-stamping

Each video frame was timestamped and assigned a corresponding frame number, which was recorded in a corresponding CSV file. This file also contained information on frame size, zoom factors, and camera matrix. The timestamping process ensured that the video data could be accurately synchronized with the EEG data, allowing for a robust multimodal analysis.

Significance of Video Data in Contextual EEG Dataset

The integration of video data into the Contextual EEG Dataset is critical for developing advanced signal-denoising techniques and improving BCI performance in real-world settings. Combining video data with EEG recordings enables researchers to explore the correlations between facial expressions, head movements, and brain activity, leading to a better understanding of various (neuro-)physiological phenomena. By capturing these problematic artefacts encountered in naturalistic environments, researchers can better understand the factors affecting EEG data quality and develop solutions to mitigate their impact on BCI performance.

3.1.4 Data Records

In this work, EEG data from 10 participants were collected, which is labeled as Dataset A. For illustrative purposes, basic pre-processing steps were conducted on the raw EEG data, resulting in a modified version referred to as Dataset B. Alongside the EEG recordings, video data were simultaneously captured, which were independently processed to extract relevant information. These video-derived extractions are denoted as Dataset C. The availability of multiple datasets enables comprehensive analysis, allowing us to explore the relationship between EEG signals, video

data, and their potential combined insights. Detailed descriptions of each dataset are provided below.

Dataset A: Raw EEG Data

The raw EEG data recorded during each task is stored in EDF file format, which contains all the relevant information, including the complete EEG signal as well as the events that occurred during the task. For each participant, there are 11 EDF files corresponding to each session, which results in a total of 44 files for all four sessions.

The RSVP tasks have additional information about the target and standard image triggers, which are given in the form of CSV files. There are five RSVP tasks per session (two standard and three noisy RSVP tasks), which results in 20 CSV marker files for each participant. This combination of EDF and CSV files provides a comprehensive dataset that allows for detailed analysis of EEG signals and their corresponding events during the RSVP tasks. An overview of the files (i.e., EDF and CSV) associated with each task in a session can be seen in Figure 3.3.

During each RSVP task, a total of 360 images were presented to participants. As a result, each corresponding CSV marker file contains 360 rows of information having two distinct labels, "1" and "2". The label "1" indicates the presence of a target image in the corresponding epoch for the RSVP task, while the label "2" indicates the absence of a target image and, therefore, a standard non-target image. These labels provide vital information for the analysis of the EEG data as they allow the identification of the specific moments in time when the target and non-target stimuli were presented during the task.

Dataset B: Processed EEG Data

In addition to the raw data (i.e., Dataset A), an illustrative pipeline has been provided that encompasses crucial preprocessing steps to make the EEG data amenable to analysis, visualisation, and machine learning. This transformation procedure was

proposed as an example to restructure the continuous raw data into a more compact dataset and to make it easier to use.

The pipeline includes essential stages such as filtering, resampling, re-referencing, and epoching. By following this example pipeline, researchers can effectively transform and structure the data, allowing for a deeper understanding of its characteristics and facilitating further analysis. Python (version 3.9.12) was mainly used along with the MNE (version 0.24.0) to implement the processing. This example pipeline is available at `https://github.com/meharahsanawais/AMBER-EEG-Dataset/blob/main/Example_Code.ipynb`, allowing interested researchers to modify the processing setup as they wish.

Raw data loading

A function was developed in Python to quickly load the raw EEG data for a particular participant and session. This function is helpful for efficiently accessing the EEG data for analysis and processing. The raw data is stored in the EDF file format, which contains all the relevant information related to the EEG recordings, including the event markers. Further information about event markers is stored in separate CSV files for each RSVP task.

Digital filtering

A low-pass IIR filter with a cut-off frequency of 30 Hz was applied to the EEG data to remove any high-frequency noise or artefacts that may be present in the signal above 30 Hz. This filtering step helped to improve the signal-to-noise ratio and enhance the quality of the EEG data.

Re-sampling

The original signal, sampled at 1000 Hz, was down-sampled to a new sampling rate of 100 Hz. The down-sampled signal can help reduce the computational load of subsequent processing steps while preserving the essential features of the signal.

Re-referencing

The raw EEG data were initially recorded using the default CPz reference. However, in order to improve the quality of the data, common average referencing (CAR) was applied using the MNE re-reference function.

Epoching

The continuous EEG recordings were segmented into epochs by extracting the time series data from -0.2 seconds to 1 second relative to the onset of the visual stimulus.

After undergoing the aforementioned pre-processing steps, the data is saved into CSV files. Each session comprises 11 CSV files, resulting in a total of 44 files for each participant. Each file contains comprehensive information about the channels, epochs, and their respective categories. By providing access to this pre-processed data as an example, users can explore the raw datasets in a multitude of ways, unlocking various possibilities for analysis and interpretation.

Data C: Video Data

This section describes the process of extracting metadata from video recordings using OpenCV libraries and synchronizing it with EEG data recorded in the European Data Format (EDF).

Video Processing and Metadata Extraction

Video recordings can provide valuable information on an individual's facial expressions, head position, and skeletal movements. To extract this information, OpenCV libraries were utilised, which offer a wide range of functionalities for image and video analysis. The metadata extracted includes:

1. Three-dimensional head position
2. Eye opening
3. Mouth opening

4. Two-dimensional eye position

5. Three-dimensional skeleton movement

Each of these metadata elements is timestamped to ensure accurate synchronisation with the EEG data.

Data Conversion and Representation

Once the metadata is extracted from the video, it is converted into a suitable format for integration with the EEG data. Each metadata element is represented as a graph line to facilitate the visualisation and analysis of the combined data. This graphical representation allows researchers to easily identify patterns and correlations between video metadata and EEG activity.

Synchronisation with EEG Data

In order to accurately integrate the video metadata with the EEG data, the video frame timestamps are used as a reference for aligning the metadata with the EEG data. This ensures that each metadata element is correctly associated with the corresponding EEG activity.

The combined analysis of video metadata and EEG data has numerous applications in various fields, such as neuroscience, psychology, and human-computer interaction. By understanding the relationship between facial expressions, head movements, and brain activity, researchers can gain insights into emotion recognition, attention, and cognitive processes. Moreover, this approach can also be applied to develop advanced human-computer interfaces and improve the accuracy of brain-computer interfaces [247].

3.2 Dataset 2 – AMBER 2.0

Despite the extensive use of RSVP paradigms for P300 elicitation, most studies rely on traditional computer monitor displays. This experimental setup offers a

consistent and familiar visual environment, reducing distractions and delivering accurate data for P300 detection. However, real-world applications of BCIs frequently go beyond the boundaries of a standard laboratory setting, necessitating systems that work in dynamic and engaging environments. For example, BCIs designed for immersive gaming or other settings that mimic real-world situations.

Building upon the foundation of AMBER, the AMBER 2.0 dataset was introduced to simulate more immersive real-world scenarios. The primary innovation in AMBER 2.0 is the use of a head-mounted display for RSVP image presentation, replacing the traditional monitor setup. This change was motivated by the need to explore BCI performance in virtual and augmented reality environments, which are increasingly relevant in applications such as assistive technology, gaming, and training simulations. The RSVP experiment was displayed using the "VisionHMD bigeyes H3" head-mounted display device, as shown below.

Figure 3.4: VisionHMD bigeyes H3 head mounted display

3.2.1 Dataset Structure

AMBER 2.0 also includes data collected from four sessions for each participant, but instead of eleven blocks per session as in AMBER, it contains seven blocks per session, resulting in a total of 28 blocks across all four sessions. On average, participants were given a break of about 5 minutes between sessions, similar to the AMBER dataset. A description of each task in a single session is given in Table 3.2.

As seen in Table 3.2, the data collection excluding the baselines is split into two

Table 3.2: Description of tasks performed in a single session

No.	Task ID	Task Description	Duration
1.	B1	Baseline Eyes Open	60s
2.	B0	Baseline Eyes Movement	10s
3.	B2	Baseline Eyes Close	60s
4.	X2	RSVP Task	90s
5.	X4	RSVP + Body Movement	90s
6.	X6	RSVP + Talking	90s
7.	X8	RSVP + Head Movement	90s

sections:

1. The standard RSVP image search paradigm, in which the subjects perform the target search task while wearing a head mount display (X2);

2. The RSVP paradigm with participant-induced noise, in which the participants, through following a protocol, generate three distinct types of EEG artefactual noise through movements (i.e., head movement, body movement, and talking) while doing the RSVP task (X4, X6, and X8).

Similar to AMBER, a total of 360 images per RSVP block were presented to participants in AMBER 2.0. The image selection procedure was the same as in AMBER; however, these images were displayed using a head-mounted display instead of a monitor screen. Consequently, each corresponding CSV marker file contains 360 rows of information, with two distinct labels: "1" and "2." The label "1" denotes the presence of a target image in the corresponding epoch for the RSVP task, while the label "2" signifies the absence of a target image, representing a standard non-target image.

There were a total of 144/1296 target/standard trials captured for each intentionally as well as non-intentionally contaminated RSVP task per participant.

3.2.2 Data Records

The raw EEG data recorded from 10 participants during each task are stored in EDF file format, which contains all the relevant information, including the complete

EEG signal as well as the events that occurred during the task. For each participant, there are 7 EDF files corresponding to each session, which results in a total of 28 files for all four sessions.

The RSVP tasks have additional information about the target and standard image triggers, which are given in the form of CSV files. There are total four RSVP tasks per session (one traditional and three noisy RSVP tasks), which result in 16 CSV marker files for each participant. This combination of EDF and CSV files provides a comprehensive dataset that allows a detailed analysis of EEG signals and their corresponding events during RSVP tasks. An overview of the AMBER 2.0 files (that is, EDF and CSV) associated with each task in a session can be seen in Figure 3.5.

3.2.3 Differences from AMBER

Display Technology: The use of an HMD instead of a monitor introduces a more immersive and dynamic testing environment.

1. Task Design: The exclusion of tasks involving only noise (e.g., X3, X5, X7 from AMBER) simplifies the dataset while maintaining its focus on RSVP tasks under noisy conditions.

2. Video Recordings: Unlike AMBER, AMBER 2.0 does not include video recordings, emphasizing EEG data collection and analysis.

3.2.4 Applications and Advantages

AMBER 2.0 reflects the growing demand for datasets that replicate real-world BCI scenarios. By incorporating HMD-based testing, it facilitates:

1. Research into BCIs for virtual and augmented reality settings.

2. Comparative studies on the impact of display technology on BCI performance.

3. Development of robust algorithms tailored to immersive environments.

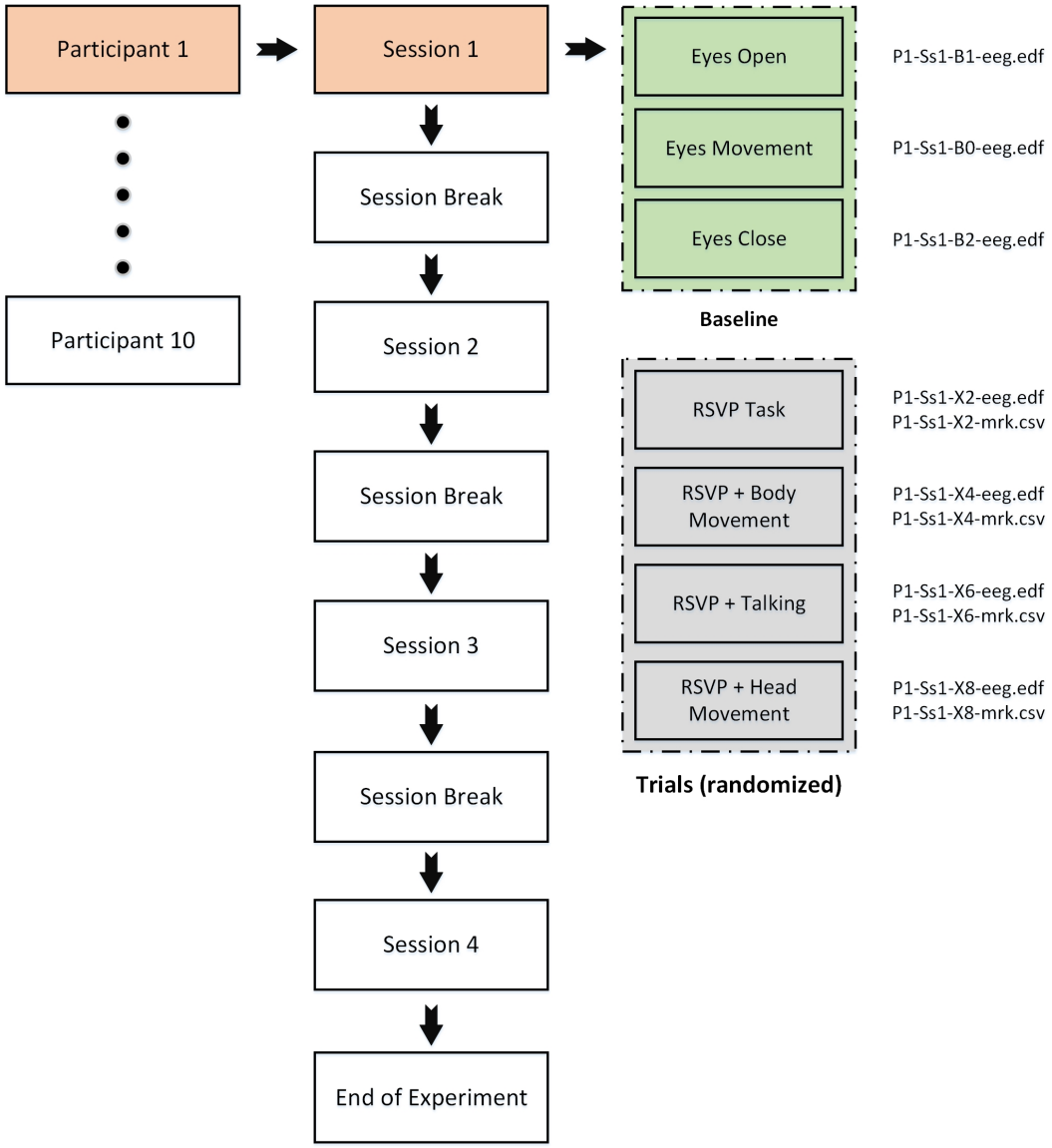


Figure 3.5: AMBER 2.0: organisation of the EEG recordings for each participant

3.3 Summary

As part of this work's core contributions, the AMBER and AMBER 2.0 datasets were systematically designed and collected to support the development of real-world BCI systems. These datasets provide comprehensive resources for advancing BCI research in naturalistic and ecologically valid environments. By addressing the limitations of lab-constrained datasets, these datasets aim to facilitate the development of robust and reliable BCIs capable of handling the complexities of everyday use.

AMBER relies on a traditional computer monitor display offering a consistent

and familiar visual environment, supporting multimodal analysis with video synchronisation, while AMBER 2.0 introduces immersive scenarios beyond the boundaries of a standard laboratory setting, through HMD-based RSVP tasks, marking significant strides toward practical BCI deployment.

Both datasets serve as benchmarks for evaluating BCI algorithms and exploring the interplay of environmental factors, task demands, and user performance. Together, they lay the groundwork for the development and validation of next-generation, noise-resilient applications in neurotechnology.

Chapter 4

The Impact of Noise on P300 Detection

BCIs could enable numerous opportunities in both clinical and non-clinical applications; however, deploying BCIs outside of controlled lab settings introduces challenges due to environmental noise and movement artifacts. Removing unwanted noise from EEG recordings would enable a broader application of BCIs in real-world settings, particularly in clinical and health-focused application spaces, as well as those in human-computer interaction, including Virtual Reality.

Traditional analysis methods have predominantly relied on cleanly recorded EEG data collected within controlled laboratory settings that aim to minimise noise. Nevertheless, a significant research gap persists in the investigation of the P300 ERP within real-world conditions. In real-world conditions, EEG data collection can occur in environments where complex and uncontrolled scenarios can generate significant noise in the EEG recording that hinders the detection and use of the P300 in applications such as BCIs. Mitigating this issue to allow effective detection of important ERP phenomena in EEG, such as the P300, is not only essential for expanding the understanding of cognitive processes in naturalistic environments but also for the development of applications in fields such as human-computer interaction.

In order to measure and understand the impact of different sources of movement-

related noise on P300 detection in a RSVP-BCI image search application, detection performance was assessed using both EEG data captured in clean recording conditions (i.e., where participants were instructed to withhold movements) and a variety of movement conditions where participants intentionally produced movements to contaminate the EEG recordings. This demonstrates variations in classification accuracy based on various systematically induced types of noise.

The primary objective of this research was to predict target versus non-target single-trial responses in a novel dataset, specifically, RSVP-based P300 datasets [248] [249], which included both clean and intentionally induced noisy recording conditions. The impact of noise and artifacts on detection performance has been systematically examined, highlighting their influence on prediction accuracy. In addition, the importance of using clean data for training machine learning models was investigated, along with the effects of utilising contaminated data on detection accuracy.

The key contributions of this chapter are as follows:

1. Assessing the impact that different noisy recording conditions (i.e., body movement, talking, head movement) have on single-trial P300 detection performance.

2. Investigating the generalisability of models trained on clean EEG data by evaluating their performance on artifact-contaminated test data.

3. Assessing the impact of noisy data when used to train machine learning models, and measuring this in terms of the performance of an RSVP-BCI using ROC-AUC.

This chapter addresses two key research questions focused on the robustness of RSVP-P300 BCI systems in the presence of real-world noise and behavioral variability. **Research Question 1 (RQ.1)** investigates how specific types of intentional behavioral artifacts namely, body movement, head movement, and talking, affect EEG signals and BCI performance. **Research Question 2 (RQ.2)** explores how

well BCI models trained on clean versus artifact-contaminated EEG data generalise to real-world conditions, and can training on noisy data yields comparable performance when tested on similarly noisy inputs. By systematically analyzing these questions, the chapter provides critical insights into the resilience and practical viability of traditional BCI systems beyond controlled laboratory settings.

4.1 Methodology

In this section, I briefly describe the dataset used, the preprocessing steps carried out, the analysis techniques employed, and the metric used for evaluation and comparison.

During EEG data acquisition, participants engaged in an RSVP task presented on a computer screen (in the AMBER dataset) and inside an HMD display (in the AMBER 2.0 dataset). Following data acquisition, preprocessing procedures were applied, and subsequent analyses involved target versus non-target prediction to examine the influence of noise on classification performance. Further elaboration on these methodological aspects is provided in the following subsections.

4.1.1 Dataset

The following two datasets were utilised in this study, the detailed descriptions of which can be found in Chapter 3:

1. AMBER: Advancing Multimodal Brain-Computer Interfaces for Enhanced Robustness – A Dataset for Naturalistic Settings [248]

   This dataset involves RSVP task presentations on a traditional monitor display. A detailed description is provided in Section 3.1.

2. AMBER 2.0: A Dataset for Naturalistic Settings with HMD-Based RSVP Tasks [249]

   This dataset involves RSVP task presentations using a head-mounted display (HMD). A detailed description is provided in Section 3.2.

4.1.2 Preprocessing

The preprocessing of EEG data is an essential step as it helps enhance the quality and interpretability of the recorded signals [238]. It involves a series of procedures aimed at removing noise and preparing the data for subsequent analysis. In this study, Python, along with the MNE library (version 1.6.0), was utilised for data preprocessing. The preprocessing steps in this study comprise the following steps:

1. Digital Filtering: To mitigate high-frequency noise, an Infinite Impulse Response (IIR) band-pass filter with cut-off frequencies of 0.1 Hz and 30 Hz was applied to the EEG data.

2. Re-sampling: The original data, initially sampled at 1000 Hz, was downsampled to 100 Hz. This reduction in sampling frequency allows for more manageable data processing while preserving critical temporal information.

3. Re-referencing: The EEG data was re-referenced from the initial CPz electrode to a common average reference, a technique that aids in reducing common-mode noise and enhancing signal quality [250].

4. Epoching: The continuous EEG recordings were segmented into epochs by extracting time series data from -0.2 to 0.8 seconds relative to the onset of the visual stimulus. This step is crucial for aligning the data with specific events of interest, facilitating subsequent analysis focused on event-related potentials and other cognitive processes.

4.1.3 Classification and Evaluation

The investigation in this chapter did not intend to surpass any existing single-trial P300 detection models. Instead, the classifier was merely a tool to enable systematic evaluation, where the primary objective of this research was to comprehensively examine the impact of noise and artifacts on classification performance for RSVP-P300 datasets.

Bayesian Ridge Regression:

Bayesian machine learning approaches have gained widespread acceptance among researchers in the field of Brain-Computer Interfaces due to their effectiveness in addressing classification challenges [251]–[254].

Bayesian Ridge Regression is a machine learning approach that combines Bayesian principles with ridge regression shrinks the coefficients to achieve greater numerical stability, resulting in improved computational accuracy [255]. It is particularly useful when dealing with datasets that have properties such as multicollinearity, i.e., where the input features are highly correlated [256].

In traditional regression analysis, model parameters are estimated based solely on the observed data. However, Bayesian ridge regression incorporates prior information about the parameters into the model, allowing more informed estimates. It assigns prior distributions to coefficients and combines them with a likelihood function to yield a posterior distribution through Bayes' theorem. It then estimates parameters based on this posterior distribution, and introduces regularisation to prevent overfitting [257], [258].

In this work, the Bayesian Ridge algorithm [259] [260] was used for single-trial P300 detection as similar to [163], [261], [262], for both IC-RSVP (noisy) and RSVP (clean) conditions using the AMBER dataset. Scikit-learn [263] version 1.3.2 was used for model training, hyperparameter tuning, and evaluation.

The default configuration for Bayesian ridge was employed. RandomizedSearchCV, a random search approach, was utilised to tune two alpha and two lambda parameters, optimising the model's performance.

Receiver Operating Characteristic - Area Under Curve:

Model performance was assessed using ROC-AUC (Receiver Operating Characteristic - Area Under Curve) [264]. This metric enables measurement of a model's discriminative ability.

ROC-AUC is insensitive to class imbalances, ensuring that the performance eval-

uation is not skewed by the dominance of one class over the other, making it particularly useful in scenarios where ranking or prioritisation is a key objective. For Rapid Serial Visual Presentation EEG research, ROC-AUC is a commonly employed evaluation metric [9], [265]–[267].

4.1.4 Analysis Procedure

In this study, P300-based Rapid Serial Visual Presentation trials from both datasets were categorized into two types: RSVP (i.e., traditional clean RSVP) and IC-RSVP (i.e., RSVP trials subjected to intentional artifact production using body movements, talking, and head movements).

The data (both traditional RSVP and intentionally contaminated IC-RSVP) were each further categorized into "clean" and "bad" on the basis of peak-to-peak thresholding, following the recommendation by Steven J. Luck [113], where a threshold between 50µV and 200µV is commonly employed for ERP analysis. This categorisation was made to investigate the impact of using clean, bad, or a combination of both types of data (clean + bad) on training and testing, ultimately influencing the classification performance. A 100µV threshold was applied, consistent with various studies [268] [269] [270] [271].

As this study explores P300 detection to understand how noise affects detection performance, a classifier was trained using clean trials from traditional RSVP, serving as a baseline for understanding the influence of noise on prediction outcomes. After training the model on clean trials, the classifier was used to test on both bad trials and a combination of clean and bad trials from the same RSVP dataset. In the case of artifact-laden RSVPs (i.e., IC-RSVP), predictions were conducted on all trials (clean and bad), only bad trials, and only clean trials for each of the three different artifact-producing behaviors (body movement, talking, and head movement).

An overview of the total number of trials, the percentage of retained trials (hereafter, referred to as Rt%), and the percentage of dropped trials (hereafter, referred to as Dr%) resulting from the rejection procedure using the peak-peak threshold

(as presented in Table 4.1 and 4.2) is crucial in understanding the size of the dataset used for training and testing the model. For instance, in subject 1 (AMBER dataset), out of a total of 2880 trials, 73.30% (2111 trials) were retained as clean trials, while 26.70% (769 trials) were identified as bad trials. With a sampling rate of 100 and an epoch size of 1 second across 32 channels, the model was trained on the clean trials, equivalent to a dataset size of 2111x3200, and tested on the bad trials, amounting to 769x3200.

Next, to highlight the significance of training models on clean trials from traditional RSVP and to assess how detection accuracies are affected when intentionally contaminated RSVPs are used for training, a comparative analysis was conducted. This involved employing different strategies to train and test models on both RSVP and IC-RSVP datasets, aiming to evaluate detection performance under various conditions.

1. **Condition 1:** Training and testing on clean trials from the RSVP dataset, i.e., **train** (*clean* RSVP) and **test** (*clean* RSVP). Additionally, for comparative analysis, training on clean trials from the IC-RSVP dataset was performed, and the testing was executed using the same subset from RSVP, i.e., **train** (*clean* IC-RSVP) and **test** (*clean* RSVP).

   In IC-RSVP, "clean trials" refer to those that are below the peak-to-peak rejection threshold, and not necessarily completely free from contamination.

2. **Condition 2:** Training and testing on the combination of clean+bad trials from the RSVP dataset, i.e., **train** (*clean+bad* RSVP) and **test** (*clean+bad* RSVP) [1]. Additionally, for comparative analysis, training on the combination of clean+bad trials from the IC-RSVP dataset was performed, and the testing was executed using the same subset from RSVP, i.e., **train** (*clean+bad* IC-RSVP) and **test** (*clean+bad* RSVP).

3. **Condition 3:** Training on clean trials from the IC-RSVP dataset and testing

[1] Although participants were instructed to minimise behaviors that would induce noise in the EEG recording during the RSVP condition, some trials containing artifacts were still present.

it on the subset within the same category, i.e., **train** (*clean* IC-RSVP) and **test** (*clean* IC-RSVP). Additionally, to explore the impact of bad trials within intentionally contaminated RSVP, training on a combination of clean and bad trials from the IC-RSVP dataset was performed, and the testing phase was then executed on the subset within the same category, i.e., **train** (*clean+bad* IC-RSVP) and **test** (*clean+bad* IC-RSVP).

The objective was to assess the impact concerning choices when training models on various trial combinations from intentionally contaminated RSVPs and traditional RSVPs. I maintained consistency across all training models by using an equal number of trials for training (1080 trials) and the remaining trials for testing purposes. Models were trained and evaluated on a participant-per-participant basis.

Given the utilisation of a 100µV threshold for trial rejection, further analysis was conducted by systematically varying the threshold from 50µV to 500µV. This analysis serves as the concluding exploration, where the objective was to observe the impact of threshold variations on the percentage of rejected trials, as this factor significantly influences the training and testing phases of the classification.

4.2 Results and Discussion

Figure 4.1 presents target versus standard averaged epochs for two different subjects at channel Pz. The first column shows examples from subject 4 (AMBER dataset), while the second column shows data from subject 8 (AMBER 2.0 dataset). The first row shows averaged epochs in a traditional lab setting, while the subsequent three rows show averaged epochs from RSVP contaminated with talking, body movements, and head movements, respectively.

Figure 4.2 shows grand averages across all subjects at channel Pz and compares the traditional RSVP with intentionally contaminated RSVP, illustrating all artifacts combined, as well as each artifact separately, all within a single figure. In this representation, dotted lines represent the grand average of standard epochs, while

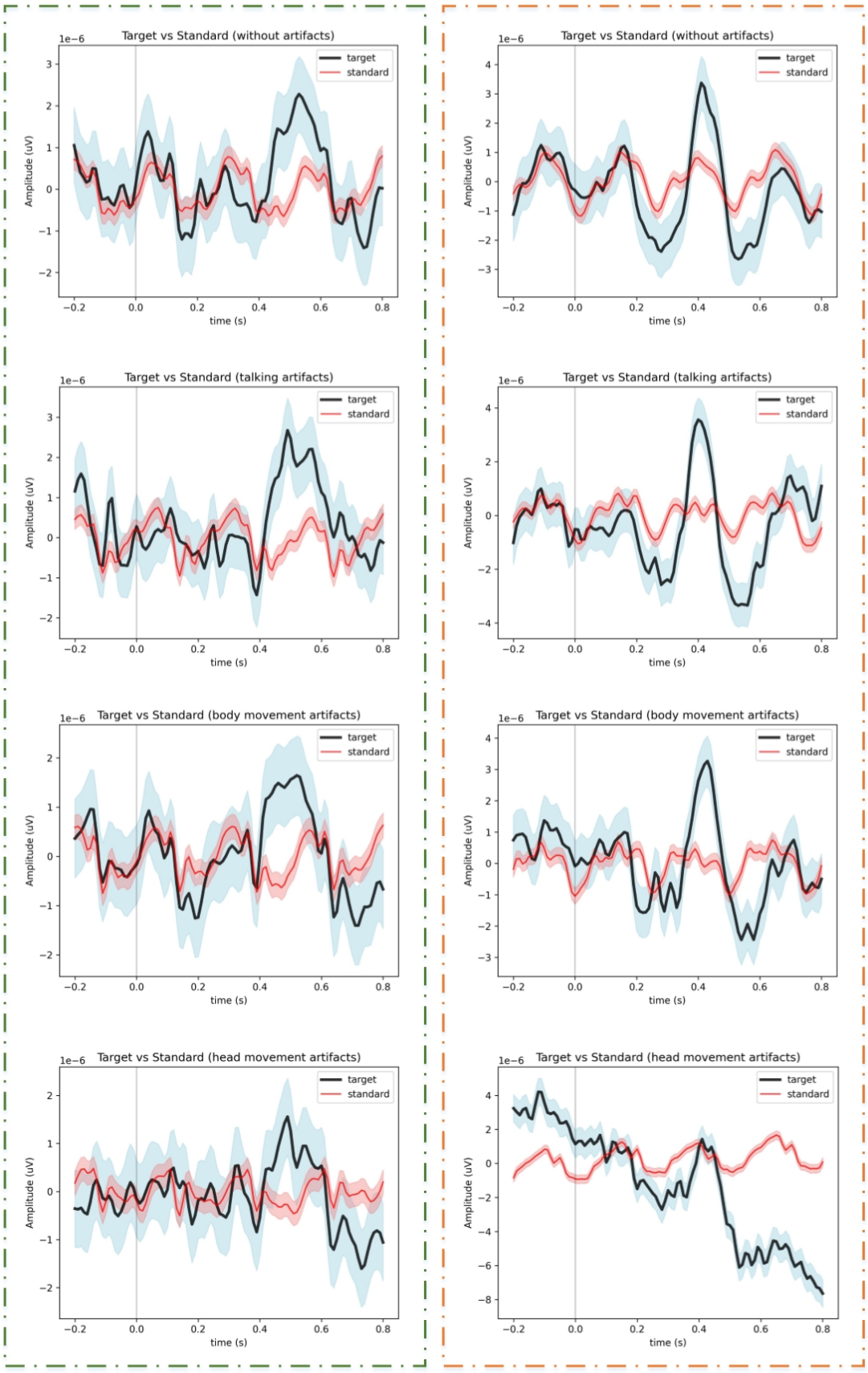


Figure 4.1: Visualisation of averaged epochs at channel Pz for RSVP, talking, body movement, and head movement conditions (top to bottom). The first column shows examples from subject 4 (AMBER dataset), while the second column shows data from subject 8 (AMBER 2.0 dataset).

the solid lines represent the grand average of target epochs. The high amplitude of noise caused by head movement in the AMBER dataset (Figure 4.2 (a)) made it impractical to display the comparison in the combined figure. Therefore, the IC-RSVP (head movements) condition was scaled down to fit within the comparison.

In order to visually analyze it further in depth, Figure 4.3 presents butterfly plots using all the electrode channels instead of only Pz as used in Figures 4.1 and 4.2. The butterfly plots show the difference between the grand averages of target epochs and standard epochs across all blocks. In addition, topographic analysis has also been made part of the butterfly plot for better understanding. Characteristic P3b activity can be seen at posterior scalp sites approximately between 300ms and 600ms following target detection; however, it cannot be as readily identified in the topographic maps of intentionally contaminated RSVPs.

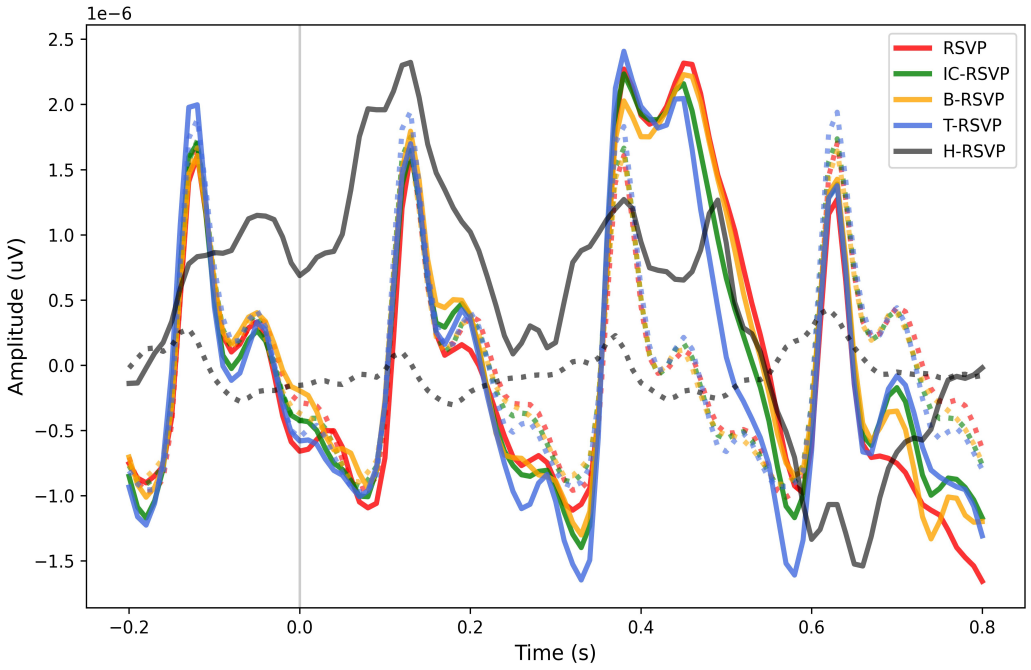
From these figures, it becomes clear that studying the effect of artifacts on P300 detection solely through visual inspection of the averaged epochs is challenging. Nevertheless, the impact is clearly visible in topographic plots related to the intentionally contaminated RSVPs, specifically of the head movement artifacts, i.e., Figure 4.3.

4.2.1 Bad Trials Rejection

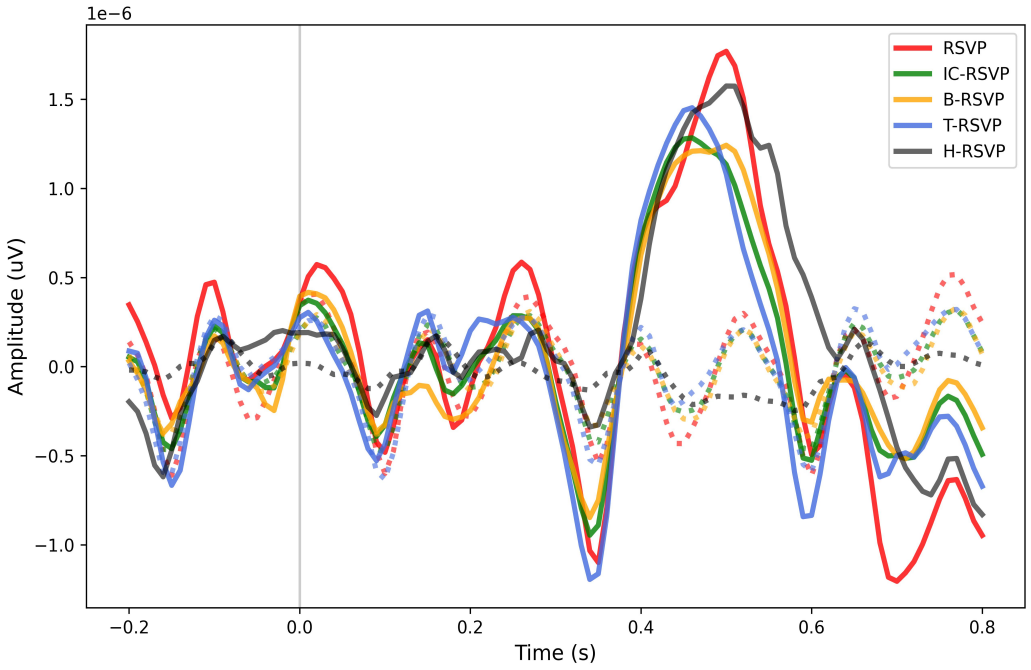
The initial step in the analysis involved rejecting bad trials using a 100µV peak-to-peak threshold. Table 4.1 (AMBER dataset) and 4.2 (AMBER 2.0 dataset) detail the outcomes of the bad trials rejection process, providing information on total trials, the percentage of retained trials after rejection (i.e., clean trials), and the percentage of dropped trials (i.e., bad trials). In these tables, the percentage of retained trials is referred to as Rt% and the percentage of dropped trials is referred to as Dr%.

AMBER dataset

For the RSVP condition, a total of 2880 trials were recorded. In contrast, for IC-RSVP, each artifact category for each participant had 1440 trials. Table 4.1 displays

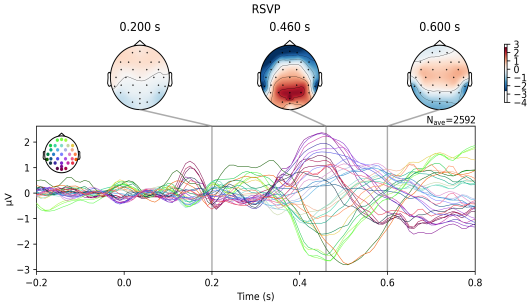


(a) AMBER dataset

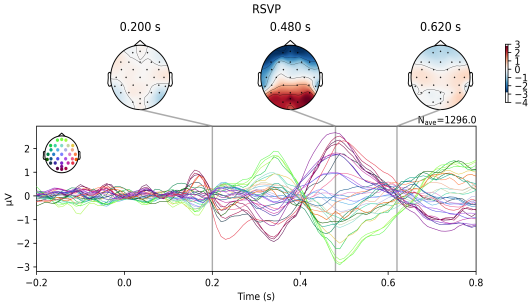


(b) AMBER 2.0 dataset

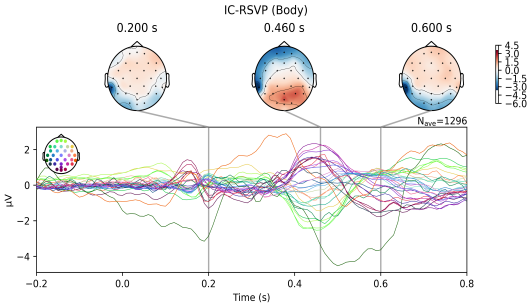
Figure 4.2: Target vs Standard grand average comparisons under different conditions at channel Pz. The solid lines on the plot represent target epochs, whereas the dotted lines refer to the standard epochs.
IC—intentionally contaminated, T—talking, B—body movement, H—head movement



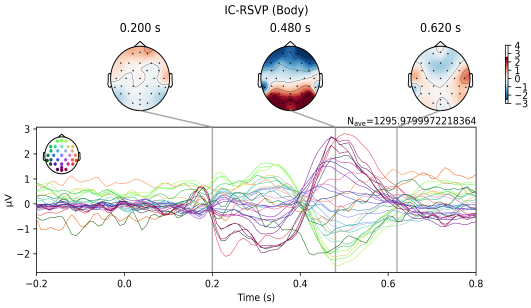
(a)Traditional RSVP



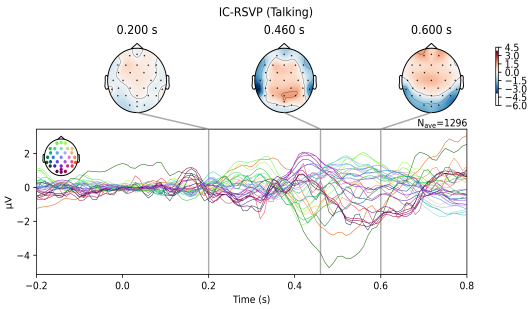
(b) Traditional RSVP



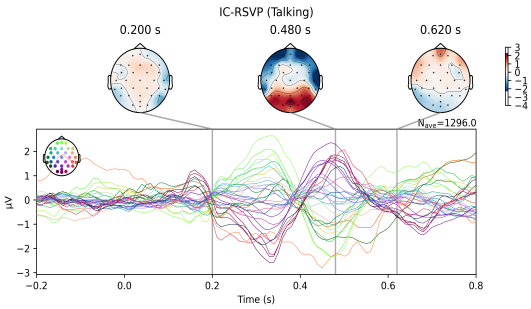
(c) IC-RSVP body movement)



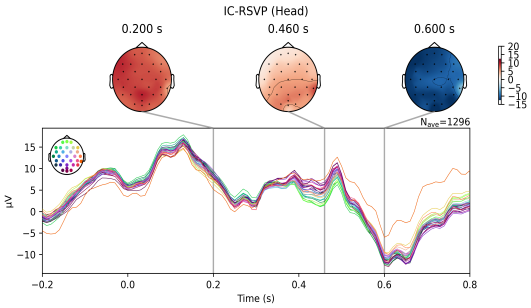
(d) IC-RSVP body movement)



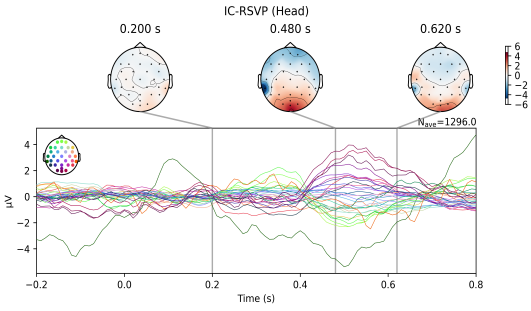
(e) IC-RSVP (talking)



(f) IC-RSVP (talking)



(g) IC-RSVP (head movement)



(h) IC-RSVP (head movement)

Figure 4.3: Butterfly plot (ERP averages) of target epochs minus averaged standard epochs across all blocks. The colors on time-series plots refer to the electrode locations on the scalp.
Column 1 – AMBER dataset , Column 2 – AMBER 2.0 dataset

Table 4.1: Trial rejection using peak-to-peak threshold of 100µV in AMBER dataset.

Subject	RSVP			IC-RSVP						
	Total Trials	Trials Rt (%)	Trials Dr (%)	Total Trials	Body Movement		Talking		Head Movement	
					Trials Rt (%)	Trials Dr (%)	Trials Rt (%)	Trials Dr (%)	Trials Rt (%)	Trials Dr (%)
1	2880	73.30	26.70	1440	13.61	86.39	53.96	46.04	1.25	98.75
2	2880	95.66	4.34	1440	91.25	8.75	94.93	5.07	60.97	39.03
3	2880	77.26	22.74	1440	55.69	44.31	81.53	18.47	0.14	99.86
4	2880	87.81	12.19	1440	90.76	9.24	95.90	4.10	0	100
5	2880	99.31	0.69	1440	98.75	1.25	95.35	4.65	0.42	99.58
6	2880	97.05	2.95	1440	71.81	28.19	15.07	84.93	0.07	99.93
7	2880	90.90	9.10	1440	44.65	55.35	86.60	13.40	5.97	94.03
8	2880	77.85	22.15	1440	68.19	31.81	58.06	41.94	2.92	97.08
9	2880	97.95	2.05	1440	97.15	2.85	98.33	1.67	27.85	72.15
10	2880	93.58	6.42	1440	95.07	4.93	33.26	66.74	0	100
Average	2880	89.07	10.93	1440	72.69	27.31	71.30	28.70	9.95	90.04

Trials Rt (%)—Percentage of trials retained, and Trials Dr (%)—Percentage of trials dropped RSVP—traditional RSVP, and IC-RSVP—intentionally contaminated RSVP

subject-wise trial rejections, revealing that subject 1 had the highest number of trials rejected for RSVP and IC-RSVP with body movement. However, for talking, subject 6 exhibited the most rejections, while for the head movement category, subjects 4 and 10 had all trials rejected.

IC-RSVP with head movement showed the highest trial rejections, averaging 90.04% across all 10 subjects. Only one subject retained 61%, and another retained approximately 28% of the trials, while the remaining subjects had over 90% of the trials rejected. In contrast, the RSVP condition, owing to its non-artefactual nature, had the fewest trials rejected, averaging 91% retention. For IC-RSVP with body movement and talking, an average of 72.7% and 71.3% of trials were retained, respectively.

AMBER 2.0 dataset

For both RSVP and IC-RSVP conditions in AMBER 2.0, a total of 1440 trials were recorded. Table 4.2 displays subject-wise trial rejections, revealing that subject 5 had the highest number of trials rejected for RSVP and IC-RSVP with talking. However, for body movement, subject 9 exhibited the most rejections, while for the head movement category, subjects 1 and 3 had almost all trials rejected.

IC-RSVP with head movement showed the highest trial rejections, averaging

Table 4.2: Trial rejection using peak-to-peak threshold of 100µV in AMBER 2.0 dataset.

Subject	RSVP			IC-RSVP						
	Total Trials	Trials Rt (%)	Trials Dr (%)	Total Trials	Body Movement		Talking		Head Movement	
					Trials Rt (%)	Trials Dr (%)	Trials Rt (%)	Trials Dr (%)	Trials Rt (%)	Trials Dr (%)
1	1440	94.72	5.28	1440	93.19	6.81	90.69	9.31	2.15	97.85
2	1440	98.61	1.39	1440	93.54	6.46	76.04	23.96	13.54	86.46
3	1440	96.25	3.75	1440	86.81	13.19	84.17	15.83	1.67	98.33
4	1440	91.04	8.96	1440	71.51	28.49	92.22	7.78	16.74	83.26
5	1440	88.54	11.46	1440	69.17	30.83	71.60	28.4	19.38	80.62
6	1440	94.65	5.35	1440	70.14	29.86	90.00	10.00	30.62	69.38
7	1440	96.11	3.89	1440	78.12	21.88	100	0	31.94	68.06
8	1440	96.25	3.75	1440	87.43	12.57	95.42	4.58	74.51	25.49
9	1440	97.71	2.29	1440	37.36	62.64	77.08	22.92	31.39	68.61
10	1440	99.44	0.56	1440	85.42	14.58	86.81	13.19	29.79	70.21
Average	1440	95.33	4.67	1440	77.27	22.73	86.40	13.60	25.17	74.83

Trials Rt (%)—Percentage of trials retained, and Trials Dr (%)—Percentage of trials dropped RSVP—traditional RSVP, and IC-RSVP—intentionally contaminated RSVP

74.83% across all 10 subjects. Only one subject retained 74%, while the remaining subjects had over 68% of trials rejected. In contrast, the RSVP condition, owing to its non-artefactual nature, had the fewest trials rejected, averaging 95.33% retention. For IC-RSVP with body movement and talking, an average of 77.27% and 86.40% of trials were retained, respectively.

These results from both datasets highlight the substantial impact of artifact-producing movement, particularly head movement, on the quality of trials. The observed trial rejection rates highlight the challenges posed by such artifacts and emphasize the need for robust preprocessing techniques in P300-based studies involving real-world scenarios.

4.2.2 P300 Detection

After performing trial rejection to categorize trials into clean and bad, single-trial P300 detection was carried out by training only on clean trials from RSVP, as it serves as a baseline for understanding the influence of noise on prediction outcomes.

Considering the percentage of dropped trials outlined in Table 4.1, it's noteworthy that for some subjects, prediction was not feasible using clean trials due to insufficient data for testing after trial rejection. For instance, as shown in Table 4.3,

Table 4.3: ROC-AUC scores (AMBER dataset). The model was trained on traditional RSVP (clean trials) and tested across different conditions from RSVP and IC-RSVP.

Subject	**RSVP**		**IC-RSVP**								
			Body Movement			**Talking**			**Head Movement**		
	Clean + Bad	**Bad**	**Clean + Bad**	**Clean**	**Bad**	**Clean + Bad**	**Clean**	**Bad**	**Clean + Bad**	**Clean**	**Bad**
1	0.66	0.61	0.62	*	0.62	0.68	0.74	0.64	0.53	*	0.53
2	0.94	0.68	0.90	0.92	0.69	0.88	0.89	0.69	0.85	0.87	0.81
3	0.77	0.69	0.79	0.84	0.75	0.83	0.87	0.67	0.57	*	0.57
4	0.83	0.77	0.87	0.89	0.72	0.87	0.89	0.46	0.62	*	0.62
5	0.91	0.92	0.91	0.92	0.71	0.88	0.88	0.83	0.57	*	0.57
6	0.89	0.69	0.77	0.79	0.73	0.76	*	0.76	0.63	*	0.63
7	0.85	0.62	0.73	0.85	0.65	0.79	0.80	0.73	0.58	*	0.58
8	0.77	0.56	0.74	0.79	0.66	0.71	0.75	0.66	0.66	*	0.65
9	0.89	0.62	0.90	0.90	0.71	0.90	0.90	0.57	0.74	0.76	0.74
10	0.95	0.69	0.85	0.86	0.66	0.75	0.89	0.70	0.56	*	0.56
Average	0.85	0.69	0.81	0.86	0.69	0.81	0.85	0.67	0.63	0.82	0.63

* There were insufficient clean trials available to perform the prediction.
Given that the model was exclusively trained on clean trials from traditional RSVP, the table does not include results for the testing scenario using clean trials from RSVP.

no predictions could be made for Subject 1 under the body movement condition, and for Subject 6 under the talking condition, where more than 80% of the trials were discarded, leaving insufficient data for testing.

Notably, head movement IC-RSVP tasks posed additional challenges, limiting predictions to only two subjects (i.e., 2 and 9) from the AMBER dataset, as indicated in Table 4.3. Similarly, in the AMBER 2.0 dataset, clean predictions under head movement were only possible for five subjects, as shown in Table 4.4. This highlights the challenges of conducting accurate predictions in real-world scenarios, where noise and artifacts can significantly impact data availability and model performance.

AMBER dataset

The results, presented in Table 4.3 as ROC-AUC scores, provide an overview of detection performance under various conditions in the AMBER dataset.

When tested on a combination of clean and bad trials from RSVP, the model exhibited an average ROC-AUC score of 0.85, reflecting its ability to distinguish target and non-target stimuli effectively. However, when subjected to predictions

exclusively on bad trials within the RSVP, the average ROC-AUC score decreased to 0.69, highlighting the impact of noise on detection accuracy, even in controlled settings.

Expanding the analysis to trials from artifact-laden RSVP tasks (i.e., IC-RSVP), predictions on clean trials from body movement artifacts demonstrated an average ROC-AUC score of 0.86. However, introducing bad trials alongside clean trials in the same category slightly reduced the average ROC-AUC score to 0.81, indicating the challenges associated with mitigating noise effects during predictions in real-world scenarios. Similarly, for talking artifacts, the model achieved an average ROC-AUC score of 0.85 when tested on clean trials. Yet, introducing bad trials alongside clean trials in the same category led to a marginal decrease in the average ROC-AUC score to 0.81, which highlights the importance of addressing verbal interactions as potential noise sources in P300 detection.

The analysis of head movement artifact IC-RSVPs revealed distinct challenges. Predictions on clean trials achieved an average ROC-AUC score of 0.82 for the limited subset of subjects with sufficient data. However, when considering bad trials from all subjects, the average ROC-AUC score dropped to 0.63, emphasizing the challenges in handling intentional head movements as noise sources.

In Figure 4.4, boxplots are used to visually compare ROC-AUC scores obtained when the model trained on clean trials from traditional RSVP is tested across various conditions.

AMBER 2.0 dataset

The results, presented in Table 4.4 as ROC-AUC scores, provide an overview of detection performance under various conditions in the AMBER 2.0 dataset.

When tested on a combination of clean and bad trials from RSVP, the model exhibited an average ROC-AUC score of 0.78, reflecting its ability to distinguish target and non-target stimuli effectively. However, when subjected to predictions exclusively on bad trials within the RSVP, the average ROC-AUC score decreased

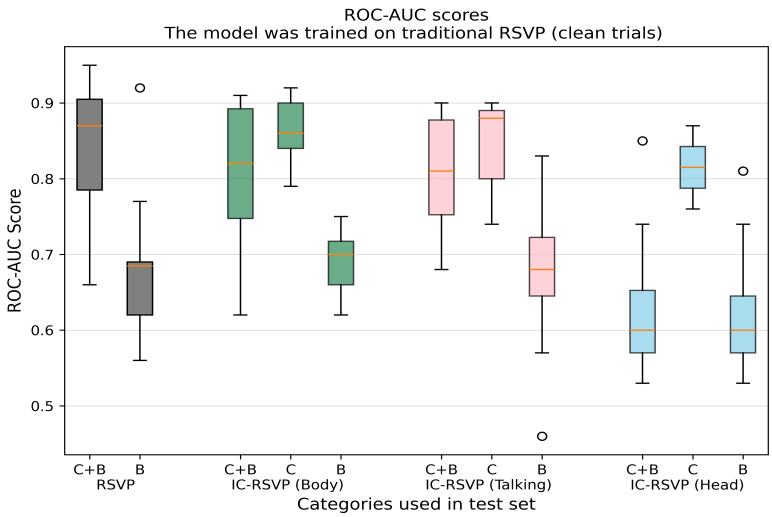


Figure 4.4: Box plots of ROC-AUC scores across subjects obtained when the model was trained on the clean trials from traditional RSVP (AMBER dataset). C—Clean trials, B—Bad trials and C+B—Clean+Bad trials

to 0.65, highlighting the impact of noise on detection accuracy, even in controlled settings. Moreover, when performing predictions on bad trials, there were insufficient data available for some subjects, specifically Subjects 2, 3, and 10, due to a low number of rejected trials, which limited the analysis in these cases.

Expanding the analysis to trials from artifact-laden RSVP tasks (i.e., IC-RSVP), predictions on clean trials from body movement artifacts demonstrated an average ROC-AUC score of 0.75. However, introducing bad trials alongside clean trials in the same category slightly reduced the average ROC-AUC score to 0.73, indicating the challenges associated with mitigating noise effects during predictions in real-world scenarios. Similarly, for talking artifacts, the model achieved an average ROC-AUC score of 0.76 when tested on clean trials. However, introducing bad trials alongside clean ones resulted in a slight decrease in performance, with the average ROC-AUC score dropping to 0.75.

The analysis of head movement artifact IC-RSVPs revealed distinct challenges. Predictions on clean trials achieved an average ROC-AUC score of 0.66 for the limited subset of subjects with sufficient data. However, when considering bad trials from all subjects, the average ROC-AUC score dropped to 0.61, emphasizing the challenges in handling intentional head movements as noise sources.

Table 4.4: ROC-AUC scores (AMBER 2.0 dataset). The model was trained on traditional RSVP (clean trials) and tested across different conditions from RSVP and IC-RSVP.

Subject	**RSVP**		**IC-RSVP**								
			Body Movement			**Talking**			**Head Movement**		
	Clean + Bad	**Bad**	**Clean + Bad**	**Clean**	**Bad**	**Clean + Bad**	**Clean**	**Bad**	**Clean + Bad**	**Clean**	**Bad**
1	0.80	0.73	0.83	0.84	0.67	0.82	0.83	0.72	0.59	*	0.59
2	0.91	*	0.79	0.80	0.64	0.74	0.77	0.67	0.58	*	0.57
3	0.75	*	0.69	0.72	0.56	0.72	0.73	0.68	0.61	*	0.62
4	0.82	0.60	0.74	0.74	0.79	0.83	0.84	0.69	0.66	*	0.66
5	0.70	0.67	0.65	0.68	0.59	0.72	0.74	0.68	0.60	*	0.60
6	0.56	0.60	0.61	0.62	0.59	0.61	0.60	0.65	0.52	0.64	0.48
7	0.72	0.46	0.66	0.66	0.65	0.72	0.72	*	0.59	0.64	0.58
8	0.84	0.59	0.76	0.78	0.61	0.78	0.79	0.56	0.69	0.71	0.64
9	0.87	0.87	0.80	0.87	0.77	0.76	0.77	0.73	0.68	0.69	0.68
10	0.83	*	0.76	0.82	0.54	0.8	0.82	0.73	0.60	0.63	0.58
Average	0.78	0.65	0.73	0.75	0.64	0.75	0.76	0.68	0.61	0.66	0.60

* There were insufficient trials available to perform the prediction.

In Figure 4.5, boxplots are used to visually compare ROC-AUC scores obtained when the model trained on clean trials from traditional RSVP is tested across various conditions.

Relationship between ROC-AUC Scores and % of dropped trials

The percentage of dropped trials, as outlined in Table 4.1, significantly impacts detection performance, as reflected in the ROC-AUC scores in Table 4.3 and 4.4. Figure 4.6 illustrates the inverse relationship between the percentage of dropped trials and the ROC-AUC scores achieved when tested with a combination of clean and bad trials from each category. A higher percentage of dropped trial results in lower ROC-AUC scores, and vice versa. The red color represents the ROC-AUC scores for traditional RSVP, while green, blue, and black represent the scores for body movement, talking, and head movement RSVPs, respectively.

To statistically assess this correlative relationship, the Pearson correlation coefficient and the p-value was calculated between the percentage of dropped trials and ROC-AUC scores showing a significant negative correlation resulting in ($r = -0.907$, $p = 7.924e-16$) for AMBER dataset and ($r = -0.663$, $p = 3.201e-06$) for AMBER

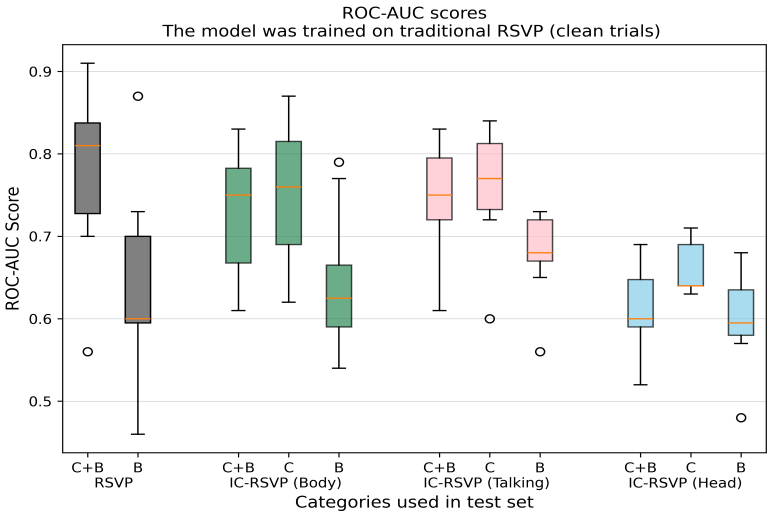


Figure 4.5: Box plots of ROC-AUC scores across subjects obtained when the model was trained on the clean trials from traditional RSVP (AMBER 2.0 dataset). C—Clean trials, B—Bad trials and C+B—Clean+Bad trials

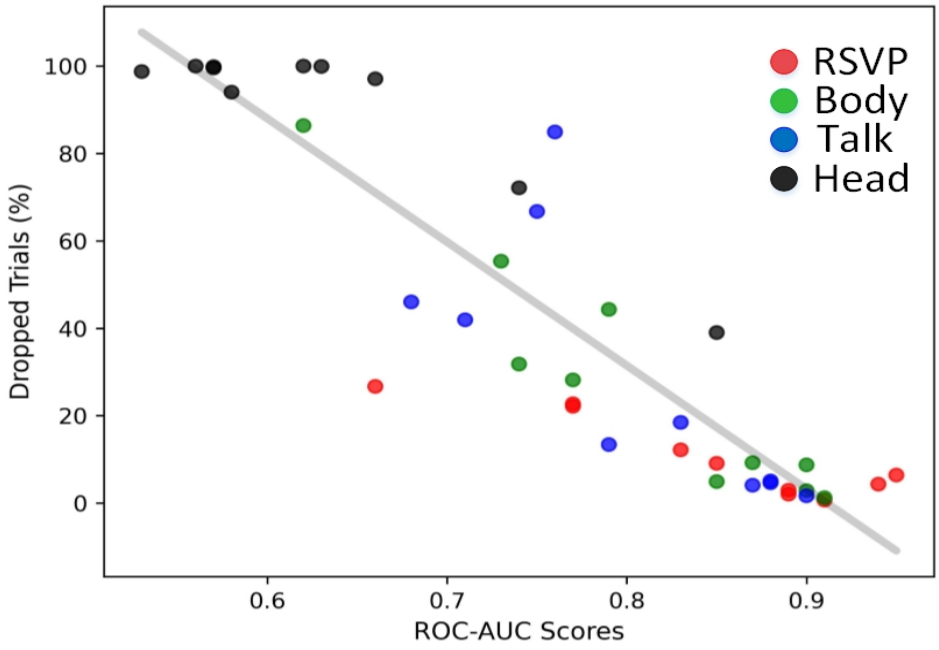
2.0 dataset.

As shown in Table 4.1, the percentage of dropped trials varied across subjects, resulting in differing training set sizes. To address this variability, the training size was standardized to 1595 trials for the AMBER dataset and 1000 trials for the AMBER 2.0 dataset, corresponding to the subject with the highest number of dropped trials. Predictions were then recomputed based on the standardized data, and the Pearson correlation coefficient and p-value were recalculated using these predictions and the percentage of dropped trials: (r = - 0.911, p = 3.738e - 16) for AMBER dataset and (r = - 0.678, p = 1.534e - 06) for AMBER 2.0 dataset.

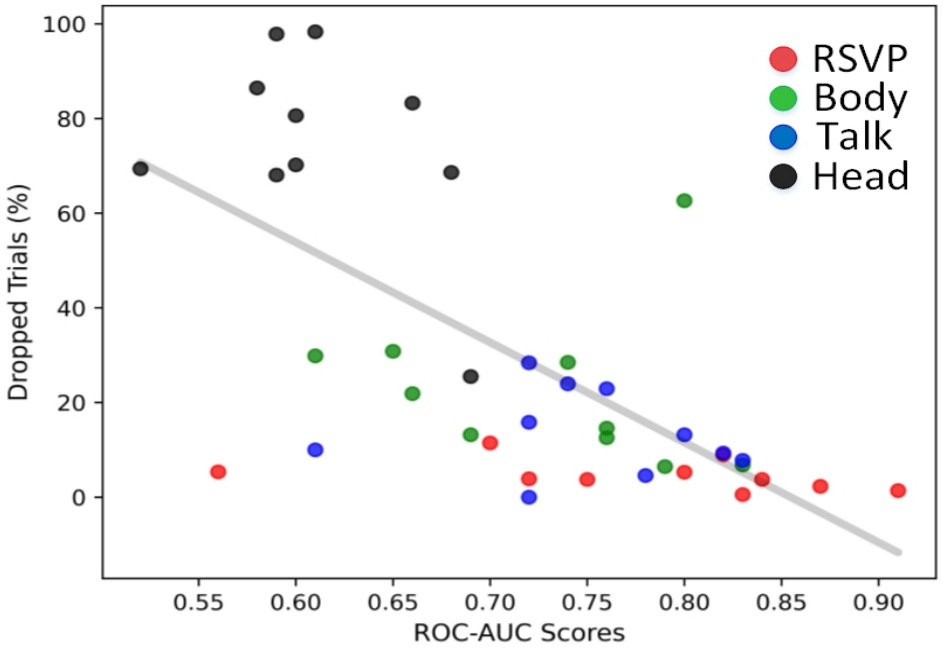
4.2.3 Impact of Bad Trials on Training and Testing Evaluation

To emphasize the importance of training models on clean trials from RSVP, a comparative analysis wa conducted. The outcomes of the investigation into various conditions, encompassing both RSVP and IC-RSVP used for training and testing, are presented in Table 4.5 (AMBER dataset) and 4.6 (AMBER 2.0 dataset).

The observations for each training and testing setup are detailed below:



(a) AMBER dataset



(b) AMBER 2.0 dataset

Figure 4.6: Scatterplot showing the relationship between the percentage of dropped trials and ROC-AUC scores. For each subject, the model was trained on clean trials from the traditional RSVP and tested on a combination of clean and bad trials from both traditional and intentionally contaminated RSVP conditions.

Table 4.5: ROC-AUC scores (AMBER dataset). The models were trained in different conditions, including only clean trials for training and then training the model using a combination of clean and bad trials.

Subject	Condition 1		Condition 2		Condition 3	
	Train: RSVP (Clean)	**Train: IC (Clean)**	**Train: RSVP (Clean+Bad)**	**Train: IC (Clean+Bad)**	**Train: IC (Clean)**	**Train: IC (Clean+Bad)**
	Test: RSVP (Clean)	**Test: RSVP (Clean)**	**Test: RSVP (Clean+Bad)**	**Test: RSVP (Clean+Bad)**	**Test: IC (Clean)**	**Test: IC (Clean+Bad)**
1	0.68	*	0.61	0.52	*	0.54
2	0.93	0.91	0.92	0.92	0.92	0.87
3	0.82	*	0.78	0.60	*	0.55
4	0.86	*	0.82	0.80	*	0.78
5	0.90	*	0.87	0.88	*	0.86
6	0.86	*	0.88	0.79	*	0.76
7	0.81	*	0.78	0.73	*	0.62
8	0.78	*	0.78	0.69	*	0.71
9	0.86	0.87	0.88	0.84	0.87	0.86
10	0.92	*	0.90	0.86	*	0.81
Average	0.84	0.89	0.82	0.76	0.90	0.74

RSVP—traditional RSVP, IC—intentionally contaminated RSVP
* There were insufficient clean trials available to train the model (i.e., more than 80% of the trials were dropped using a 100µV peak-to-peak threshold).

Table 4.6: ROC-AUC scores (AMBER 2.0 dataset). The models were trained in different conditions, including only clean trials for training and then training the model using a combination of clean and bad trials.

Subject	Condition 1		Condition 2		Condition 3	
	Train: RSVP (Clean)	**Train: IC (Clean)**	**Train: RSVP (Clean+Bad)**	**Train: IC (Clean+Bad)**	**Train: IC (Clean)**	**Train: IC (Clean+Bad)**
	Test: RSVP (Clean)	**Test: RSVP (Clean)**	**Test: RSVP (Clean+Bad)**	**Test: RSVP (Clean+Bad)**	**Test: IC (Clean)**	**Test: IC (Clean+Bad)**
1	0.89	*	0.89	0.84	*	0.78
2	0.88	0.85	0.9	0.81	0.82	0.64
3	0.83	*	0.73	0.68	*	0.68
4	0.84	0.83	0.83	0.75	0.85	0.78
5	0.79	0.74	0.82	0.62	0.73	0.57
6	0.61	0.61	0.68	0.59	0.66	0.61
7	0.84	0.76	0.82	0.71	0.68	0.65
8	0.79	0.8	0.82	0.72	0.76	0.73
9	0.88	0.85	0.85	0.83	0.78	0.77
10	0.88	0.87	0.91	0.89	0.82	0.79
Average	0.82	0.63	0.83	0.74	0.61	0.7

RSVP—traditional RSVP, IC—intentionally contaminated RSVP
* There were insufficient clean trials available to train the model (i.e., more than 80% of the trials were dropped using a 100µV peak-to-peak threshold).

1. **Condition 1** involved training the model with clean trials from RSVP and testing it on a subset from the same category, resulting in an average ROC-AUC score of 0.84 for the AMBER dataset. To investigate the impact of noise in training data, the model was trained on clean trials from IC-RSVP [2] and tested it on the clean trials from RSVP used in Condition 1. Due to the insufficient number of trials for training in all subjects except subjects 2 and 9 in the AMBER dataset, the analysis could not be extended. For these two subjects, the average ROC-AUC score was 0.89.

   In the AMBER 2.0 dataset, similar procedures were followed. The model was trained on clean trials from the traditional RSVP task and tested on the same category, achieving an average ROC-AUC score of 0.82. Next, the model was trained using clean IC-RSVP trials, and unlike AMBER, the AMBER 2.0 dataset allowed for a broader analysis, as sufficient trials were available for most subjects except Subjects 1 and 3. For the participants with sufficient trials, the average ROC-AUC score was 0.63.

   This decrease in performance in both datasets illustrates the difficulty of achieving reliable classification when training data is insufficient or affected by artifacts, despite the clean trials appearing to have minimal contamination.

2. **Condition 2** encompassed training the model with a combination of clean and bad trials from RSVP and testing it on the same category, resulting in an average ROC-AUC score of 0.82 and 0.83 for AMBER and AMBER 2.0 datasets, respectively. To explore the influence of noise in the training dataset, the model was then trained on clean and bad trials from IC-RSVP and tested on the clean trials from RSVP used in Condition 2. The average ROC-AUC score dropped to 0.76 (for AMBER) and 0.74 (for AMBER 2.0), showing a decline in performance.

[2] In the context of intentionally contaminated RSVP, "clean" trials refer to those below the peak-to-peak rejection threshold, rather than truly clean trials.

3. **Condition 3** examined model training using only the clean trials from IC-RSVP and testing on that same clean subset. In the AMBER dataset, this analysis was feasible for only two subjects, yielding an average ROC-AUC of 0.90, because the 100 µV peak-to-peak threshold led to the rejection of too many trials for the remaining eight subjects. In contrast, in the AMBER 2.0 dataset, sufficient clean trials were available for all but Subjects 1 and 3, resulting in an average ROC-AUC of 0.61.

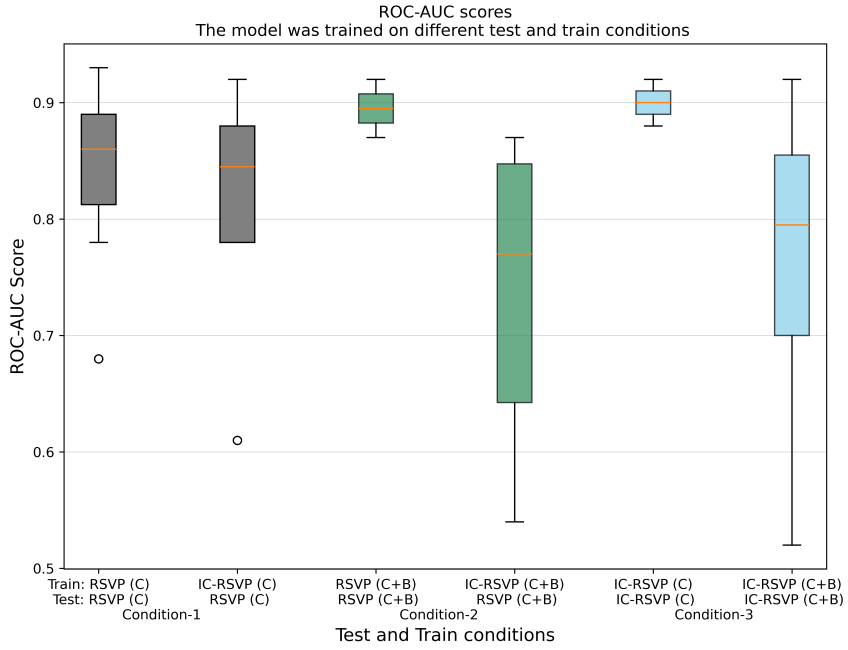
   When both clean and bad IC-RSVP trials were combined for training and tested on the corresponding combined subset, the average ROC-AUC scores dropped to 0.74 for AMBER and 0.70 for AMBER 2.0, highlighting the difficulty of training reliable models on artefactual data.

This finding aligns with the earlier discussions on the challenges posed by intentional artifacts. Connecting this analysis to the previous section of this chapter, where the model was exclusively trained on clean trials from RSVP, the comparison underlines the importance of using non-artifactual data for model training. The performance deterioration when the model was trained on data subjected to intentional artifacts, as compared to the ROC-AUC scores measured by training the model on non-intentionally contaminated data, signifies the importance of having clean and noise-free data for training purposes.

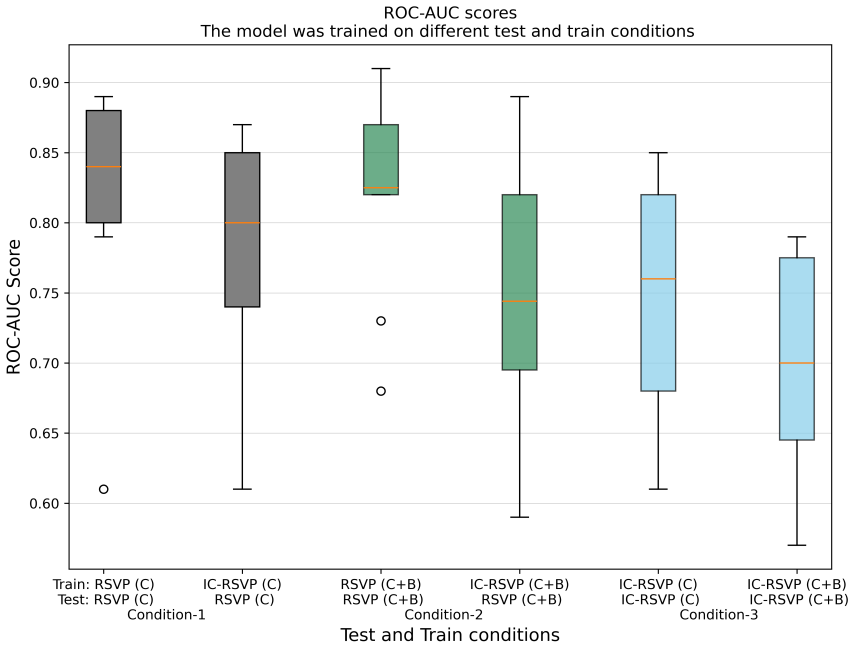
Figure 4.7 displays a graphical comparison of ROC-AUC scores derived from the three aforementioned conditions as boxplots.

4.2.4 Trials Rejection Using Different Peak-to-Peak Thresholds

Following the bad trial rejection at the 100µV threshold and the subsequent prediction analysis performed using this rejection threshold, an extended version of Table 4.1, denoted as Table 4.7 and 4.8, has been included. These tables provide insights into the percentage of dropped trials, presenting both average and standard



(a) AMBER dataset



(b) AMBER 2.0 dataset

Figure 4.7: Box plots of ROC-AUC scores obtained when the model was trained and tested on different conditions incorporating both traditional RSVP and IC-RSVP. C—Clean trials and C+B—Clean+Bad trials

deviation values. The rejection thresholds were systematically varied from 50µV to 500µV to cover the range commonly applied in related studies, with particular emphasis on the frequently used 100µV threshold.

While this study did not set out to determine the optimal threshold for bad trial rejection, the focus was on utilising a previously established peak-to-peak threshold, namely 100µV, to conduct this analysis. The primary objective was to highlight the influence of artifacts on prediction accuracies, emphasizing the necessity of addressing these common types of noise that occur in real-world conditions. The aim is rooted in the broader goal of ensuring EEG data integrity beyond controlled laboratory environments.

4.3 Summary

Addressing **Research Question 1 (RQ.1)** and **Research Question 2 (RQ.2)**, this chapter comprehensively examined the impact of specific intentional behavioral artifacts on EEG signals and BCI performance. Additionally, it evaluated the generalisation capabilities of models trained on clean versus artifact-contaminated EEG data, with a particular focus on whether training on noisy inputs can support robust performance when tested under similarly noisy conditions.

To distinguish between clean and artifact-contaminated trials, a peak-to-peak amplitude threshold of 100 µV, commonly used in the literature [268] [269] [270] [271], was applied. This led to substantial trial rejection, especially in the head movement condition, with some subjects left with insufficient clean data. The Bayesian Ridge model was then trained exclusively on clean trials from traditional RSVP settings and subsequently tested on diverse subsets of the dataset, examining clean trials, bad trials, and a mix of clean and bad trials, each associated with different artifact condition categories. ROC-AUC scores were employed to evaluate classification performance, consistently demonstrating degraded detection in the presence of noisy data.

Further experiments investigated different training–testing conditions across RSVP

Table 4.7: Trial rejection in AMBER dataset using different peak-to-peak thresholds, showing percentage of rejected trials.

Threshold (µV)	RSVP		IC-RSVP					
			Body Movement		Talking		Head Movement	
	Avg (%)	SD (%)	Avg (%)	SD (%)	Avg (%)	SD (%)	Avg (%)	SD (%)
50	33.36	25.09	58.26	35.69	68.37	27.97	99.83	0.46
75	16.46	13.75	37.96	31.45	39.74	33.64	97.06	6.48
100	10.93	9.60	27.31	27.96	28.70	29.62	90.04	19.85
150	4.89	5.28	16.46	19.37	14.81	19.62	77.12	33.70
200	2.39	3.22	11.00	14.52	8.70	13.78	65.79	36.61
250	0.97	1.02	8.30	12.75	5.60	11.24	52.71	36.53
300	0.30	0.29	6.81	11.04	3.98	9.16	42.94	33.67
350	0.16	0.22	5.84	9.62	3.28	8.34	37.13	31.50
400	0.09	0.16	5.20	8.56	2.86	7.43	33.22	29.30
450	0.05	0.14	4.34	7.13	2.43	6.38	30.01	26.92
500	0.05	0.14	3.73	6.08	2.17	5.65	27.34	24.62

Avg—Average, and SD—Standard deviation

Table 4.8: Trial rejection in AMBER 2.0 dataset using different peak-to-peak thresholds, showing percentage of rejected trials.

Threshold (µV)	RSVP		IC-RSVP					
			Body Movement		Talking		Head Movement	
	Avg (%)	SD (%)	Avg (%)	SD (%)	Avg (%)	SD (%)	Avg (%)	SD (%)
50	22.15	14.28	68.33	23.88	56.67	22.79	97.15	5.29
75	7.75	4.58	38.43	22.94	24.00	12.77	86.29	15.26
100	4.67	3.35	22.73	16.74	13.60	9.13	74.83	20.80
150	1.92	1.99	8.87	8.50	6.85	5.74	56.24	25.17
200	0.90	0.99	4.10	4.35	3.35	3.22	43.35	25.12
250	0.51	0.60	2.08	2.39	1.62	1.71	33.58	22.42
300	0.36	0.51	1.13	1.53	1.04	1.19	25.97	20.34
350	0.26	0.48	0.60	0.98	0.77	1.01	20.83	18.23
400	0.15	0.35	0.33	0.60	0.63	0.86	16.88	16.69
450	0.10	0.31	0.21	0.40	0.53	0.80	13.83	14.94
500	0.07	0.22	0.13	0.29	0.45	0.72	10.77	13.40

Avg—Average, and SD—Standard deviation

and IC-RSVP modalities under clean and mixed-quality data regimes. While models trained and tested on clean data yielded the highest performance, including noisy trials in the training set did not always improve generalisability, particularly when the noise levels differed between training and testing phases. These findings underscore the complex trade-offs involved in noise-handling strategies. Statistical analyses using the Pearson correlation coefficient confirmed a significant relationship between the number of rejected trials and ROC-AUC scores, reinforcing the idea that signal quality is a critical factor in model performance.

Given that such artifacts are unavoidable in real-world BCI applications, this study emphasizes the need for robust preprocessing and adaptive model approaches that minimise information loss while maintaining performance. Overall, these findings highlight the foundational role of noise resilience in extending EEG-based BCI systems from the laboratory to more dynamic, everyday environments.

Chapter 5

Subject-Independent Benchmarking

Despite recent advancements in machine learning and signal processing techniques, training models for robust P300 classification continues to be a substantial challenge. One crucial obstacle lies in the variability of EEG signals across individuals. Factors such as differences in brain anatomy, cognitive state, electrode placement, and external noise significantly influence the recorded signals [106]. Consequently, machine learning models trained on subject-specific data often fail to generalise effectively to new users, limiting the scalability and usability of BCIs in real-world scenarios [178].

Traditionally, machine learning models for P300 classification are trained on subject-specific data. While this approach delivers high accuracy within the training set, it often results in models that are overly specialized to the specific characteristics of the training subject's EEG responses. Such models often exhibit poor performance when applied to data from new subjects due to inter-subject variability. This limitation poses a significant barrier to developing generalised BCIs capable of serving diverse populations without requiring extensive retraining for each user.

Most existing BCIs rely on subject-specific training protocols, requiring significant time and effort to calibrate for each new user. These protocols limit scal-

ability and hinder the widespread adoption of BCI technology. To address this, subject-independent classification methods have emerged as a promising solution [178], [200], [229]. By training models on data from multiple users, these methods aim to learn invariant features of P300 signals, enabling BCIs to generalise across individuals without extensive retraining. This approach not only minimises the need for lengthy calibration sessions but also makes BCIs more practical and accessible for real-world applications.

Subject-independent classification also aligns with the goal of building inclusive and adaptive technologies by enabling generalisation across different users. This study investigates the effectiveness of deep learning and other performant machine learning algorithms for subject-independent P300 classification in RSVP paradigms, highlighting their potential to improve classification accuracy and adaptability across individuals.

This work addresses **Research Question 3 (RQ.3)**, "Which advanced machine learning models are most effective in achieving subject-independent P300 detection robust to real-world noise?" by evaluating the robustness of EEG-based brain-computer interface systems under naturalistic conditions using a subject-independent classification framework.

5.1 Methodology

Single-trial classification of P300 responses elicited during RSVP tasks was performed using a leave-one-subject-out (LOSO) approach, where data from 19 subjects were used for training, and the remaining subject was held out for testing in each iteration, as shown in Figure 5.14. This strategy provides a realistic assessment of model generalisability across individuals, a critical requirement for practical BCI deployment in real-world, noisy environments.

The development of EEG decoding models has progressed from traditional machine learning approaches to advanced deep learning architectures, as comprehensively reviewed in Section 2.4. Early work relied on Bayesian methods and shallow

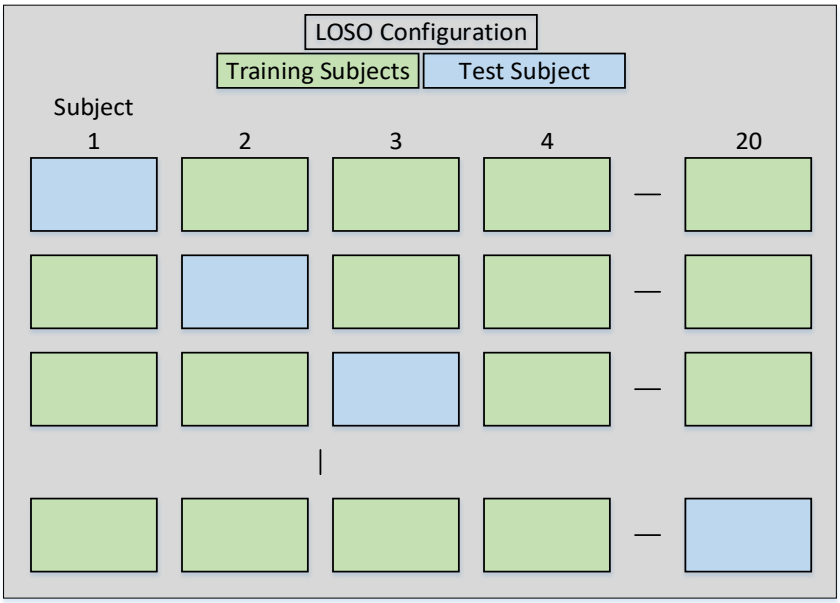


Figure 5.1: Leave-one-subject-out (LOSO) configuration

classifiers, which were later surpassed by convolutional neural networks that better capture spatiotemporal patterns in EEG signals. Recent advances have introduced Transformer-based models and their hybrids, combining the local feature extraction of CNNs with global attention mechanisms. These architectures have demonstrated superior performance in capturing both transient evoked potentials and long-range dependencies in EEG data [222], [227], [272].

This work evaluates both established and novel EEG decoding architectures, each carefully optimised for P300 classification performance. The investigation encompasses: (1) Bayesian methods [259], [260] as traditional ML baselines, (2) CNN-based architectures (CNN-1 [166], EEGNet [165]), (3) Transformer architectures, and (4) Transformer hybrids (EEGNet-Transformer [227], EEG Conformer [222]). While a few of the models were proposed in prior work, all implementations in this study were systematically tuned for my specific task through hyperparameter optimisation to act as suitable baselines.

Beyond these baseline adaptations, I further explored several architectural variants, including frequency-band specialized models and Mixture-of-Experts (MoE) extensions. The motivation for incorporating multi-band models stems from the

well-established neurophysiological principle that different EEG frequency bands capture distinct cognitive and perceptual processes. Processing these bands separately enables the model to learn frequency-specific features that are often obscured in full-band representations, thereby enhancing both the interpretability and precision of EEG decoding [93], [273]. Prior studies have shown that frequency decomposition can enhance classification performance in BCIs [231]–[233]. Leveraging this evidence, a multi-branch architecture was adopted to enable frequency-aware learning tailored to the spectral characteristics of brain activity.

Complementing this, the Mixture-of-Experts framework was employed to further enhance model adaptability and generalisation. MoE architectures allow the network to dynamically assign specialized "experts" to different subspaces of the input, enabling context-sensitive feature processing [237], [274]. Instead of forcing a one-size-fits-all representation, MoE lets the model learn diverse patterns that cater to different subjects or noise profiles. Notably, recent literature has demonstrated the effectiveness of MoE approaches in improving classification performance EEG EEG-based BCI applications [235], [275]. Building on this evidence, my design integrates MoE to enable flexible and robust decision-making across diverse recording conditions.

The overall methodology consists of three key stages, each detailed in subsequent sections. First, EEG data were collected from two experimental datasets, AMBER and AMBER 2.0, which differ in display modalities (see Section 5.2). Next, the raw EEG data underwent a standardized preprocessing pipeline to enhance signal quality and reduce noise (detailed in 5.3). Finally, evaluation of deep learning models for subject-independent P300 classification, with architectural details provided in Section 5.4.

5.2 Dataset

The dataset used in this study includes EEG recordings from a total of 20 participants, collected as part of two datasets: AMBER: Advancing Multimodal BCIs

for Enhanced Robustness – A Dataset for Naturalistic Settings [248], and AMBER 2.0: A Dataset for Naturalistic Settings with HMD-based RSVP Tasks [249]. These datasets are comprehensively described in Sections 3.1 and 3.2, respectively.

The AMBER dataset comprises data from the first 10 subjects who performed RSVP tasks using a conventional monitor display. In contrast, AMBER 2.0 contains data from another set of 10 subjects who engaged in the same RSVP paradigm but viewed the stimuli through a head-mounted display, allowing for an immersive and more naturalistic user experience.

This dataset was originally designed to evaluate BCIs under both clean and noisy conditions by incorporating RSVP-BCI tasks in naturalistic settings, where a protocol involving instructed movements was used in order to systematically contaminate particular blocks of trials.

For the purposes of this study, however, the training phase strictly utilised only the clean RSVP P300 trials, which were extracted from blocks where subjects remained stationary and no additional artifacts were introduced. These trials represent the ideal, noise-free condition and serve as the foundation for building baseline classification models. During the testing phase, model performance was evaluated not only on clean trials but also across all noisy conditions (including those affected by talking, head movement, and body movement). This approach enabled a detailed investigation into the generalisability and robustness of the models when faced with EEG data of varying quality and contamination levels.

5.3 Data Preparation

To ensure high-quality input for subsequent classification tasks, several preprocessing steps were applied to the EEG data. First, the raw EEG signals were band-pass filtered between 0.1 Hz and 30 Hz to remove slow drifts and high-frequency noise, preserving the frequency components relevant for P300-based analysis. Next, stimulus-locked epochs were extracted from the continuous EEG data, capturing a time window from 0 ms (stimulus onset) to 800 ms post-stimulus. This interval was

chosen to encompass the P300 component, which typically occurs between 250 ms and 600 ms.

To reduce computational complexity and improve model efficiency, the data were then downsampled to 50 samples per second (SPS). Following this, a common average reference was applied across all channels to minimise the influence of global noise and enhance the spatial resolution of the EEG signal. Finally, the data were normalised using min-max normalisation, which rescaled the input values to a standard range between 0 and 1. This normalisation step helped stabilize training and ensured consistent input for the machine learning models.

5.4 Classification

Single-trial subject-independent classification was performed using a leave-one-subject-out approach, evaluating a diverse range of models spanning traditional machine learning techniques to advanced neural architectures. The evaluated models included: (1) traditional machine learning approaches (e.g., Bayesian Ridge Regression [259], [260]); (2) convolutional neural networks such as CNN-1 [166] and EEGNet [165], which are widely used in EEG decoding due to their efficiency in capturing spatiotemporal features; (3) advanced transformer-based models, including simple transformer baselines, Multiband Transformer, and Mixture-of-Experts Transformer variants, which leverage attention mechanisms to capture long-range temporal dependencies; and (4) hybrid models that combine CNNs and attention layers, such as the EEG Conformer [222] and EEGNet Transformer variants [223], [227], which have demonstrated improved performance by integrating local feature extraction with global context modeling.

As described in Section 5.1, a few neural architectures, such as CNN-1, EEGNet, and Conformer, have been previously applied to EEG decoding tasks and have shown promising results across a range of BCI applications. Their inclusion in this study enables both replication of established baselines and benchmarking under real-world noise conditions. Additionally, multiband processing and Mixture-of-

Experts extensions were introduced to several architectures. The motivation behind these modifications stems from prior work highlighting the benefits of frequency-specific processing and ensemble modeling strategies (see Section 5.1). Multiband approaches enable the network to learn discriminative patterns within distinct EEG frequency bands (e.g., delta, theta, alpha, beta), which are known to carry task-relevant information. Similarly, MoE models enable dynamic specialisation by assigning different sub-networks (experts) to learn complementary features, thereby improving robustness and generalisation.

All models were trained for a maximum of 150 epochs with early stopping implemented to prevent overfitting and ensure computational efficiency. Training was terminated if the validation loss failed to improve for 20 consecutive epochs (patience=20), with the best weights being restored from the epoch with optimal validation performance. This approach not only maintained model generalisation but also automatically determined the optimal training duration for each subject-independent model during the leave-one-subject-out configuration.

Experiments were conducted using an NVIDIA GeForce RTX 3090 Ti GPU using the Keras (version 3.8.0) and TensorFlow (version 2.18.0) frameworks for deep learning based model development.

To guide readers through the various model architectures explored in this study, Table 5.1 presents a structured summary that includes each model's category, name, corresponding section number, and page number.

A detailed description of each model architecture and its implementation is provided in the following sub-sections.

5.4.1 Traditional ML

Bayesian Ridge Regression

Bayesian machine learning approaches have gained widespread acceptance among researchers in the field of Brain-Computer Interfaces due to their effectiveness in addressing classification challenges [251] [252] [253] [254].

Table 5.1: Directory of classification methodologies for rapid navigation

Model No.	Category	Model Name	Section	Page Number
CM-1	Traditional ML	Bayesian Ridge	5.4.1	106
CM-2	CNN-based Architectures	CNN-1	5.4.2	108
CM-3		EEGNet		108
CM-4		MoE-EEGNet		111
CM-5		Multiband EEGNet		111
CM-6		Multiband MOE EEGNet		112
CM-7	Transformers	EEGTransformer	5.4.3	114
CM-8		MOE-Transformer		115
CM-9	Hybrid Transformers	EEG Conformer	5.4.4	116
CM-10		CNN1-Transformer		118
CM-11		EEGNet-Transformer		119
CM-12		Multiband EEGNet-Transformer		121
CM-13		EEGNet-MoE Transformer		123
CM-14		Multiband EEGNet MoE-Transformer		124

Bayesian Ridge Regression is a machine learning approach that combines Bayesian principles with ridge regression shrinks the coefficients to achieve greater numerical stability, resulting in improved computational accuracy [255]. It is particularly useful when dealing with datasets that have properties such as multicollinearity, i.e., where the input features are highly correlated [256].

In traditional regression analysis, model parameters are estimated based solely on the observed data. However, Bayesian ridge regression incorporates prior information about the parameters into the model, allowing more informed estimates. It assigns prior distributions to coefficients and combines them with a likelihood function to yield a posterior distribution through Bayes' theorem. It then estimates parameters based on this posterior distribution, and introduces regularisation to prevent overfitting [257] [258].

In this work,the Bayesian Ridge algorithm [259] [260] is used for single-trial P300 detection, similar to [261] [262] [163]. Scikit-learn [263] version 1.3.2 was used for model training, hyperparameter tuning, and evaluation.

The default configuration for Bayesian ridge was employed. RandomizedSearchCV, a random search approach, was utilised to tune two alpha and two lambda parameters, optimising the model's performance.

5.4.2 Convolutional Neural Networks

CNN-1

In 2010, Cecotti et al. [166] introduced CNN-1, a 4-layer, 1D convolutional neural network designed for P300 detection. The model begins with a spatial convolution layer that learns linear combinations of input EEG channels, effectively acting as a spatial filter. This is followed by a temporal convolution layer, which extracts local temporal patterns while subsampling the feature maps to reduce dimensionality. Both convolutional layers employ the scaled hyperbolic tangent activation function ($1.7159 \cdot \tanh(2x/3)$), with layer weights initialized using the Cecotti-normal distribution.

A flattening layer transforms the 3D feature maps into a 1D vector, which is then processed by two dense layers with sigmoid activations. The first dense layer (100 units) performs nonlinear feature integration, while the final single-unit layer produces a probabilistic output for binary classification.

The model is trained using the Adam optimiser with binary cross-entropy loss. While the original architecture was largely retained in this study, hyperparameters were systematically tuned to optimise model performance. The final set of parameters selected after tuning is summarised in Appendix B.1.

EEGNet

Lawhern et al. [165] in 2018 proposed a compact and versatile convolutional neural network model specifically designed for BCI applications. This CNN model is effective for various tasks, including motor imagery classification [276], event-related potentials [277] and steady-state visual evoked potential analysis [278]. EEGNet employs separable and depthwise convolutions to create an EEG-specific model, in-

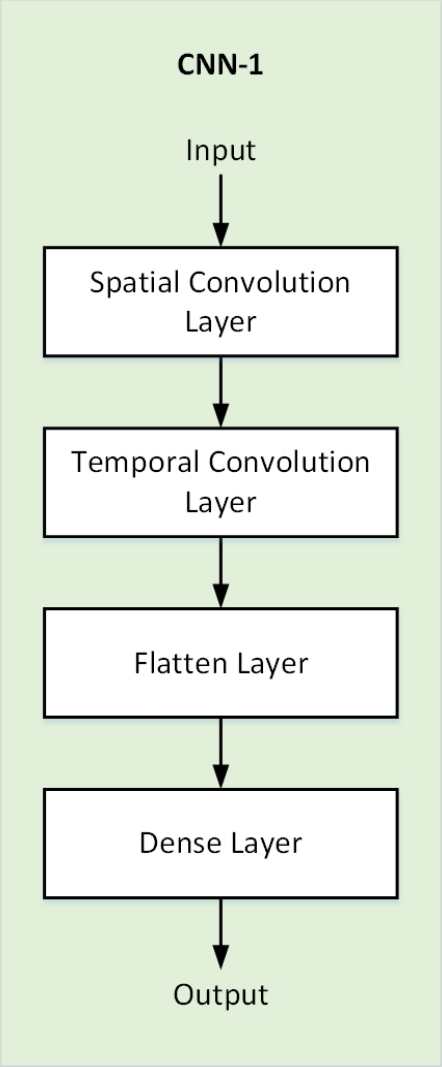


Figure 5.2: Network architectures: CNN-1 (based upon [166])

corporating established feature extraction principles for BCI. It is highly efficient, with significantly fewer trainable parameters compared to other deep learning architectures commonly used in BCI classification.

EEGNet employs a three-stage architecture that begins with a temporal convolution layer to process time-domain features, immediately followed by batch normalisation for stabilisation. This temporal block feeds into the second stage, a depthwise convolution layer that operates independently on each EEG channel to capture spatial relationships. Each depthwise operation is succeeded by batch normalisation, ELU activation for nonlinear transformation, average pooling for dimensionality reduction, and dropout for regularisation. The final stage employs separable convo-

lution to efficiently combine spatial and temporal features, followed again by batch normalisation, ELU activation, average pooling, and dropout. The architecture concludes with a classification block where 3D feature maps are flattened into a 1D vector and processed by dense layers for final decision-making.

While the core architecture of EEGNet was preserved, hyperparameter tuning was performed to optimise the performance. The final set of parameters selected after tuning is summarised in Appendix B.1.

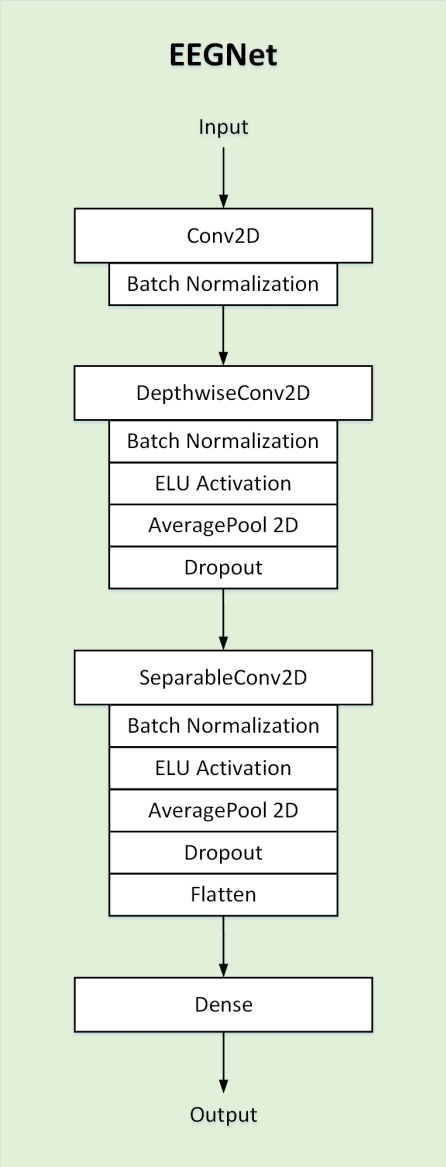


Figure 5.3: Network architectures: EEGNet (based upon [165])

MoE-EEGNet

The Mixture-of-Experts (MoE) architecture was first proposed in 1991 [279], implementing a divide-and-conquer approach to complex learning problems. This framework decomposes the input space into specialized regions, with each region handled by a dedicated expert network. A trainable gating network serves two critical functions: (1) dynamically selecting the most relevant experts for each input, and (2) combining their outputs through learned weighted aggregation [280]. This adaptive specialisation enables the model to maintain high representational capacity while preserving computational efficiency through sparse expert activation. MoE models have been used in EEG-based BCIs for applications like emotion recognition [235], ADHD prediction [234], and seizure detection [236].

In this work, a mixture-of-experts enhanced EEGNet architecture called MoE-EEGNet was proposed. The EEGNet backbone first processes raw EEG signals through its signature pipeline: (1) temporal convolution for time-domain patterns, (2) depthwise convolution for spatial filtering, and (3) separable convolution for joint spatiotemporal features.

The flattened features are then processed by the MoE layer, where a softmax-activated gating network allocates inputs to N parallel expert networks (ReLU-activated dense layers), whose outputs are weighted and aggregated.

The final prediction is generated through a sigmoid classification head. Critical hyperparameters like the number of experts and their dimensionality were systematically tuned to optimise the balance between model specialisation and generalisation capability. The final set of parameters selected after tuning is summarised in Appendix B.1.

Multiband EEGNet

Multiband EEGNet is introduced as an extension of EEGNet, incorporating explicit frequency-band specialisation to enhance performance. The model processes four canonical EEG frequency bands (delta, theta, alpha, beta) through dedicated

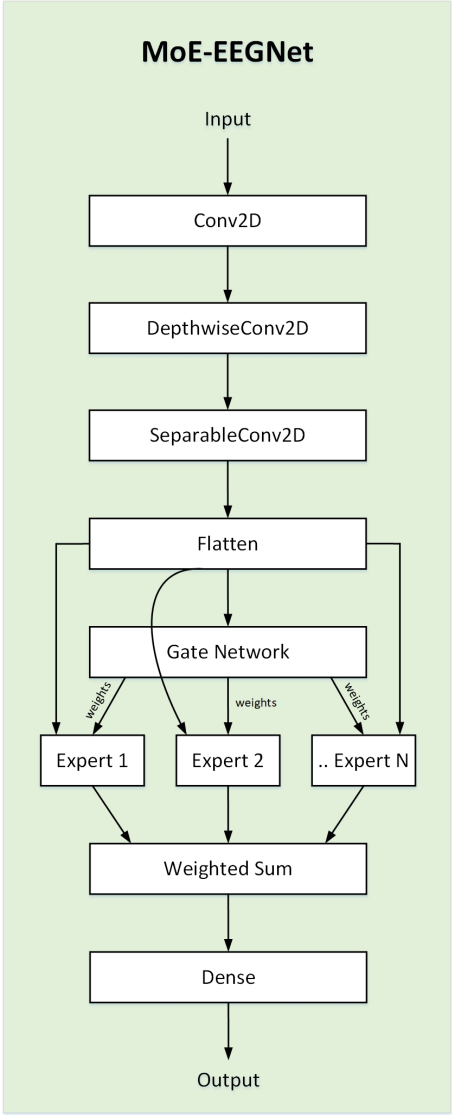


Figure 5.4: Network architectures: MoE-EEGNet

EEGNet subnetworks, each learning band-specific spatiotemporal features. The outputs of these parallel networks are concatenated and fed into a final classification layer, enabling explicit modeling of frequency-dependent patterns while maintaining EEGNet's core architectural benefits. The complete set of model hyperparameters is documented in Appendix B.1.

Multiband MoE-EEGNet

The multiband mixture-of-experts EEGNet (MB-MoE EEGNet) is introduced as a novel variant that synergistically integrates three powerful paradigms: (1) EEGNet's spatiotemporal feature extraction, (2) frequency-band specialisation, and (3)

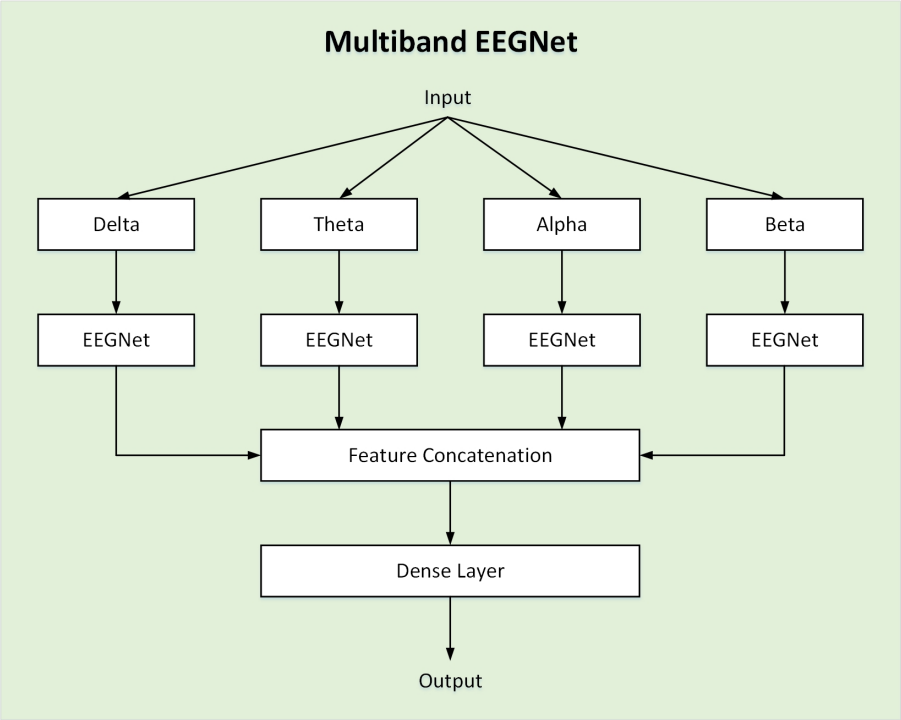


Figure 5.5: Network architectures: Multiband EEGNet

Mixture-of-Experts adaptive processing.

This architecture achieves hierarchical spectral-temporal modeling through four key stages. First, parallel EEGNet branches independently process delta, theta, alpha, and beta frequency bands. Second, each branch connects to a dedicated mixture-of-experts layer, where a softmax gating network dynamically routes features to specialized ELU-activated sub-networks (experts), enabling band-adaptive representation learning. Third, a multi-head attention mechanism fuses time-averaged outputs from all MoE layers, modeling cross-band interactions through learned attention weights. Finally, the attended features are aggregated and passed to a unified classification layer. Hyperparameter tuning was performed to optimise the model's performance. The complete set of model hyperparameters is documented in Appendix B.1.

5.4.3 Transformers

The Transformer architecture [281], originally developed for sequence-to-sequence tasks [282], has become foundational across machine learning domains. Following

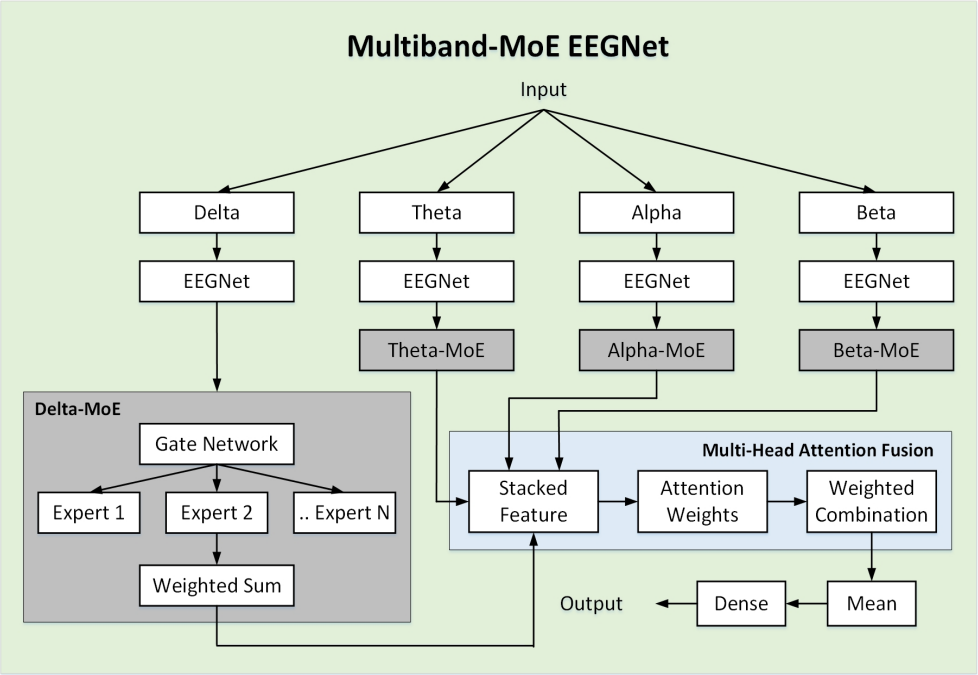


Figure 5.6: Network architectures: Multiband MoE-EEGNet

its success in natural language processing and computer vision, the paradigm has been adapted for temporal data processing in speech and biomedical applications [283]. Its self-attention mechanism is particularly suited for modeling long-range dependencies in time-series data [284], making it an ideal candidate for EEG analysis where precise temporal relationships between neural events are critical.

EEGTransformer

Building on these advancements in the field of deep learning, an EEG-specific Transformer architecture is proposed that adapts the self-attention mechanism to the unique challenges of neural time-series analysis. Unlike language or vision Transformers that process discrete tokens or image patches, my model treats individual timepoints as sequence elements while preserving channel-wise relationships through learned embeddings.

The proposed Transformer model processes EEG data by first reshaping the input EEG signal into a sequence format, where each time step across all channels is projected into a high-dimensional feature space via a linear embedding layer. To provide temporal context, a learnable positional encoding is added to these features,

informing the model of the sequence order.

This prepared sequence is then fed into the core of the architecture: a stack of Transformer Encoder layers. Within each encoder, a multi-head self-attention mechanism captures complex short- and long-range dependencies by weighing the relevance of different time steps, while a feed-forward network applies further non-linear transformations.

After passing through the encoder stack, the entire output sequence is aggregated into a single, fixed-size feature vector for the EEG trial using global average pooling. Finally, this consolidated vector is passed to a linear classification head to produce the final output scores for each class.

Architectural parameters such as the number of encoder layers, attention heads, and feedforward dimension were systematically tuned to optimise model performance (see Appendix B.1 for details).

MOE-Transformer

Another variant of the EEGTransformer, termed MoE-Transformer, was developed by integrating a Mixture-of-Experts layer with the Transformer architecture to enhance EEG data processing.

The model first flattens the input EEG signal and processes it through a Mixture-of-Experts layer, where parallel expert networks compete to process input features, with a trainable gating network dynamically weighting their contributions via softmax activation. This gated output is then reshaped and passed to a Transformer block that models temporal relationships through multi-head self-attention, enhanced with layer normalisation and residual connections. Finally, the attended features are classified via a sigmoid-activated dense layer. The complete set of model hyperparameters is documented in Appendix B.1.

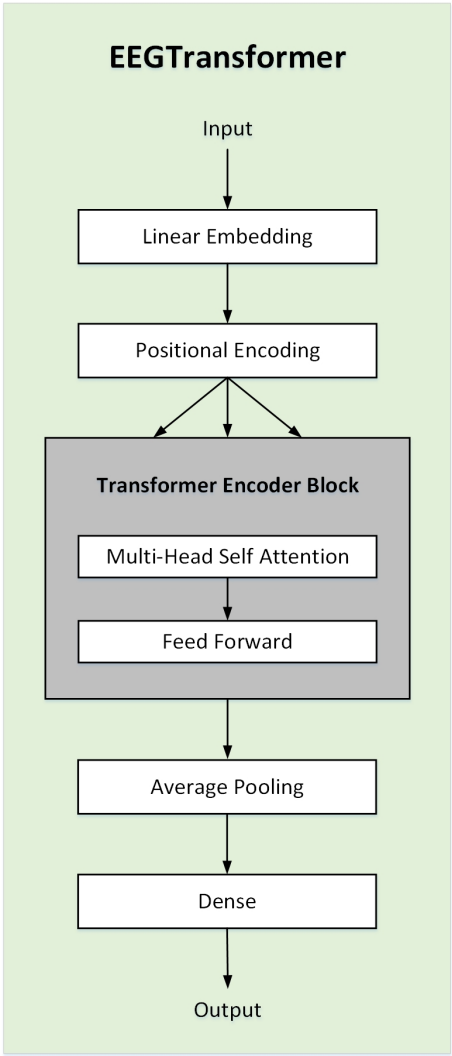


Figure 5.7: Network architectures: EEGTransformer

5.4.4 Hybrid Transformers

EEG Conformer

The EEG Conformer, introduced by Song et al. **[222]**, is a hybrid deep learning architecture designed specifically for enhanced decoding of EEG signals. This model synergistically combines the strengths of Convolutional Neural Networks and Transformer-based self-attention mechanisms, effectively addressing the limitations found in earlier EEG decoding methods.

The EEG Conformer architecture consists of three key components: a convolutional module, a self-attention module, and a fully connected classification layer. The convolutional module first processes the raw 2D EEG input by applying tem-

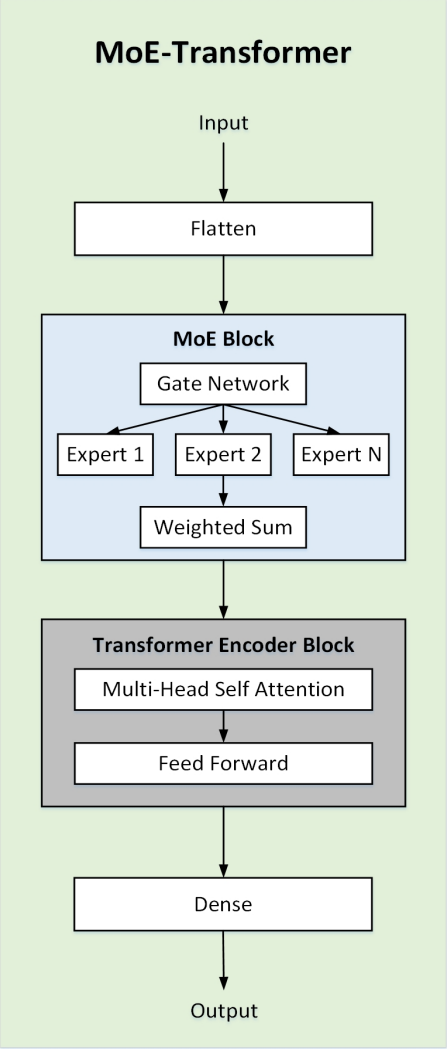


Figure 5.8: Network architectures: MoE-Transformer

poral convolutions to capture time-related patterns and spatial convolutions across electrode channels to extract inter-channel relationships. Average pooling is used to reduce noise and improve generalisation. The output is then passed to the self-attention module, which captures long-range temporal dependencies by modeling global correlations across time points. Finally, the enriched features are flattened and passed through a dense classification layer to generate the final output.

In this work, the EEG Conformer was implemented, and its hyperparameters, including convolutional filter sizes, number of attention heads, transformer layer depth, and dropout rates, were carefully tuned to optimise classification performance for the specific EEG dataset. The complete set of model hyperparameters is given in Appendix B.1.

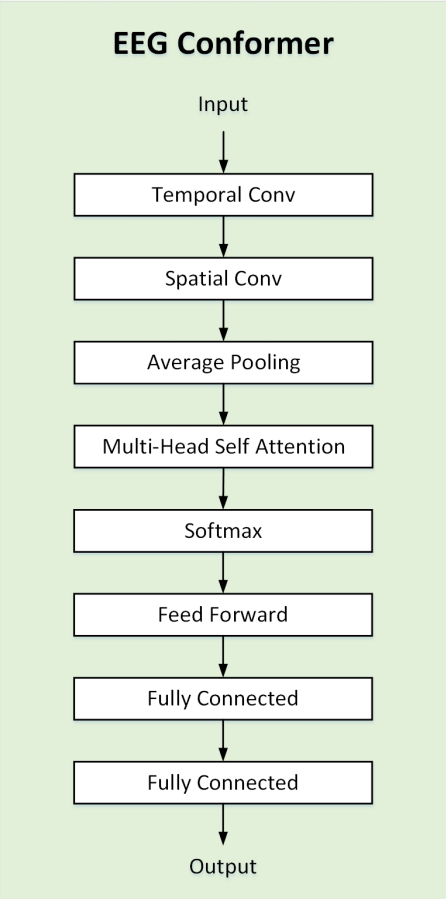


Figure 5.9: Network architectures: EEG Conformer (based upon [222])

CNN1-Transformer

A hybrid architecture is proposed that leverages a Convolutional Neural Network, specifically the CNN-1 model discussed in Section 5.4.2, as an effective feature extractor. This CNN then feeds into a Transformer Encoder to model temporal dependencies.

The model first processes the input EEG data through two layers of CNN-1, utilising its custom scaled-tanh activation function. The initial pointwise convolutional layer transforms the input channels, while the second temporal convolution extracts local features across time and downsamples the sequence.

This sequence of extracted features is then passed to a Transformer Encoder block. The Transformer's multi-head self-attention mechanism analyzes the entire sequence, capturing complex and long-range relationships between the features extracted by the CNN.

Finally, the output from the Transformer is flattened into a single vector and passed through a dense classification head with a sigmoid activation function to produce the final binary prediction. See Appendix B.1 for the full hyperparameter configuration.

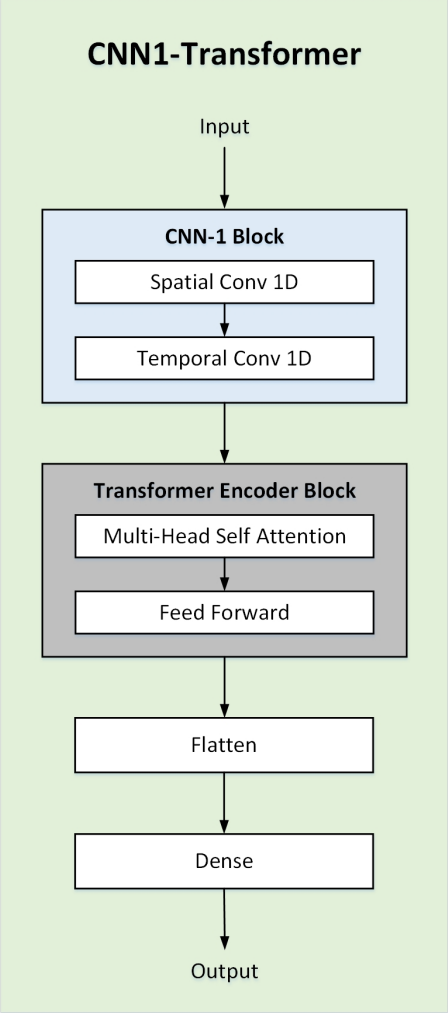


Figure 5.10: Network architectures: CNN1-Transformer

EEGNet-Transformer

A hybrid architecture, termed EEGNet-Transformer, is proposed to effectively combine the strengths of the lightweight yet powerful EEGNet for efficient feature extraction with the sequence-modeling capabilities of a Transformer Encoder..

The EEGNet-Transformer architecture is designed to robustly capture both spatial and temporal characteristics of EEG data for classification. It effectively com-

bines the strengths of the lightweight yet powerful EEGNet for efficient feature extraction with the sequence-modeling capabilities of a Transformer Encoder.

The processing begins with the raw EEG input. This data is first fed into the EEGNet component, which acts as a specialized feature extractor. EEGNet operates in two main blocks. The first block applies a temporal convolution to capture frequency-specific features, followed by a depthwise separable convolution across the channels to learn spatial filters for each feature map. This block also incorporates batch normalisation, an ELU activation function, average pooling for downsampling, and dropout for regularisation. The second block further processes these features using a separable convolution, which involves a depthwise temporal convolution followed by a pointwise convolution. This block similarly uses batch normalisation, ELU activation, average pooling, and dropout to refine the temporal representations and reduce dimensionality.

The output of the EEGNet is then reshaped, permuting and flattening the features to create a sequence of feature vectors suitable for the Transformer, where each vector represents a time-step's extracted characteristics.

Subsequently, these feature sequences are passed to a Transformer Encoder. This encoder is composed of multiple stacked Transformer Encoder layers. Each layer consists of a multi-head self-attention mechanism, which allows the model to weigh the importance of different parts of the input sequence and capture long-range temporal dependencies within the extracted EEG features. Following the attention mechanism, a position-wise feed-forward network is applied to further process the information. Layer normalisation and dropout are applied within each sub-layer to stabilize training and prevent overfitting. The Transformer Encoder's ability to model global relationships across the entire sequence of features from EEGNet is crucial for understanding complex temporal patterns in EEG signals.

Finally, the output sequence from the Transformer Encoder undergoes global average pooling across the time dimension, summarising the learned temporal representations into a single fixed-size feature vector for each sample. This consolidated

vector is then fed into a fully connected classification head, which performs the final prediction by mapping the extracted features to the desired number of output classes.

While previous studies have explored combining EEGNet with transformers, my model improves this hybrid design by making it significantly more lightweight and computationally efficient, making it feasible for real-time or resource-constrained applications. The final set of parameters selected after tuning is summarised in Appendix B.1.

Moreover, this hybrid architecture has not been systematically explored or adapted in real-world contexts involving both clean and noisy EEG signals, including motion artifacts, speech interference, and naturalistic environments. My work addresses this gap by validating the proposed model across diverse and challenging EEG recording conditions, bringing hybrid EEG classification closer to practical BCI deployment.

Multiband EEGNet-Transformer

A Multi-Band EEGNet-Transformer is proposed, combining frequency-band specialisation with hierarchical spatiotemporal feature learning.

The model begins by decomposing the raw EEG signals into four canonical frequency bands, delta, theta, alpha, and beta, each of which is passed through an independent EEGNet module. The EEGNet model consists of two main blocks: the first applies temporal and spatial filtering via depthwise convolution, while the second employs separable convolution to reduce parameters and capture more complex representations. The output of each EEGNet is reshaped into a sequence of embeddings suitable for processing with transformers.

Following the band-specific EEGNet feature extraction, each stream's feature sequence is passed to an independent Transformer Encoder. These parallel Transformer Encoders are responsible for modeling long-range temporal dependencies within their respective frequency bands. Each encoder consists of multiple layers, where multi-head self-attention mechanisms analyze the sequence to identify crucial

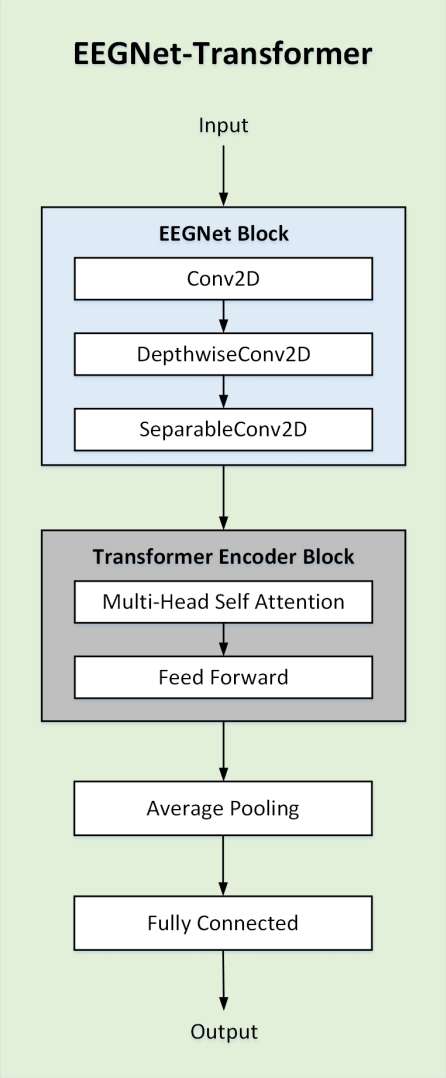


Figure 5.11: Network architectures: EEGNet-Transformer

temporal patterns and relationships, complemented by position-wise feed-forward networks, layer normalisation, and dropout for robust learning.

After temporal modeling, the outputs of the four frequency bands are stacked together, forming a combined representation across all frequency bands. This stacked input is fed into a shared MultiheadAttention layer, which serves as an adaptive fusion mechanism. This attention layer learns to dynamically weigh and combine the information from different frequency bands, effectively determining which band's features are most relevant for the final classification task at any given moment. Finally, the fused representation is passed through a fully connected layer for final classification. See Appendix B.1 for the full hyperparameter configuration.

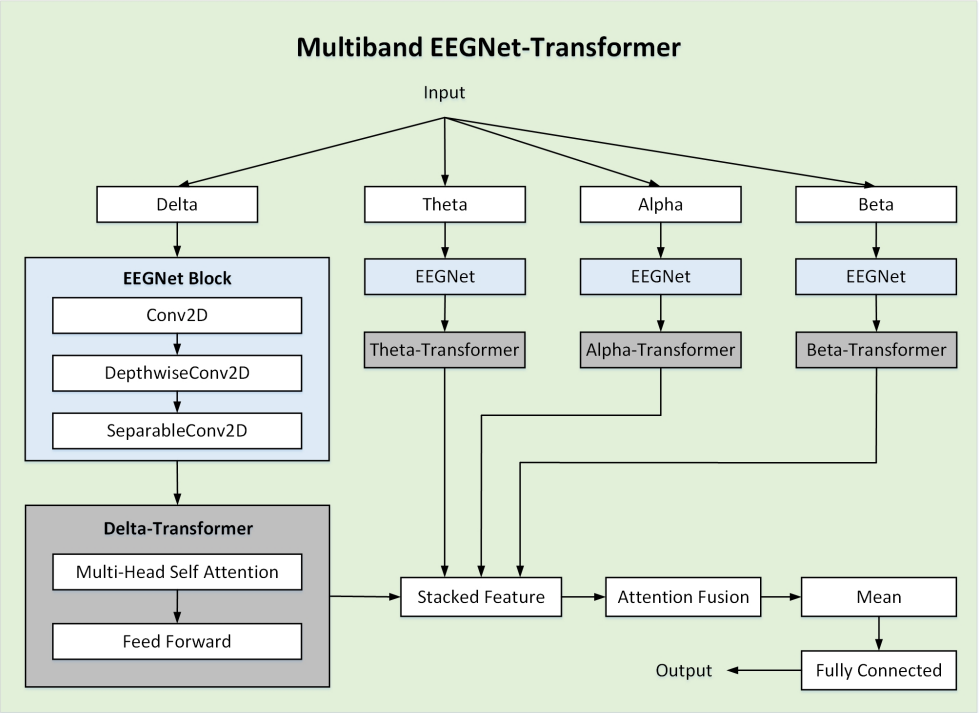


Figure 5.12: Network architectures: Multiband EEGNet-Transformer

EEGNet-MoE-Transformer

An EEGNet-MoE-Transformer is proposed, synergistically combining CNN-based feature extraction, MoE gating, and Transformer attention for EEG processing.

The model first processes raw EEG signals through an EEGNet backbone (temporal → Depthwise spatial → separable convolutions with ELU activation and average pooling), extracting spatiotemporal features.

After feature extraction, the flattened output is passed through a Mixture-of-Experts layer. The MoE layer consists of multiple independent "expert" linear networks and a "gating" network. The gating network takes the input feature vector and generates a set of probabilistic weights, indicating the relevance of each expert for the current input. The final output of the MoE layer is a weighted sum of the outputs from all experts, where the weights are determined by the gating network.

The output from the MoE layer as a single feature vector is unsqueezed and fed into a Transformer block, which introduces self-attention mechanisms to model higher-level interactions between learned representations. The transformer layer includes multi-head self-attention, followed by a position-wise feedforward network

and residual layer normalisation. Finally, the output from the Transformer Block is squeezed and passed through a dense Classification Head with a sigmoid activation function. All architectural hyperparameters and their optimised values are provided in Appendix B.1.

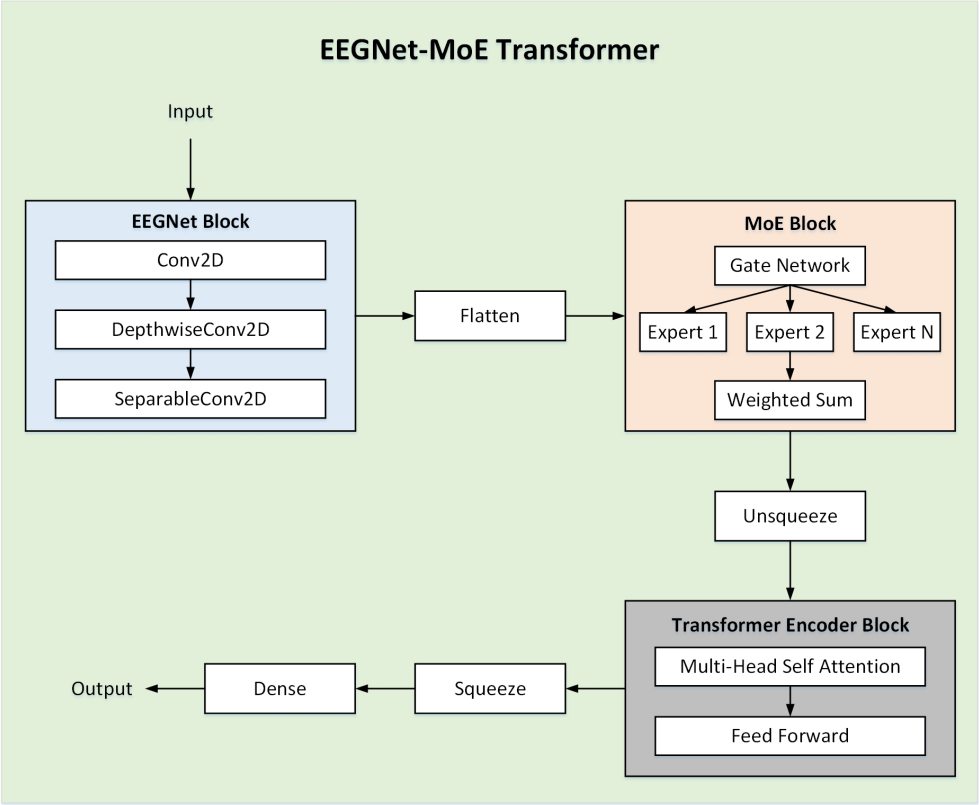


Figure 5.13: Network architectures: EEGNet-MoE-Transformer

Multiband EEGNet-MoE-Transformer

A Multiband EEGNet-MoE-Transformer model is proposed, integrating multiple advanced deep learning modules, including EEGNet, Mixture of Experts, Transformer Encoders, and Attention-based Fusion, to effectively classify EEG signals across different frequency bands.

The architecture processes four EEG bands (delta, theta, alpha, beta) through parallel pipelines. The process begins with the Multi-Band Input, where raw EEG data is separated into its constituent frequency bands (Delta, Theta, Alpha, and Beta).

Each of these distinct frequency bands has a dedicated EEGNet instance that

acts as the initial feature extractor. Following the EEGNet feature extraction, each band-specific feature sequence is fed into its own MoE Layer.

Next, the expert-processed feature sequences from each MoE layer are passed to independent Transformer Encoders. Each transformer's output is positionally encoded and passed through multi-head attention and feedforward sublayers.

The band-specific transformer outputs are aggregated using an Attention-based Fusion mechanism. First, the time dimension is averaged for each frequency band's representation. Then, the representations are stacked and passed through a multi-head attention layer, which learns inter-band dependencies and relative importance. Finally, a fully connected layer maps the fused feature vector to the output space, providing class predictions for target versus non-target EEG trials. The complete set of model hyperparameters is documented in Appendix B.1.

5.5 Evaluation Parameters

To evaluate the performance of the proposed models, a comprehensive set of metrics is used:

1. **Primary Metric**

   ROC-AUC score serves as the principal evaluation metric in this study and is the only metric reported in the main body of results.

   - **ROC-AUC (Receiver Operating Characteristic - Area Under Curve):**

     It measures the model's ability to distinguish between target and non-target classes across all possible thresholds. As discussed in Section 4.1.3, ROC-AUC is particularly well-suited for imbalanced classification tasks, such as the RSVP-P300 paradigm, where the ratio of target to standard trials is approximately 1:10. It is insensitive to class imbalance, making it a reliable indicator of model performance. Additionally, ROC-AUC is

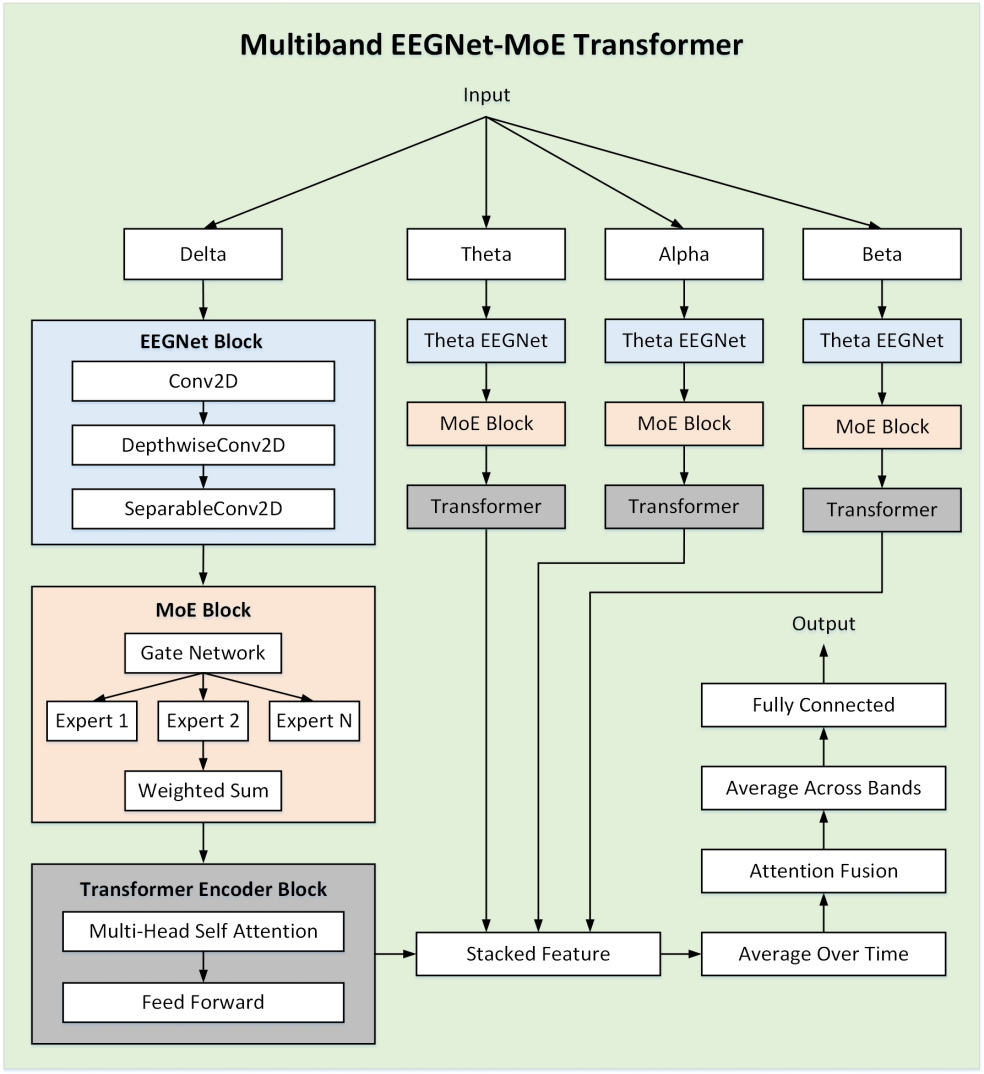


Figure 5.14: Network architectures: Multiband EEGNet-MoE-Transformer

valuable in scenarios involving ranking or prioritisation, as it evaluates the full range of classification thresholds.

2. **Secondary Metrics**

To complement the primary ROC-AUC metric, a threshold-dependent secondary metric was also computed, which provides additional insights into model behavior but is not reported in the main body. These metrics are instead detailed in the Appendix for reference:

- **Accuracy:** Represents the proportion of correctly classified trials among all trials. However, in imbalanced settings, it can be misleading as it may be biased toward the majority class.

- **Balanced Accuracy:** Computes the average of recall obtained on each class, thereby accounting for imbalanced class distributions more fairly than overall accuracy.
- **Precision:** Indicates the proportion of correctly identified target trials out of all trials classified as targets.
- **Recall (Sensitivity or True Positive Rate):** Measures the proportion of actual target trials correctly identified by the model.
- **F1-score:** The harmonic mean of precision and recall, offering a single metric that balances both false positives and false negatives.

Given the imbalanced nature of the RSVP-P300 dataset, with a target-to-standard trial ratio of approximately 1:10, the ROC-AUC score serves as a particularly valuable metric. It evaluates the model's ability to discriminate between classes across all possible thresholds, making it robust to class imbalance and suitable for assessing ranking performance without reliance on a fixed decision boundary.

In addition to ROC-AUC, threshold-dependent secondary metrics such as Accuracy, Balanced Accuracy, Precision, Recall, and F1-score are reported. Rather than relying on a fixed threshold, an ROC curve–based thresholding approach was adopted. The threshold was specifically selected as the point on the ROC curve that provides the optimal balance between sensitivity (True Positive Rate) and specificity (1 – False Positive Rate), typically the point closest to the top-left corner. This data-driven thresholding strategy ensures that the evaluation is adaptive and sensitive to variations in classifier output distributions across subjects, leading to more meaningful metric computation.

All reported threshold-dependent metrics were derived using this ROC curve–based optimal thresholding method. This unified approach improves interpretability and reflects a more realistic evaluation framework aligned with potential real-world BCI applications.

For consistency and generalisability, only ROC-AUC results are reported in the

Table 5.2: Performance comparison of classification methods: ROC-AUC scores (average-avg and standard deviation-SD) for subject-independent evaluation across 20 participants, assessed under both traditional RSVP and intentionally contaminated RSVP paradigms

Subject Independent Classification **Average (Avg) and Standard Deviation (SD) across 20 subjects**									
Sr. No	**Classification**	**RSVP**		**IC-RSVP**					
				Body Movement		**Head Movement**		**Talking**	
	Model	**Avg**	**SD**	**Avg**	**SD**	**Avg**	**SD**	**Avg**	**SD**
CM-1	Bayesian Ridge	0.732	0.091	0.685	0.087	0.574	0.053	0.689	0.084
CM-2	CNN-1	0.763	0.087	0.727	0.083	0.612	0.073	0.721	0.079
CM-3	EEGNet	0.767	0.080	0.728	0.073	0.606	0.077	0.735	0.070
CM-4	MoE-EEGNet	0.751	0.831	0.720	0.086	0.616	0.078	0.708	0.087
CM-5	Multiband EEGNet	0.779	0.078	0.729	0.075	0.609	0.068	0.739	0.073
CM-6	Multiband MOE-EEGNet	0.777	0.071	0.730	0.067	0.621	0.062	0.741	0.062
CM-7	EEGTransformer	0.710	0.076	0.665	0.064	0.577	0.038	0.666	0.076
CM-8	MOE-Transformer	0.764	0.089	0.725	0.084	0.603	0.070	0.726	0.083
CM-9	EEG Conformer	0.764	0.097	0.717	0.086	0.590	0.057	0.733	0.076
CM-10	CNN1-Transformer	0.767	0.087	0.725	0.081	0.610	0.074	0.730	0075
CM-11	EEGNet-Transformer	0.768	0.085	0.732	0.083	0.614	0.061	0.730	0.076
CM-12	Multiband EEGNet-Transformer	0.777	0.082	0.734	0.069	0.625	0.064	0.746	0.063
CM-13	EEGNet-MoE-Transformer	0.753	0.087	0.724	0.083	0.610	0.070	0.726	0.078
CM-14	Multiband EEGNet-MoE-Transformer	0.774	0.077	0.726	0.072	0.617	0.062	0.732	0.066

Complete subject-wise performance metrics for all models are provided in Appendix B.2

main text. The full set of secondary performance metrics computed using subject-specific optimal thresholds is presented in the Appendix B.2 for reference.

5.6 Results and Discussion

The performance of different models presented in Section 5.4 was evaluated under both clean/traditional RSVP conditions and three noisy IC-RSVP conditions (body movement, head movement, and talking artifacts), using ROC-AUC as the evaluation metric across 20 subjects. The table 5.2 summarises the average AUC and standard deviation (SD) for each model and condition.

Bayesian Ridge achieved an average ROC-AUC of 0.732 (SD = 0.091) under the clean RSVP condition, establishing a solid baseline for a traditional linear model. However, its performance declined in noisy scenarios, with 0.685 (SD = 0.087) under body movement, 0.574 (SD = 0.053) under head movement, and 0.689 (SD = 0.084) under talking artifacts. These results suggest that while Bayesian Ridge can handle some variability due to motion artifacts, it is highly sensitive to complex, non-stationary noise, particularly head movements, which appear to disrupt the spatial

and temporal structure of EEG signals beyond the model's capacity.

Building on this, a 1D CNN (i.e., CNN-1 [166]) was employed as a baseline convolutional neural network and demonstrated a clear improvement over the linear model, leveraging convolutional inductive biases to effectively capture spatio-temporal features in the EEG data. It achieved 0.763 (SD = 0.087) under clean conditions and retained better robustness under noise: 0.727 (SD = 0.083) for body, 0.612 (SD = 0.073) for head, and 0.721 (SD = 0.079) for talking. These results reflect CNN-1's ability to extract more localized features that are somewhat invariant to artifacts, although head movement remains a challenging condition, showing a similar drop as with Bayesian Ridge.

To further optimise CNN architectures for EEG decoding, EEGNet, originally proposed by Lawhern et al. [165], was explored as a compact convolutional model specifically designed for EEG data. EEGNet demonstrated slightly better clean-condition performance at 0.767 (SD = 0.080), and showed consistent robustness under noisy conditions: 0.728 (SD = 0.073) with body, 0.606 (SD = 0.077) with head, and 0.735 (SD = 0.070) with talking artifacts. Compared to CNN-1, EEGNet yielded improved results under speech-related noise, likely due to its depthwise separable convolutions and use of temporal filters tailored to EEG frequency characteristics.

After evaluating the base EEGNet, the architecture was extended with a Mixture of Experts (MoE) mechanism, hypothesizing that different sub-networks could specialize in decoding under distinct noise conditions. The MoE-EEGNet achieved 0.751 (SD = 0.083) in clean RSVP, and demonstrated competitive results across noisy conditions: 0.720 (SD = 0.086) for body, 0.616 (SD = 0.078) for head, and 0.708 (SD = 0.087) for talking. While performance was slightly lower than base EEGNet in clean and talking settings, MoE-EEGNet maintained comparable robustness in more artifact-prone conditions, suggesting the benefit of diversity in expert pathways.

Building upon the strengths of EEGNet, the Multiband EEGNet was introduced to leverage frequency-specific decomposition for enhanced EEG feature extraction.

This model processed each EEG frequency band (e.g., delta, theta, alpha, beta) separately before fusion, allowing more targeted feature extraction. Multiband EEGNet reported the highest clean RSVP performance among base models at 0.779 (SD = 0.078), and maintained strong performance under body (0.729, SD = 0.075), head (0.609, SD = 0.068), and talking (0.739, SD = 0.073) noise. These improvements indicate that isolating frequency components enhances signal discriminability and noise robustness.

To further improve specialisation and flexibility, frequency band decomposition was combined with a mixture of experts in the Multiband MoE-EEGNet architecture. This hybrid model maintained a strong clean performance of 0.777 (SD = 0.071), and demonstrated slightly better generalisation to noise compared to its individual components: 0.730 (SD = 0.067) for body, 0.621 (SD = 0.062) for head, and 0.741 (SD = 0.062) for talking. This shows that combining multiband learning with expert selection provides more adaptive and noise-resilient decoding.

Shifting focus toward models that can capture long-range dependencies, EEG-Transformer, a baseline self-attention model, was evaluated. It achieved 0.710 (SD = 0.076) on clean RSVP data, but its performance declined significantly under noise: 0.665 (SD = 0.064) for body, 0.577 (SD = 0.038) for head, and 0.666 (SD = 0.076) for talking artifacts. While the transformer captures temporal relationships well as reported in the recent studies [219], [220], its lack of strong inductive biases may make it less effective in low-SNR EEG settings unless guided by domain-specific structure.

To address this limitation, the MoE mechanism was integrated into the transformer, resulting in the MoE-Transformer. This model achieved a much improved clean score of 0.764 (SD = 0.089), and handled noise more robustly than the base transformer: 0.725 (SD = 0.084) for body, 0.603 (SD = 0.070) for head, and 0.726 (SD = 0.083) for talking. These gains highlight that gating multiple expert paths helps the transformer adapt to heterogeneous noise patterns.

The EEG Conformer, introduced by Song et al. [222], was also evaluated for

its ability to combine CNN-based local feature extraction with self-attention-based sequence modeling. It matched CNN-1 and EEGNet in clean RSVP (0.764, SD = 0.097) and achieved 0.717 (SD = 0.086) for body, 0.590 (SD = 0.057) for head, and 0.733 (SD = 0.076) for talking artifacts. Its hybrid design showed promise, particularly in speech-related conditions, though head movement remained a challenge.

Motivated by this hybrid approach, CNN1-Transformer, combining 1D-CNN feature extraction with transformer-based attention, was proposed. This model achieved 0.767 (SD = 0.087) under clean conditions and generalised well across noise: 0.725 (SD = 0.081) for body, 0.610 (SD = 0.074) for head, and 0.730 (SD = 0.075) for talking. The combination of local feature encoding and global sequence modeling appears effective, particularly for moderate noise conditions.

EEGNet-Transformer, combining the specialized EEGNet model with a transformer self-attention mechanism, was then evaluated. This model delivered one of the best performances overall, with 0.768 (SD = 0.085) clean RSVP and robust performance under body (0.732, SD = 0.083), head (0.614, SD = 0.061), and talking (0.730, SD = 0.076) conditions. This confirms that attention mechanisms enhance EEGNet's spatial-temporal representations, particularly in noisy environments.

To extend this further, the Multiband EEGNet-Transformer was investigated, which integrates multiband decomposition into the EEGNet-Transformer architecture. This model stood out with 0.777 (SD = 0.082) RSVP, and the strongest results across noise types: 0.734 (SD = 0.069) for body, 0.625 (SD = 0.064) for head, and 0.746 (SD = 0.063) for talking. These results establish this model as the most robust and consistent across all conditions, underscoring the power of combining frequency specialisation with attention-based learning.

Next, EEGNet-MoE-Transformer was evaluated, which integrates expert gating into the EEGNet-Transformer architecture. It achieved 0.753 (SD = 0.087) clean and held stable performance across noisy settings: 0.724 (SD = 0.083) for body, 0.610 (SD = 0.070) for head, and 0.726 (SD = 0.078) for talking. Although slightly below the multiband transformer variants, this model still showed dependable performance,

indicating the benefits of adaptive expert pathways even without explicit frequency decomposition.

Lastly, the most comprehensive model in this study, the Multiband EEGNet-MoE-Transformer, combined all three strategies: frequency-band decomposition, EEG-specific feature extraction, and transformer-based attention with expert gating. It achieved 0.774 (SD = 0.077) in the traditional lab setting, and sustained solid generalisation in noisy conditions: 0.726 (SD = 0.072) for body, 0.617 (SD = 0.062) for head, and 0.732 (SD = 0.066) for talking. While slightly below the best multiband transformer variant, it demonstrated stable and consistent performance across all noise types, making it a strong candidate for real-world BCI applications where robustness is critical.

- **Cross-Model Insights on Subject-Independent EEG Classification Across Varying Environmental Conditions**

A cross-model comparison is presented to better understand the trade-offs between different model types and their suitability for real-world BCI deployment. Building on the model-wise insights discussed above under both clean and noisy conditions, this analysis synthesizes the results to highlight how performance evolves from classical machine learning to advanced deep learning architectures, particularly in the presence of real-world artifacts.

In order to provide a more in-depth comparison and highlight the effect of each artifact separately, 5.15 illustrates the performance of different models under individual noise conditions in the form of a bar chart.

Building on this, performance was then averaged across all three noise conditions for each model and compared with their clean RSVP scores to give a consolidated view (see Figure 5.16). This allows us to assess each model's generalisation capacity and resilience to real-world EEG artifacts.

Moving from traditional machine learning to advanced deep learning architectures reveals a clear trend in performance and robustness across clean and noisy EEG conditions. When averaged across the three noise conditions, traditional Bayesian

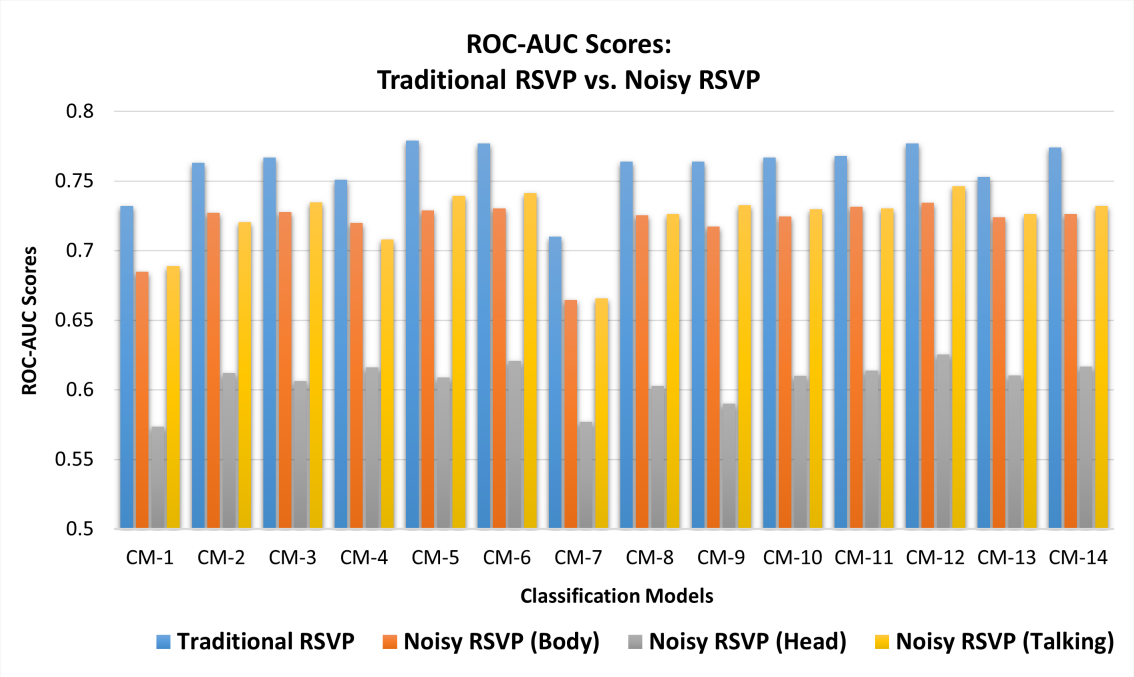


Figure 5.15: Average ROC-AUC scores of each classification model under individual noisy conditions (body, head, and talking artifacts). Models are anonymized as CM-1 to CM-14 for comparison.

CM-1: Bayesian Ridge, CM-2: EEGNet, CM-3: CNN-1, CM-4: MoE-EEGNet, CM-5: Multiband EEGNet, CM-6: Multiband MoE-EEGNet, CM-7: EEGTransformer, CM-8: MoE-Transformer, CM-9: EEG Conformer, CM-10: CNN1-Transformer, CM-11: EEGNet-Transformer, CM-12: Multiband EEGNet-Transformer, CM-13: EEGNet-MoE Transformer, CM-14: Multiband EEGNet-MoE Transformer.

Ridge Regression exhibited the lowest resilience to noise, with a notable drop from a clean ROC-AUC of 0.732 to an average noisy AUC of 0.649. This highlights the model's limited capacity to generalise under non-stationary and artifact-laden inputs, reflecting its reliance on linear assumptions and handcrafted features.

Transitioning to convolutional neural networks, CNN-1 delivered a significant improvement in both clean and noisy settings, reaching 0.763 on clean RSVP and 0.687 on noisy IC-RSVP. This improvement illustrates the benefits of using data-driven feature extraction via convolutions, which can better model the spatial-temporal structure of EEG signals. However, the gap between clean and noisy conditions remains substantial, suggesting that CNN-1, while effective, still struggles to generalise fully in the presence of real-world artifacts.

EEGNet further improved robustness, scoring 0.767 (clean) and 0.690 (noisy average), and benefited from its compact architecture designed specifically for EEG

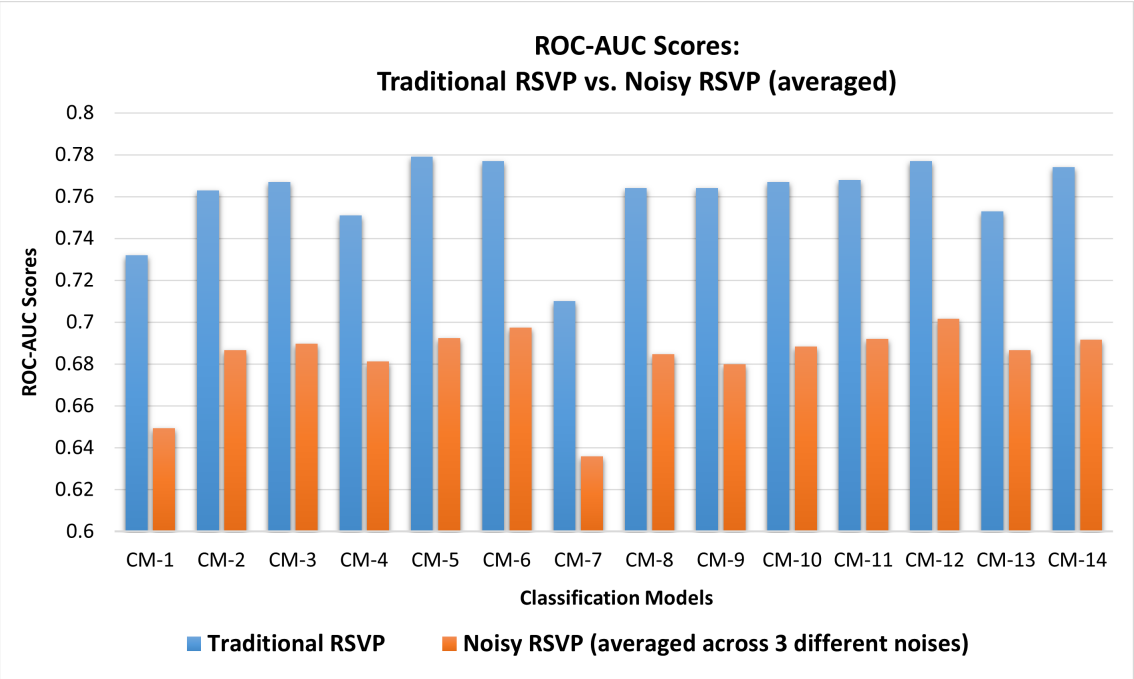


Figure 5.16: Average ROC-AUC scores of each classification model under clean (RSVP) and noisy (IC-RSVP) conditions. Noisy scores represent the average across body, head, and talking artifact conditions. Models are anonymized as CM-1 to CM-14 for comparison.

decoding. Its variant with Mixture of Experts (MoE-EEGNet) achieved slightly lower clean (0.751) and noisy performance (0.681), indicating that while expert-based routing can be beneficial, it might not consistently outperform the simpler EEGNet when applied independently. The multiband version of EEGNet enhanced performance by decomposing EEG into distinct frequency bands, reaching 0.779 on clean and 0.692 on noisy data, showcasing the value of frequency-specific processing. In addition, combining the multiband approach with MoE (Multiband MoE-EEGNet) preserved this advantage, yielding 0.777 clean and 0.697 average noisy ROC-AUC scores, indicating slightly better generalisation and highlighting the cumulative benefits of modular specialisation and spectral decomposition.

In comparison, transformer-based models initially lagged behind CNNs in noisy settings. The base EEGTransformer scored 0.710 (clean) and 0.636 (noisy RSVP condition), suggesting that while attention mechanisms model long-range dependencies, they may require more data or inductive bias to be effective in noisy EEG

contexts. However, integrating MoE into transformers (MoE-Transformer) covered this gap, achieving 0.764 clean and 0.685 noisy AUC, indicating that adaptive routing helps mitigate overfitting and noise sensitivity.

Further improvement came from hybrid architectures. EEG Conformer, a hybrid model from the literature, achieved 0.764 clean and 0.680 noisy performance, benefiting from both local and global feature modeling. The proposed CNN1-Transformer also showed strong performance (0.767 clean, 0.688 noisy), affirming the synergy between convolutional priors and attention mechanisms.

The EEGNet-Transformer emerged as one of the most balanced and robust models, scoring 0.768 on traditional RSVP and 0.692 on artefactual RSVP, outperforming both its standalone EEGNet and transformer variants. The multiband version (Multiband EEGNet-Transformer) reached 0.777 (traditional) and 0.702 (noisy), the highest noisy average across all models, confirming that combining spectral decomposition with attention leads to superior generalisation in real-world conditions.

Finally, the most comprehensive hybrid model, Multiband EEGNet-MoE Transformer, offered a 0.774 lab-constrained ROC-AUC score and 0.692 noisy ROC-AUC score. Although slightly below the simpler multiband transformer variant in noisy settings, it maintained excellent stability and performance across all conditions, suggesting that architectural complexity helps in certain scenarios but might not always outperform more streamlined alternatives.

Overall, the cross-model comparison highlights that while basic ML models like Bayesian Ridge serve as informative baselines, deep learning architectures, especially those incorporating convolutional and attention mechanisms with spectral decomposition, demonstrate superior robustness and generalisation to real-world EEG noise. Notably, models like Multiband EEGNet-Transformer and EEGNet-Transformer strike the best trade-off between clean and noisy performance, positioning them as strong candidates for practical, subject-independent BCI systems.

While the proposed novel architectures, featuring multiband processing and the integration of Mixture of Experts, demonstrated notable gains in classification per-

formance, especially under noisy conditions, these enhancements introduced additional computational overhead due to their architectural complexity and frequency-wise processing demands. Given these trade-offs, and in the interest of identifying models that strike a practical balance between robustness and deployability, four representative models were selected for further in-depth analysis in the subsequent chapter: CNN-1, EEGNet, EEG Conformer, and EEGNet-Transformer. These models not only delivered competitive performance in both clean and noisy scenarios but also offer greater computational efficiency, making them strong candidates for real-world BCI implementations where hardware limitations and real-time responsiveness are key constraints.

5.7 Supplementary Approaches to Subject Independent Classification

While the core objective of this chapter was to explore and enhance subject-independent classification, the focus thus far has been on evaluating various model architectures under traditional lab-constrained/clean and noisy conditions. In addition to the primary methods and results discussed above, supplementary approaches were also explored to enhance cross-subject generalisation. In particular, contrastive learning was applied as a representation learning technique to improve model robustness on EEG data contaminated with real-world noise and artifacts. The second strategy, where clean and noisy data were combined during training to promote robustness, was also explored. However, neither approach yielded meaningful performance gains in the real-world noisy setting. The following subsections briefly describe these two methods and outline possible reasons for their limited effectiveness.

5.7.1 Contrastive Learning

As part of the investigation into subject-independent classification, contrastive learning, a prominent representation learning technique, was also explored. This ap-

proach was tailored to EEG data contaminated with real-world noise and artifacts, aiming to enhance feature robustness and improve generalisation across subjects.

Contrastive learning focuses on learning invariant and discriminative feature embeddings by comparing different views or augmentations of the data. The fundamental idea is to bring similar instances, referred to as positive pairs, closer together in the learned feature space, while simultaneously pushing apart dissimilar instances, known as negative pairs [285]. This process encourages the model to learn representations that are not only robust to variations in the input but also highly informative for downstream tasks such as classification or clustering [286].

Given the flexibility of my dataset, which included clean and noisy recordings under various artifact conditions, multiple pairing strategies were explored to construct contrastive examples. These included clean vs. noisy pairs, target vs. standard trials, session-to-session pairs, and subject-to-subject pairs, with the aim of learning robust and invariant EEG feature representations. EEGNet, identified earlier as an efficient architecture, was used as the backbone model for this approach.

In addition to individual pairing strategies, a progressive training pipeline was also explored. The model was first trained using clean vs. noisy pairs to learn general artifact-invariant features. This was followed by a second phase where it was fine-tuned using session-to-session pairs to adapt to inter-session variability. Finally, the model was trained on target vs. standard pairs to emphasize class-discriminative features relevant for RSVP-based classification.

Despite the promise of these strategies, particularly given the success of contrastive learning in prior EEG literature [230], [287], [288], the achieved results were not encouraging. The classification performance across all pairing schemes remained consistently below that of simpler models. Moreover, the computational overhead involved in generating contrastive pairs and training in multiple phases did not translate to meaningful performance gains. This suggests that, at least in our setting, the benefits of contrastive learning are limited when compared to well-tuned supervised approaches.

5.7.2 Introducing noisy RSVP as part of training

In another set of experiments, the impact of incorporating both clean and noisy data during the training phase of subject-independent classification was explored. Throughout this chapter, all models were consistently trained using only the clean (traditional RSVP) trials and evaluated across both clean and various noisy conditions. However, this additional analysis aimed to investigate whether including noisy data in the training set could improve generalisation, particularly under real-world noise.

The influence of clean and noisy data mixtures on classification performance has already been discussed in Chapter 4, but within a subject-specific framework. In this case, the goal was to assess the same concept under subject-independent settings.

Although it was intuitively assumed that introducing noisy trials during training might help the models learn more robust features and thus improve performance on noisy test data, the results did not consistently support this hypothesis. While there were isolated cases, particularly in head movement conditions, where a marginal improvement was observed, these gains were not consistent across subjects, models, or noise types. More importantly, the inclusion of noisy data in training often led to a degradation in overall classification performance, not only on the noisy test sets but also on the clean evaluations.

These findings suggest that, in subject-independent classification, simply mixing clean and noisy data during training is not a straightforward solution for improving model robustness. The presence of noisy data may introduce variability that hinders generalisation, especially when the models are not explicitly designed to adapt to different artifact profiles. The results from this analysis can be found in Appendix-C.

5.8 Summary

Addressing **Research Question 3 (RQ.3)**, Which advanced machine learning models are most effective in achieving subject-independent P300 detection robust

to real-world noise?, this chapter presented a comprehensive evaluation of a range of models, from traditional machine learning techniques to deep learning architectures such as CNNs and Transformers.

With the broader goal of moving brain-computer interfaces beyond controlled laboratory environments into real-world applications, subject-independent classification becomes a critical capability. To this end, a Leave-One-Subject-Out cross-validation strategy was adopted, ensuring that models were evaluated on entirely unseen subjects to simulate real-world variability. The model spectrum included traditional approaches like Bayesian Ridge, early CNNs such as CNN-1, more advanced depthwise-separable convolutional networks like EEGNet, and Transformer-based architectures ranging from basic self-attention models to hybrid designs that combined CNNs with attention mechanisms.

Given the importance of frequency-specific information in EEG data and the heterogeneity introduced by real-world artifacts, multiband processing and Mixture of Experts were also explored as architectural enhancements. These components were tested individually and in combination across several base architectures, including EEGNet, Transformers, and their hybrid variants. Due to the heavy class imbalance in the RSVP-P300 dataset, ROC-AUC was employed as the primary evaluation metric, as it remains threshold-independent and more accurately reflects model discrimination performance in imbalanced classification settings.

Our results showed that the proposed EEGNet-Transformer model enhanced with multiband processing achieved the best overall performance, outperforming all other configurations across clean and noisy testing scenarios. This underscores the benefit of combining spatial filtering (via convolution), temporal attention (via Transformers), and frequency-specific decomposition. However, it was also observed that models with multiband and MoE components came with increased computational costs.

Therefore, to balance classification performance with computational efficiency, four representative models were selected for detailed investigation in the subse-

quent chapter: (1) CNN-1, (2) EEGNet, (3) EEG Conformer, and (4) EEGNet-Transformer.

In addition to this core analysis, several auxiliary strategies were explored to enhance subject-independent classification. One such approach involved applying contrastive learning, leveraging the structured presence of clean and noisy data to construct diverse pairing schemes. Despite promising findings in the literature, the achieved outcomes were underwhelming. Similarly, another strategy involved training models with a mixture of clean and noisy data in hopes of improving generalisation to noisy test conditions; however, this strategy also failed to deliver consistent improvements and often led to degraded performance on both clean and noisy evaluations.

These outcomes highlight that naively applying complex training strategies does not guarantee performance gains in subject-independent EEG classification, especially under real-world conditions. Instead, careful model design and architectural choices, such as those combining multiband analysis and temporal attention, are more impactful, even when computational resources are limited.

Chapter 6

Practical Considerations for Real-World BCI Deployment

This chapter addresses **Research Question 4**, which focuses on identifying key practical refinements and design considerations necessary to advance BCI systems toward real-world usage. It specifically breaks down into four sub-questions: **(RQ.4a)** What is the impact of the amount of training data from different subjects on the performance of subject-independent BCI models? **(RQ.4b)** How much subject-specific calibration data is required to effectively adapt subject-independent models for improved performance in the presence of noise? **(RQ.4c)** What is the optimal number and combination of EEG channels required to maintain high classification accuracy while minimising complexity in real-world BCI applications? and **(RQ.4d)** How does the use of a head-mounted display for RSVP presentation, compared to a traditional monitor, affect EEG signals and overall BCI performance?

To explore these questions in a controlled and focused manner, only clean RSVP-P300 EEG data collected in traditional laboratory settings are utilized. The goal here is not to assess noise robustness, which has already been investigated in earlier chapters, but rather to study core design variables such as training data size, fine-tuning strategies, and channel optimisation.

Given the large number of models developed and benchmarked in the previous

chapter, a representative subset of four models was deliberately selected for inclusion in this chapter: CNN-1, EEGNet, EEG Conformer, and EEGNet-Transformer. These models span a spectrum from traditional CNN architectures to more advanced Transformer-based hybrids, offering a balance between predictive performance and computational feasibility while maintaining clarity and focus.

As outlined in Section 5.5, ROC-AUC is adopted as the primary evaluation metric throughout this chapter due to its robustness to class imbalance and its ability to reflect ranking performance across all thresholds. To maintain consistency and interpretability across experiments, all results are reported in terms of ROC-AUC scores.

6.1 Effect of Training Data Size

This section systematically investigates the effect of using different amounts of training data, e.g., training with 5, 10, 15, or 19 subjects in LOSO setups, and assesses the trade-offs between data quantity and classification accuracy, directly addressing research question **(RQ.4a)** regarding the impact of training data volume on subject-independent BCI performance. These insights are vital for planning data acquisition strategies in practical BCI deployments.

As highlighted in Section 2.5.3 of the literature review, there remains a critical gap in understanding how training data requirements scale, especially under real-world conditions. This analysis aims to address that gap by providing empirical evidence on how varying the number of training subjects influences generalisability and robustness in BCI applications.

To examine the impact of training data size on model performance in a subject-independent EEG classification setting, a series of controlled experiments is conducted by systematically varying the number of training subjects. Initially, the leave-one-subject-out (LOSO) approach was adopted, where 19 subjects were used for training and the remaining one was held out for testing. To assess the influence of training data quantity on model generalisability, the number of training subjects

was further reduced to 15, 10, and 5, while maintaining one subject as the test set in each iteration.

For each training size configuration (15, 10, and 5), the training subjects were selected randomly from the available pool, excluding the test subject. This random selection process was repeated five times per test subject to ensure that results were not biased by any particular subset of training data. The performance metrics from these multiple runs were averaged to obtain a robust estimate of the model's behavior under varying data availability conditions.

The 10-subject training configuration is illustrated in Figure 6.1, which serves as a representative example of how training subjects were randomly selected, five times per test subject, while excluding the test subject from the training pool.

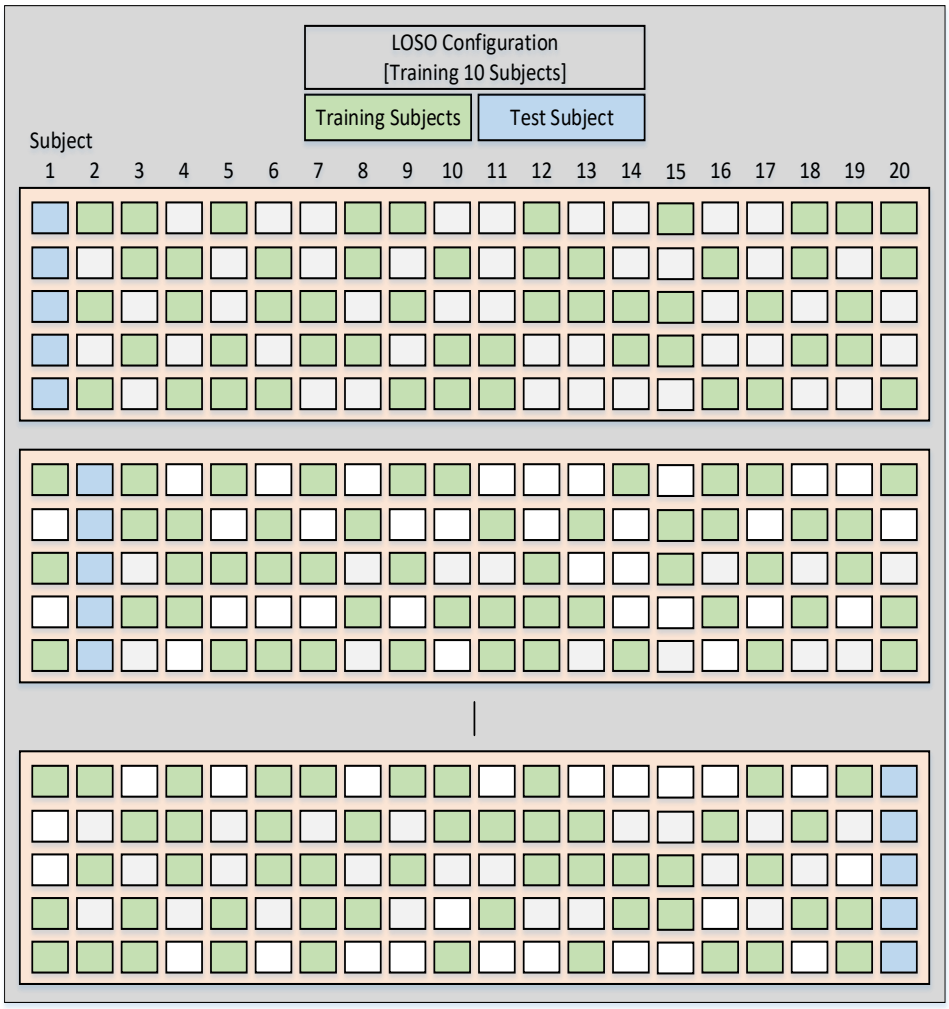


Figure 6.1: LOSO (training 10 subjects) configuration

This analysis plays a crucial role in understanding the trade-off between training data size and model performance. It reflects real-world scenarios where collecting large-scale EEG data from multiple individuals may not always be feasible. By exploring these different training configurations, practical insights are provided into how well subject-independent models can be expected to perform when trained with limited data, which is particularly valuable for applications in clinical or portable BCI systems with constrained datasets.

6.2 Fine-Tuning for Subject-specific Calibration

Fine-tuning plays a critical role in real-world BCI applications, where small amounts of subject-specific calibration data can substantially enhance performance without the need for extensive retraining.

Addressing research question **(RQ.4b)** in this section, the investigation focuses not only on the overall impact of fine-tuning for subject-independent EEG classification models but also evaluates the effectiveness of different fine-tuning strategies and determines how much calibration data is sufficient for practical adaptation. As noted in Section 2.5.2 of the literature review, a key limitation in current research is the lack of understanding around how much fine-tuning is truly effective, particularly when calibration data is affected by noise from movement, speech, or environmental factors. This analysis aims to bridge that gap by empirically evaluating fine-tuning needs under both clean and noisy conditions, which is crucial for rapid personalisation in real-world BCI deployments.

To explore this, the performance of the same four representative models, with and without fine-tuning, is compared on a subject left out during training (LOSO). This comparison highlights the significance of incorporating subject-specific adaptation. Additionally, two fine-tuning strategies were analysed: (1) random train-test splitting using all available data from the test subject and (2) session-based fine-tuning. Both approaches aimed to calibrate a model that was initially trained in a subject-independent manner using a leave-one-subject-out scheme, where data from

19 subjects were used to train the model, and the remaining subject served as the unseen test subject.

- **Random train-test splitting fine-tuning**

In the first approach, data from all four sessions of the test subject were pooled together and randomly split into training and testing sets (e.g., 30% training, 70% testing). This method offers a quick way to evaluate fine-tuning effectiveness but risks temporal leakage, where similar or temporally adjacent trials might appear in both the training and testing sets, potentially inflating performance estimates.

- **Session-based fine-tuning**

To address the limitations of the first approach, random train-test splitting, which may lead to overly optimistic performance estimates, a second strategy was adopted: session-based fine-tuning. This approach better simulates real-world calibration scenarios by leveraging temporally distinct sessions for fine-tuning and testing.

Each subject in the dataset completed four separate recording sessions. Models were initially trained in a LOSO fashion, using data from all subjects except the test subject. To adapt the pretrained model to the test subject, the model was incrementally fine-tuned using an increasing number of sessions from the test subject while keeping the remaining sessions for evaluation.

Specifically, the following were used:

1. Session 1 for fine-tuning, and Sessions 2, 3, and 4 for testing, representing a 25% fine-tune / 75% test split.

2. Sessions 1 and 2 for fine-tuning, and Sessions 3 and 4 for testing, representing a 50% fine-tune / 50% test split.

3. Sessions 1, 2, and 3 for fine-tuning, and Session 4 for testing, representing a 75% fine-tune / 25% test split.

These session-based fine-tuning configurations are visually summarised in Figure 6.2, which illustrates how different proportions of session data were allocated for fine-tuning and testing.

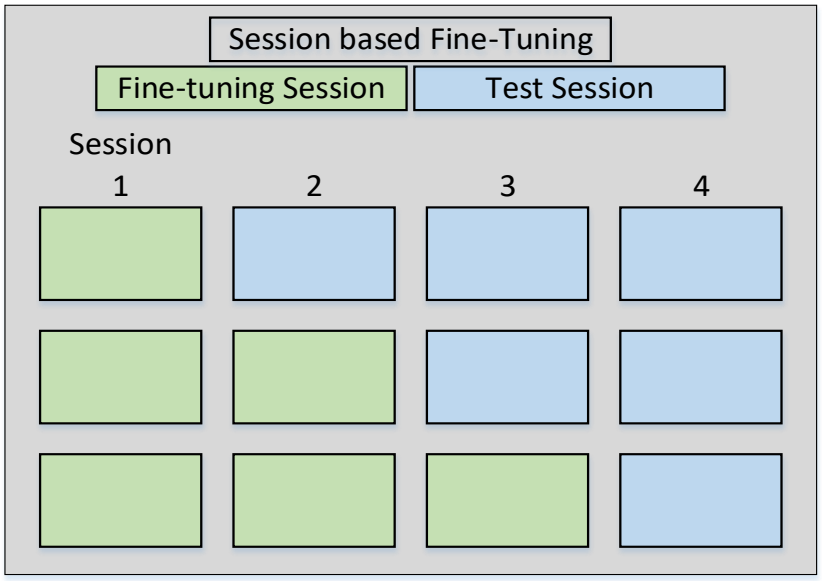


Figure 6.2: Illustration of session-based fine-tuning configurations.

This session-wise protocol was chosen over random data splitting to preserve the temporal order of EEG data collection and to simulate a realistic calibration setting, where data collected during an initial session is used to calibrate the model before real-time use in future sessions. Random splitting could introduce optimistic bias by mixing temporally adjacent trials across fine-tuning and testing sets, which does not accurately reflect real-world usage where data is collected in blocks or sessions.

This progressive fine-tuning framework allows us to systematically assess the trade-off between calibration effort and performance gain, offering insights into the minimum amount of subject-specific data needed to achieve reliable results. Such analysis is crucial for developing user-friendly and time-efficient BCI systems, especially in real-world environments where prolonged calibration is often impractical.

6.3 Optimal Channels

In real-world BCI applications, usability, portability, and user comfort become crucial factors. Unlike controlled laboratory settings, deploying BCIs outside the lab necessitates minimising setup time, reducing hardware complexity, and improving wearability. This makes channel optimisation an essential step toward practical and scalable BCI systems.

High-density EEG systems with 32 or more electrodes provide rich spatial information, but they are often impractical for real-life use due to their bulkiness and lengthy setup procedures. Therefore, identifying a minimal yet effective subset of electrodes is critical to maintain classification performance while enhancing system usability.

Addressing research question **(RQ.4c)**, a systematic exploration of reduced-channel configurations was conducted across four distinct neural network architectures to assess their impact on classification performance. By evaluating both literature-based and custom-designed electrode subsets, it aimed to determine the minimum number of electrodes required to retain high decoding accuracy. This analysis not only quantified the trade-offs between channel reduction and model performance but also revealed the differential robustness of each architecture to spatial sparsity.

As noted in Section 2.5.4 of the literature review, most prior work has focused on posterior or occipital regions and lacked systematic evaluations of progressive channel reduction, particularly under real-world conditions involving motion artifacts or integration with HMDs. This work directly addresses these limitations by validating reduced-channel models in both clean and noisy scenarios, advancing the feasibility of practical BCI deployment.

To assess the feasibility of reducing the number of EEG channels for practical BCI deployment, the impact of various electrode subsets on model performance was investigated. The baseline configuration consisted of the full set of 32 channels (see Figure 6.3) used during data acquisition:

32-channel (default):

- **Ch-32:** Fp1, Fpz, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, TP9, T7, C3, Cz, C4, T8, TP10, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, POz, O1, Oz, and O2.

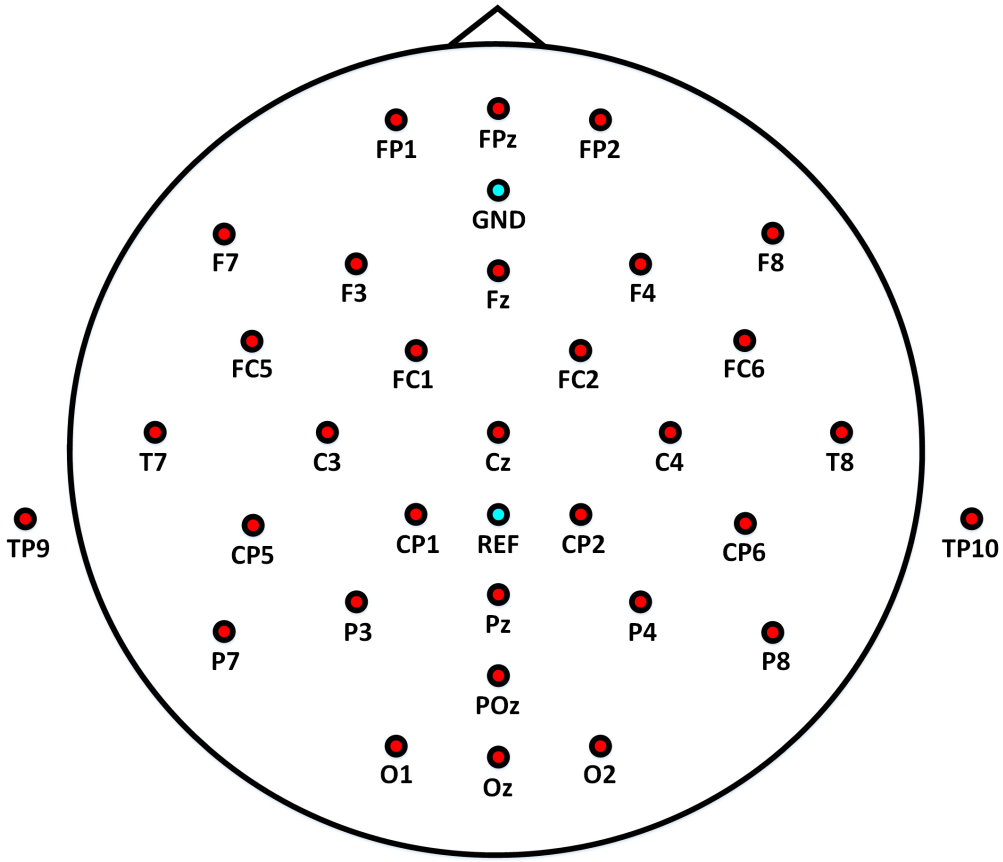


Figure 6.3: Electrode configuration showing all 32 EEG channels

To explore more compact and potentially optimal configurations, a variety of reduced-channel sets was evaluated. Some of these were from existing literature, while others were empirically designed. The combinations are grouped below. Each channel configuration is visualized in Figure 6.4, where the selected electrodes are marked in red.

1. **16-channel sets:**

   - **Ch-16a [199]:** FC1, FC2, C3, Cz, C4, CP1, CP2, P7, P3, Pz, P4, P8, POz, O1, Oz and O2

- **Ch-16b:** Fpz, FC1, FC2, TP9, TP10, CP5, CP6, P7, P3, Pz, P4, P8, POz, O1, Oz and O2

2. **12-channel set:**

   - **Ch-12:** FC2, TP10, CP5, CP6, P7, P3, Pz, P4, P8, POz, O1 and O2

3. **10-channel set:**

   - **Ch-10:** FC2, TP10, CP5, CP6, P7, P3, Pz, P4, POz and O1

4. **8-channel sets:**

   - **Ch-8a:** [**200**] Fz, Cz, P7, P3, Pz, P4, P8 and Oz
   - **Ch-8b:** FC2, TP10, P7, Pz, P4, P8, POz and O2

5. **6-channel sets:**

   - **Ch-6a [199]:** P3, Pz, P4, POz, O1 and O2
   - **Ch-6b:** P7, P3, Pz, P4, P8 and POz

6. **4-channel sets:**

   - **Ch-4a [199]:** P3, P4, O1 and O2
   - **Ch-4b:** FC2, P7, P8 and POz

These configurations were selected to assess the trade-off between classification performance and electrode reduction. Channel sets from previous studies (e.g., Noble et al. [199], Sahay et al. [200]) were included for comparative analysis, as they have demonstrated reliable performance in earlier P300 classification tasks.

In addition to the literature-supported configurations, several novel channel subsets were designed and evaluated to investigate alternate spatial coverage patterns that might better capture task-relevant neural activity in our specific experimental context. These custom sets were carefully constructed by retaining posterior-parietal and occipital electrode regions consistently implicated in P300 generation, while

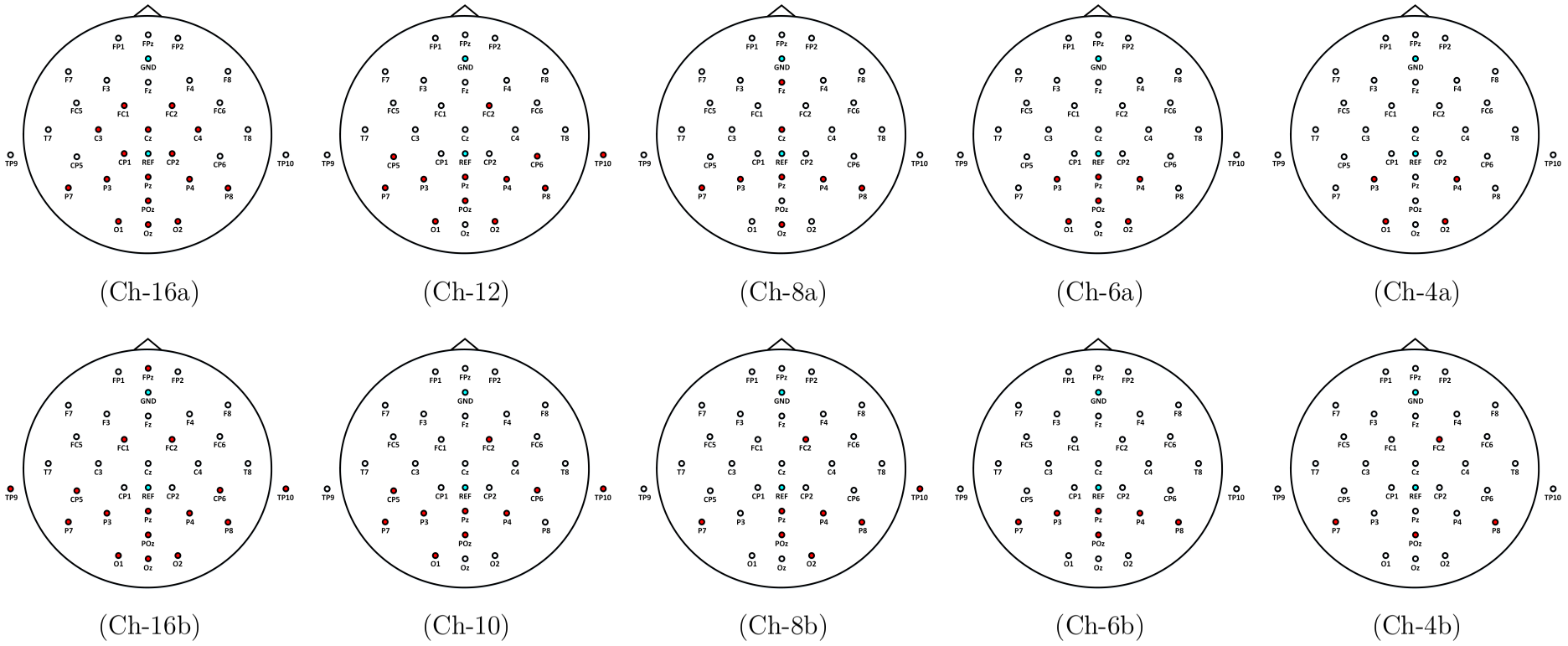


Figure 6.4: Electrode configurations for different channel subsets
Each figure corresponds to a specific configuration illustrating various combinations used for analysis. Only the selected electrodes are highlighted in red; all other electrodes remain uncolored for contrast.

varying the inclusion of lateral, central, and frontal electrodes. This exploration allows for a more comprehensive understanding of how different spatial layouts impact decoding performance. This analysis aims to identify minimal yet effective electrode subsets for real-world BCI applications, where usability, comfort, and portability are critical.

6.4 Effect of Display Modality

This section investigates the effect of display modality on the RSVP image search task, addressing research question **RQ.4d**, which explores how the use of a head-mounted display (HMD), compared to a traditional monitor, impacts EEG signals and overall BCI performance. As introduced in Chapter 3, two separate datasets were collected in this study, each involving 10 subjects. In the original AMBER dataset (see section 3.1), the RSVP task was presented on a standard monitor, while in the AMBER 2.0 dataset (see section 3.2), the same task was presented using a head-mounted display. All other aspects of the RSVP paradigm, including stimulus presentation, remained consistent across both datasets.

While most prior research has relied exclusively on monitor-based RSVP displays, as highlighted in the literature, this limits the portability and ecological validity of BCI systems in real-world contexts. By analyzing the two modalities separately, this work directly addresses that limitation and provides insight into how immersive display technologies like HMDs may support more practical and mobile BCI applications.

While previous analyses in this thesis treated both datasets collectively to maximize training diversity, this section analyzes them separately to isolate and evaluate the impact of the display modality. A Leave-One-Subject-Out configuration is still employed within each dataset, training on 9 subjects and testing on the 1 left-out subject, allowing for consistent comparison between the two setups under identical experimental conditions.

6.5 Results and Discussion

6.5.1 Effect of Training Data Size

Tables 6.1 – 6.4 present the results from the analysis of subject-independent EEG classification performance across four deep learning models: CNN-1 (Table 6.1), EEGNet (Table 6.2), EEG Conformer (Table 6.3), and EEGNet-Transformer (Table 6.4). Each table reports the average ROC-AUC scores for individual subjects when models were trained on data from 5, 10, 15, and 19 subjects, following the standard Leave-One-Subject-Out approach.

Across all four architectures, a clear decrease in accuracy was observed as the number of training subjects was reduced. Interestingly, EEGNet-Transformer consistently outperforms the other models across most training sizes, achieving the highest overall average ROC-AUC score (0.768) with 19 subjects, and maintaining competitive performance even with fewer subjects (0.760 with 15 subjects and 0.751 with 10 subjects). This suggests that the transformer-based attention mechanisms may enhance robustness to subject variability and noise, making this model particularly suitable for subject-independent EEG tasks even in low data settings.

EEGNet also shows stable and strong performance across all settings. It maintains a good balance between compact architecture and classification accuracy, especially evident with 10 and 15 training subjects, where it performs nearly on par with EEGNet-Transformer. For instance, with 15 subjects, EEGNet achieves an average AUC of 0.754 versus EEGNet-Transformer's 0.760.

In contrast, CNN-1, while effective with more data (0.763 with 19 subjects), experiences more noticeable degradation when trained with fewer subjects, dropping to 0.706 with only 5. This suggests that CNN-1, being most data-hungry, is less suited to low-data, subject-independent scenarios.

The EEG Conformer model displays a mixed pattern: while it performs strongly with 15 and 19 subjects (0.756 and 0.764), its performance degrades the most when the training data is small, achieving the lowest score (0.702) among all models with

only 5 training subjects. This could be due to the model's complexity and reliance on sufficient data to fully leverage its convolution-attention hybrid design.

Overall, EEGNet-Transformer emerges as the most robust and high-performing model across all training sizes, particularly excelling in both data-rich and data-constrained scenarios. EEGNet follows closely behind, offering a lightweight yet effective solution. CNN-1 and EEG Conformer perform well when ample data is available, but show limited generalisation with smaller subject pools.

The same trend is visually illustrated in Figure 6.5, which presents boxplots showing the distribution of ROC-AUC scores across all 20 test subjects for each model under different training set sizes. The performance pattern aligns with the earlier results, and performance consistently declines as the number of training subjects decreases. The figure also highlights variability across models, with some maintaining more stable performance under limited data conditions.

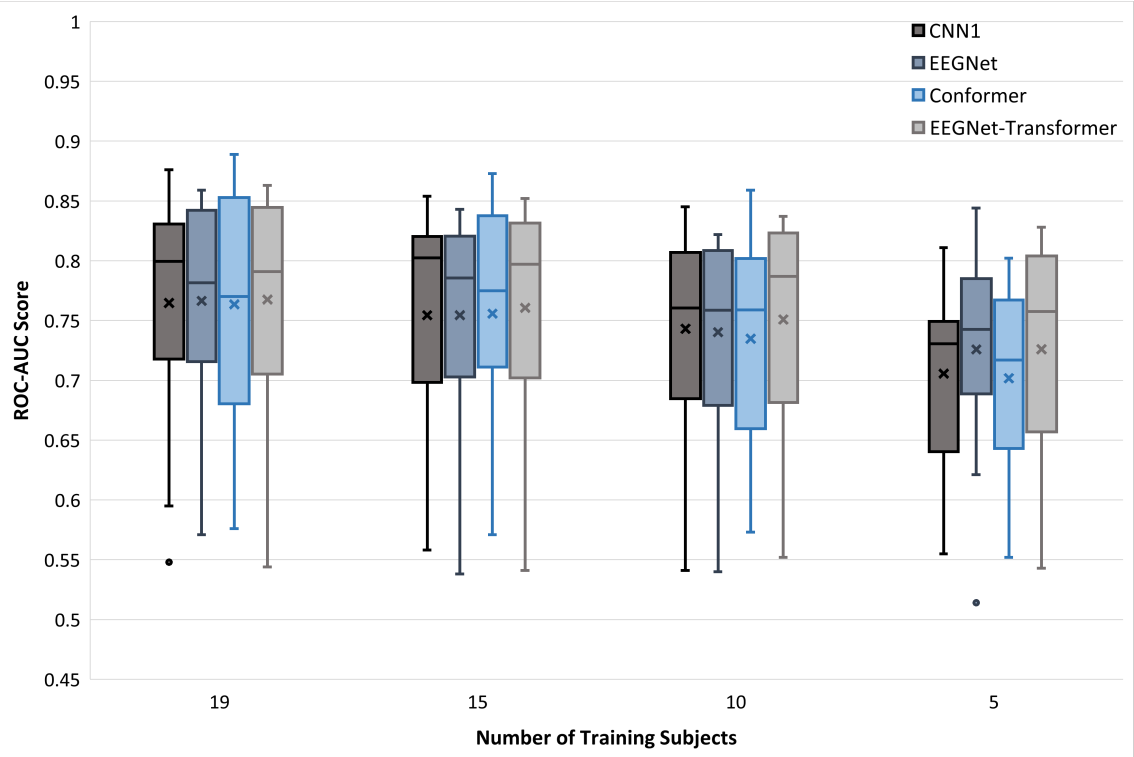


Figure 6.5: Effect of training data size on P300 classification performance. Each boxplot represents the per-subject average ROC-AUC scores distribution across 20 test subjects under leave-one-subject-out evaluation.

Table 6.1: ROC-AUC performance of CNN-1 with varying training set sizes (N=5, 10, 15, 19 subjects) using leave-one-subject-out approach on 20 total subjects

Subject	**CNN1**			
	number of training subjects			
	5	**10**	**15**	**19**
P1	0.608	0.619	0.627	0.595
P2	0.731	0.715	0.741	0.737
P3	0.631	0.675	0.628	0.678
P4	0.740	0.784	0.801	0.801
P5	0.778	0.843	0.844	0.852
P6	0.730	0.808	0.808	0.817
P7	0.668	0.750	0.742	0.747
P8	0.624	0.668	0.624	0.702
P9	0.752	0.829	0.821	0.830
P10	0.732	0.789	0.823	0.808
P11	0.811	0.824	0.842	0.876
P12	0.715	0.845	0.854	0.849
P13	0.703	0.705	0.711	0.713
P14	0.752	0.798	0.818	0.831
P15	0.769	0.771	0.804	0.804
P16	0.555	0.541	0.558	0.548
P17	0.629	0.678	0.694	0.732
P18	0.703	0.706	0.720	0.733
P19	0.740	0.804	0.816	0.844
P20	0.741	0.711	0.813	0.798
Average	**0.706**	**0.743**	**0.754**	**0.763**

Table 6.2: ROC-AUC performance of EEGNet with varying training set sizes (N=5, 10, 15, 19 subjects) using leave-one-subject-out approach on 20 total subjects

Subject	**EEGNet**			
	number of training subjects			
	5	**10**	**15**	**19**
P1	0.621	0.622	0.636	0.632
P2	0.711	0.745	0.749	0.773
P3	0.669	0.706	0.701	0.679
P4	0.768	0.795	0.787	0.785
P5	0.817	0.811	0.834	0.845
P6	0.746	0.756	0.787	0.794
P7	0.691	0.730	0.752	0.778
P8	0.643	0.658	0.654	0.671
P9	0.773	0.813	0.819	0.845
P10	0.844	0.822	0.843	0.856
P11	0.794	0.816	0.841	0.848
P12	0.789	0.801	0.821	0.834
P13	0.696	0.673	0.659	0.709
P14	0.791	0.816	0.828	0.819
P15	0.759	0.799	0.798	0.795
P16	0.514	0.540	0.538	0.571
P17	0.693	0.647	0.734	0.740
P18	0.688	0.697	0.708	0.735
P19	0.771	0.798	0.815	0.859
P20	0.739	0.761	0.784	0.761
Average	**0.726**	**0.740**	**0.754**	**0.767**

Table 6.3: ROC-AUC performance of EEG Conformer with varying training set sizes (N=5, 10, 15, 19 subjects) using leave-one-subject-out approach on 20 total subjects

Subject	**EEG Conformer**			
	number of training subjects			
	5	**10**	**15**	**19**
P1	0.586	0.624	0.629	0.613
P2	0.719	0.774	0.774	0.859
P3	0.594	0.626	0.630	0.635
P4	0.738	0.773	0.790	0.835
P5	0.802	0.824	0.853	0.857
P6	0.774	0.763	0.827	0.817
P7	0.682	0.722	0.711	0.749
P8	0.624	0.604	0.620	0.645
P9	0.679	0.723	0.751	0.674
P10	0.732	0.827	0.841	0.855
P11	0.782	0.811	0.844	0.843
P12	0.795	0.859	0.863	0.889
P13	0.707	0.697	0.739	0.720
P14	0.715	0.772	0.777	0.751
P15	0.721	0.755	0.776	0.783
P16	0.552	0.573	0.571	0.576
P17	0.670	0.713	0.711	0.700
P18	0.634	0.647	0.729	0.757
P19	0.759	0.774	0.806	0.846
P20	0.770	0.836	0.873	0.868
Average	**0.702**	**0.735**	**0.756**	**0.764**

Table 6.4: ROC-AUC performance of EEGNet-Transformer with varying training set sizes (N=5, 10, 15, 19 subjects) using leave-one-subject-out approach on 20 total subjects

Subject	**EEGNet-Transformer**			
	number of training subjects			
	5	**10**	**15**	**19**
P1	0.606	0.616	0.632	0.634
P2	0.693	0.689	0.745	0.731
P3	0.675	0.679	0.699	0.705
P4	0.755	0.806	0.813	0.811
P5	0.805	0.830	0.839	0.852
P6	0.764	0.794	0.811	0.807
P7	0.689	0.746	0.740	0.758
P8	0.612	0.662	0.665	0.693
P9	0.813	0.837	0.852	0.863
P10	0.801	0.815	0.832	0.861
P11	0.815	0.829	0.840	0.852
P12	0.828	0.826	0.843	0.837
P13	0.612	0.677	0.685	0.706
P14	0.800	0.810	0.830	0.816
P15	0.784	0.780	0.795	0.787
P16	0.543	0.552	0.541	0.544
P17	0.651	0.702	0.711	0.704
P18	0.707	0.727	0.711	0.749
P19	0.807	0.829	0.829	0.847
P20	0.760	0.810	0.799	0.795
Average	**0.726**	**0.751**	**0.760**	**0.768**

6.5.2 Fine-Tuning for Subject-specific Calibration

To evaluate subject-specific adaptation, two fine-tuning strategies, random data split and session-based adaptation, were explored across four model architectures: CNN-1, EEGNet, EEGNet-Transformer, and EEG Conformer. The ROC-AUC scores were used to quantify classification performance. Tables 6.5 – 6.8 summarise the performance of each model under the two fine-tuning protocols.

Table 6.5: ROC-AUC performance of CNN-1 with and without fine-tuning (FT), evaluated under both random-split and session-based calibration protocols.

Subject	**CNN-1**				
	No Fine Tuning	**Fine Tuning Strategy**			
		Split	**Session Based**		
		30%	**Ss1**	**Ss12**	**Ss123**
P1	0.595	0.660	0.604	0.586	0.501
P2	0.737	0.914	0.898	0.924	0.935
P3	0.678	0.749	0.766	0.780	0.804
P4	0.801	0.856	0.815	0.875	0.881
P5	0.852	0.877	0.851	0.872	0.857
P6	0.817	0.875	0.839	0.890	0.894
P7	0.747	0.790	0.750	0.755	0.772
P8	0.702	0.772	0.743	0.760	0.759
P9	0.830	0.908	0.772	0.790	0.820
P10	0.808	0.906	0.912	0.907	0.837
P11	0.876	0.892	0.861	0.826	0.789
P12	0.849	0.896	0.887	0.895	0.962
P13	0.713	0.728	0.693	0.729	0.745
P14	0.831	0.845	0.836	0.789	0.782
P15	0.804	0.800	0.807	0.793	0.763
P16	0.548	0.619	0.514	0.549	0.554
P17	0.732	0.753	0.725	0.781	0.771
P18	0.733	0.771	0.745	0.709	0.632
P19	0.844	0.846	0.859	0.832	0.932
P20	0.798	0.919	0.808	0.942	0.949
Average	**0.763**	**0.819**	**0.784**	**0.800**	**0.797**

The results demonstrate a clear and consistent improvement in model performance when fine-tuning is applied compared to models without any subject-specific adaptation. Across all architectures, fine-tuning led to higher ROC-AUC scores, underscoring its effectiveness in enhancing BCI system performance.

Without fine-tuning, the average ROC-AUC scores across all models ranged

Table 6.6: ROC-AUC performance of EEGNet with and without fine-tuning (FT), evaluated under both random-split and session-based calibration protocols.

Subject	**EEGNet**				
	No Fine Tuning	**Fine Tuning Strategy**			
		Split	**Session Based**		
		30%	**Ss1**	**Ss12**	**Ss123**
P1	0.632	0.654	0.663	0.660	0.530
P2	0.773	0.913	0.892	0.922	0.945
P3	0.679	0.803	0.750	0.792	0.781
P4	0.785	0.835	0.813	0.868	0.852
P5	0.845	0.873	0.853	0.881	0.882
P6	0.794	0.835	0.826	0.886	0.905
P7	0.778	0.813	0.711	0.715	0.705
P8	0.671	0.718	0.720	0.725	0.736
P9	0.845	0.895	0.830	0.847	0.813
P10	0.856	0.922	0.894	0.902	0.902
P11	0.848	0.846	0.809	0.779	0.723
P12	0.834	0.911	0.846	0.891	0.941
P13	0.709	0.732	0.655	0.721	0.736
P14	0.819	0.847	0.802	0.824	0.732
P15	0.795	0.801	0.780	0.778	0.751
P16	0.571	0.569	0.522	0.592	0.661
P17	0.740	0.751	0.671	0.750	0.815
P18	0.735	0.787	0.632	0.713	0.652
P19	0.859	0.855	0.795	0.821	0.855
P20	0.761	0.898	0.772	0.924	0.932
Average	**0.767**	**0.813**	**0.762**	**0.800**	**0.792**

Table 6.7: ROC-AUC performance of EEG Conformer with and without fine-tuning (FT), evaluated under both random-split and session-based calibration protocols.

Subject	**EEG Conformer**				
	No Fine Tuning	**Fine Tuning Strategy**			
		Split	**Session Based**		
		30%	**Ss1**	**Ss12**	**Ss123**
P1	0.613	0.711	0.653	0.562	0.560
P2	0.859	0.914	0.905	0.941	0.950
P3	0.635	0.841	0.707	0.732	0.840
P4	0.835	0.876	0.802	0.868	0.871
P5	0.857	0.880	0.874	0.891	0.883
P6	0.817	0.870	0.781	0.891	0.894
P7	0.749	0.826	0.778	0.787	0.779
P8	0.645	0.796	0.712	0.769	0.793
P9	0.674	0.929	0.830	0.838	0.723
P10	0.855	0.921	0.929	0.938	0.852
P11	0.843	0.861	0.827	0.800	0.730
P12	0.889	0.898	0.907	0.929	0.964
P13	0.720	0.736	0.719	0.767	0.729
P14	0.751	0.838	0.762	0.759	0.816
P15	0.783	0.793	0.787	0.766	0.659
P16	0.576	0.629	0.517	0.563	0.698
P17	0.700	0.761	0.717	0.636	0.771
P18	0.757	0.750	0.763	0.809	0.697
P19	0.846	0.862	0.836	0.839	0.907
P20	0.868	0.915	0.907	0.912	0.916
Average	**0.764**	**0.830**	**0.786**	**0.800**	**0.802**

Table 6.8: ROC-AUC performance of EEGNet-Transformer with and without fine-tuning (FT), evaluated under both random-split and session-based calibration protocols.

Subject	**EEGNet-Transformer**				
	No Fine Tuning	**Fine Tuning Strategy**			
		Split	**Session Based**		
		30%	**Ss1**	**Ss12**	**Ss123**
P1	0.634	0.680	0.617	0.571	0.460
P2	0.731	0.933	0.927	0.950	0.939
P3	0.705	0.812	0.781	0.790	0.830
P4	0.811	0.879	0.763	0.885	0.867
P5	0.852	0.898	0.683	0.888	0.877
P6	0.807	0.893	0.800	0.881	0.889
P7	0.758	0.817	0.753	0.688	0.748
P8	0.693	0.802	0.755	0.762	0.779
P9	0.863	0.906	0.835	0.846	0.817
P10	0.861	0.917	0.915	0.915	0.865
P11	0.852	0.876	0.801	0.757	0.814
P12	0.837	0.877	0.889	0.921	0.927
P13	0.706	0.726	0.705	0.731	0.729
P14	0.816	0.866	0.823	0.797	0.779
P15	0.787	0.816	0.795	0.801	0.726
P16	0.544	0.635	0.542	0.573	0.618
P17	0.704	0.771	0.774	0.795	0.823
P18	0.749	0.781	0.764	0.773	0.659
P19	0.847	0.827	0.847	0.839	0.889
P20	0.795	0.885	0.907	0.919	0.936
Average	**0.768**	**0.830**	**0.784**	**0.804**	**0.799**

between 0.763 and 0.768, indicating moderate but insufficient performance for reliable BCI operation. However, with fine-tuning, regardless of the strategy employed (random splits, Ss1, Ss12, or Ss123), the average performance improved significantly, reaching AUC values between 0.784 and 0.830. This improvement was particularly pronounced in subjects such as P2, where fine-tuning increased the AUC from 0.737 (no fine-tuning) to 0.935–0.950 (with session-based fine-tuning), and P20, where performance rose from 0.798 to as high as 0.949. These gains highlight the necessity of subject-specific adaptation to achieve optimal decoding accuracy.

Notably, the worst-performing subjects without fine-tuning (P1, P16) still exhibited improvements after fine-tuning, though their absolute performance remained lower than others.

The benefits of fine-tuning were evident even with minimal subject-specific data. For instance, fine-tuning with just Ss1 (a single session) improved the average AUC from 0.763–0.768 (no fine-tuning) to 0.762–0.786, with some subjects (P10, P19) showing near-peak performance with this limited calibration. More extensive fine-tuning (Ss12 or Ss123) further boosted performance, but the gains were often incremental, reinforcing that even small amounts of subject-specific data can yield substantial improvements.

However, the extent of improvement varied across models. While transformer-based architectures (EEG Conformer, EEGNet-Transformer) showed strong gains with minimal data, EEGNet exhibited weaker adaptation, with its average AUC staying almost the same ($0.767 \rightarrow 0.762$) and even decreasing for some subjects (e.g., P7: $0.778 \rightarrow 0.711$). This indicates that EEGNet may require more calibration data for reliable subject-specific tuning.

This suggests that while fine-tuning consistently enhances generalisability compared to non-adapted models, the final performance ceiling depends on three key factors: (1) the amount of fine-tuning data available, (2) the specific architecture's adaptability to the amount of calibration data, and (3) inherent subject variability. Transformer-based models demonstrated robust performance even with minimal

fine-tuning, while EEGNet showed sensitivity to data scarcity, with its gains diminishing when only one session was available.

Moreover, individual differences persisted across all conditions - some subjects (e.g., P10, P19) reached near-optimal performance with just Ss1, while others (e.g., P1, P16) remained challenging cases regardless of fine-tuning amount or model choice. These results emphasize that while increasing fine-tuning data (from Ss1 to Ss12/Ss123) generally improves performance, the relationship is non-linear and mediated by both model architecture and subject-specific factors, suggesting that optimal calibration strategies should consider this three-way interaction in real-world BCI deployments.

- **Random Fine-Tuning Yields Higher Accuracy but Less Realistic Evaluation**

As shown in Tables 6.5 – 6.8, the random split strategy consistently outperformed all session-based variants across all four models. Average ROC-AUC scores for CNN-1 (0.819), EEGNet (0.813), EEG Conformer (0.830), and EEGNet-Transformer (0.830) were the highest under random sampling. This performance gain, however, must be interpreted with caution.

The inflated scores from random splitting can be attributed to temporal leakage and autocorrelation in EEG data. EEG signals are inherently non-independent and temporally structured. When data are randomly sampled from across all sessions, trials that are temporally or contextually similar often end up in both the training and testing sets. This leakage artificially simplifies the classification task, as the model encounters highly correlated examples during testing that it has already learned from during fine-tuning.

In addition, random splits preserve the underlying data distribution while masking critical sources of neural variability inherent to real-world usage. By pooling data across sessions, this approach artificially minimises the impact of natural signal fluctuations, including amplitude variations, latency shifts, and morphological changes, caused by factors such as evolving mental states, fatigue, head movement,

or electrode impedance drift. These variations alter the signal characteristics across recording sessions but become homogenized through random sampling, effectively obscuring the very challenges that make real-world BCI applications difficult.

Consequently, while random fine-tuning demonstrates high accuracy, it provides an overly optimistic estimate of generalisation performance by failing to account for both the temporal non-stationarity and session-to-session variability that fundamentally challenge operational BCI systems. The resulting models may excel at classifying carefully curated data splits but lack robustness against the dynamic neural patterns encountered in actual deployment scenarios.

- **Session-Based Fine-Tuning: A Realistic Yet Challenging Paradigm**

In contrast, session-based fine-tuning provides a more ecologically valid evaluation. This approach mimics realistic deployment scenarios, where a user performs an initial calibration session (e.g., Session 1), and the trained system is expected to function in subsequent sessions without repeated calibration. However, this strategy consistently led to lower performance compared to the random split. The average ROC-AUC scores dropped by 0.03-0.05 points (CNN-1: 0.784, EEGNet: 0.762, EEG Conformer: 0.786, EEGNet-Transformer: 0.784), highlighting the challenges introduced by temporal and physiological variability across sessions.

The performance degradation is primarily due to session-to-session variability in EEG signals. Each session introduces shifts in brain dynamics and recording conditions, such as changes in user attention, emotional state, or even subtle variations in cap placement. Since the fine-tuning and testing sessions are strictly separated in this setting, no temporal leakage is possible, and the model must generalise to previously unseen temporal contexts, making the task more difficult; however, more realistic.

Despite this challenge, session-based strategies offer a more trustworthy estimate of model performance over time. Importantly, they reflect the operational realities of BCIs in real-world environments, where calibration burden must be minimised, and long-term performance consistency is essential.

Figure 6.6 illustrates the average ROC-AUC performance for each model under four fine-tuning strategies: random data split and session-based fine-tuning using one (Ss1), two (Ss12), and three (Ss123) sessions.

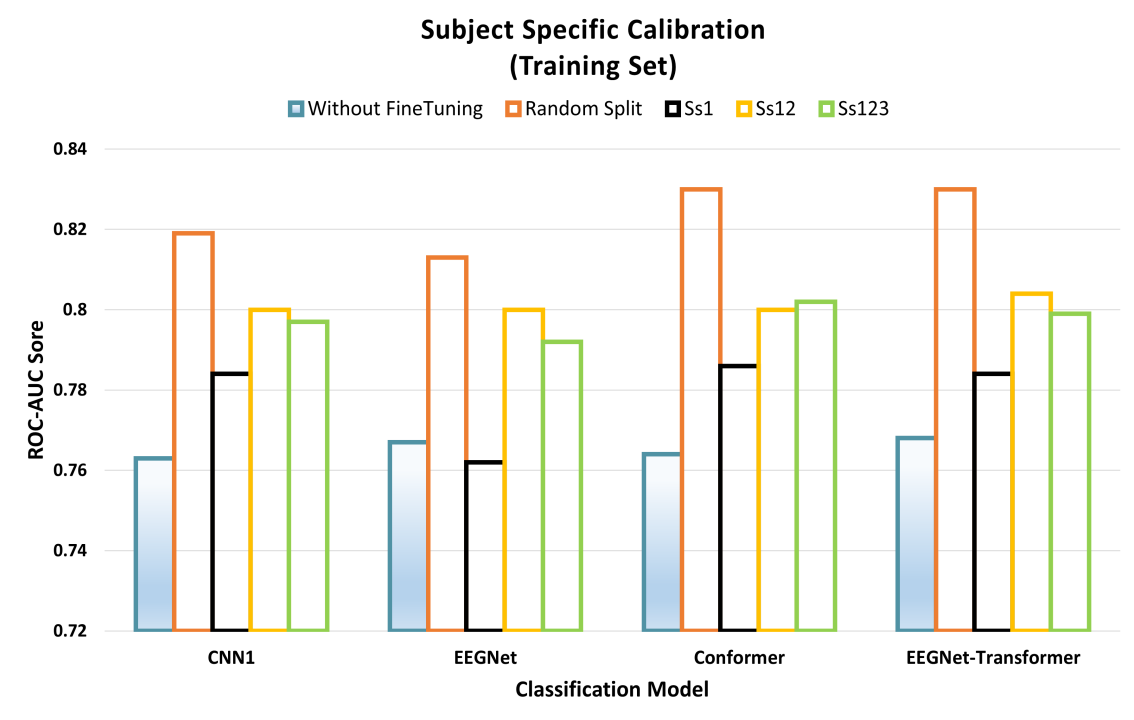


Figure 6.6: Comparison of average ROC-AUC scores across different fine-tuning strategies.

- **Optimal Calibration Trade-Off: Two Sessions May Be Sufficient**

A key finding from this analysis is that increasing the number of sessions used for fine-tuning improves performance to a certain extent, but gains tend to saturate beyond two sessions. Specifically, fine-tuning using only Session 1 (Ss1) consistently resulted in the lowest performance across all models. The average ROC-AUC scores were relatively poor; CNN-1: 0.784, EEGNet: 0.762, EEG Conformer: 0.786, and EEGNet-Transformer: 0.784. This suggests that relying on a single session for calibration is insufficient, likely due to the inability to capture the full variability in EEG signal dynamics and task engagement across time.

In contrast, fine-tuning with two sessions (Ss12) led to an improved performance across all models; CNN-1: 0.800, EEGNet: 0.800, EEG Conformer: 0.800, EEGNet-Transformer: 0.804.

These results indicate that incorporating data from a second session enhances the model's ability to generalise to unseen sessions by capturing a broader range of neural and noise-related variability. Importantly, adding a third session (Ss123) yielded only marginal or inconsistent gains, with some models even showing slight performance drops; CNN-1: 0.797, EEGNet: 0.792, EEG Conformer: 0.802, EEGNet-Transformer: 0.799.

This variance in performance implies that while a second session is necessary to boost generalisability, a third session does not offer statistically meaningful improvements. It is possible that adding more data beyond a certain point introduces redundant or noisy information, or that diminishing returns are reached due to the model already capturing the most informative inter-session variability with two sessions.

Overall, these findings highlight that calibration using only a single session is insufficient for reliable BCI operation. However, two-session fine-tuning (Ss12) strikes an optimal balance between calibration effort and model performance, offering a practical trade-off for real-world BCI systems. This insight is valuable for designing user-friendly calibration protocols, especially in applications where time, fatigue, or attention span are limiting factors.

In order to validate the claim that using two sessions for fine-tuning offers an optimal balance between calibration effort and model performance, a series of paired t-tests was conducted, comparing the classification scores obtained under different fine-tuning configurations: Ss1 (Session 1 only), Ss12 (Sessions 1 and 2), and Ss123 (Sessions 1 to 3). To ensure a robust and model-agnostic statistical analysis, the classification results across all four models were pooled together, yielding 80 samples for each session configuration.

The results revealed a statistically significant improvement when comparing Ss1 to Ss12, $t(79) = -4.07$, $p < 0.001$, indicating that including a second session for fine-tuning leads to significantly better model performance than relying on a single session. In contrast, the comparison between Ss12 and Ss123 yielded no significant

difference, t(79)= 0.53, p= 0.598, suggesting that adding a third session does not provide a meaningful benefit. However, comparing Ss1 directly to Ss123 still showed a statistically significant improvement, t(79)= -2.18, p= 0.032, although the effect size was smaller than that observed between Ss1 and Ss12.

These findings empirically support the earlier observation of this work that using only one session for fine-tuning results in suboptimal generalisation, while fine-tuning with two sessions provides sufficient adaptation without unnecessary overhead. The lack of further improvement with a third session confirms that Ss12 offers the best trade-off between performance and practicality for real-world BCI deployment.

- **Noise Adaptation**

To explore real-world adaptation of BCI systems, fine-tuning experiments were extended by incorporating noisy data into the calibration process. In the earlier analysis using clean (traditional RSVP) data, different fine-tuning strategies were evaluated for subject-specific calibration. The results indicated that session-based fine-tuning using the initial two sessions (i.e., Ss12) provided a balanced trade-off, offering a significant performance boost while avoiding unnecessary calibration effort, as including a third session did not yield further improvements. Building on this, the same configuration was adopted to fine-tune models using intentionally contaminated data, simulating real-world conditions. Specifically, subject-specific fine-tuning was performed using calibration data contaminated by three different artifact categories: body movement, head movement, and talking. This approach allows for evaluating how effective real-world noisy calibration data can be in adapting subject-independent models for robust BCI performance.

The results in Figure 6.7 reflect the model performances on contaminated data before and after applying fine-tuning (FT) calibration using sessions Ss1 and Ss2. Across all three noise categories, consistent performance improvements were observed following fine-tuning. This confirms the importance and effectiveness of personalized calibration, even when the calibration data itself contains real-world noise.

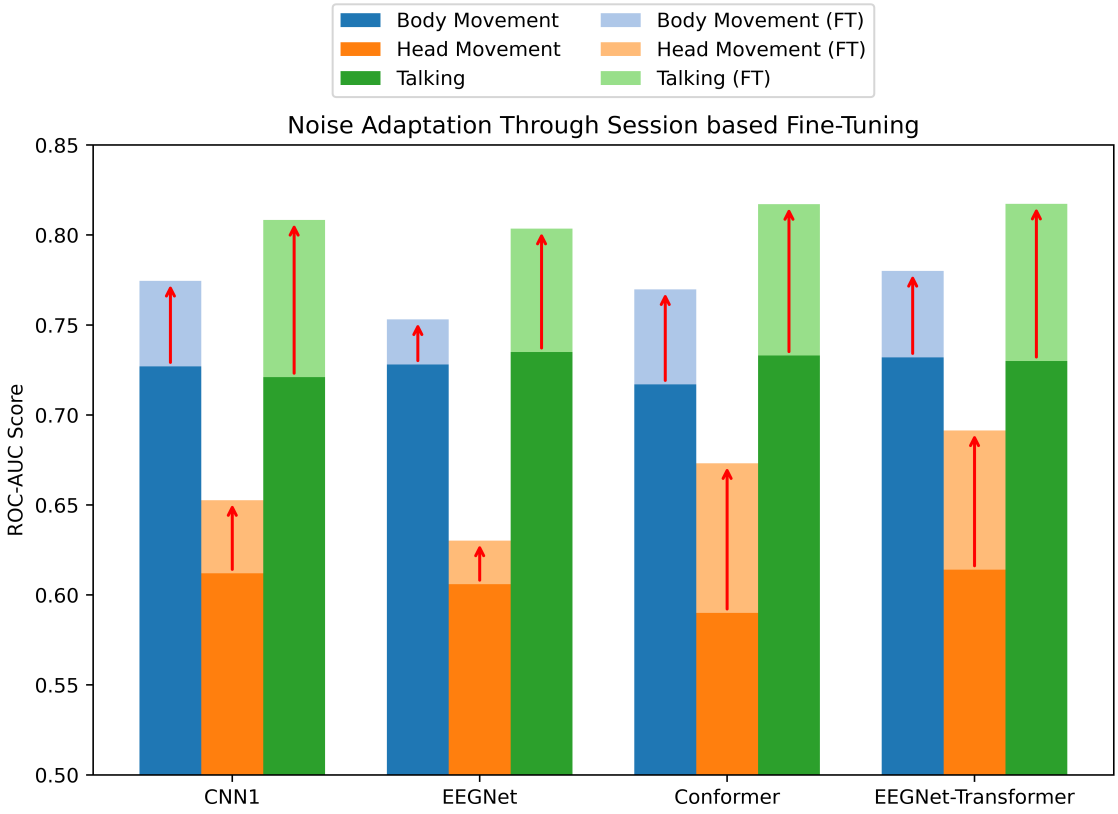


Figure 6.7: Finetune calibration with noisy data.

Among the three artifact types, head movement proves to be the most disruptive, showing the lowest baseline AUC scores across all models (ranging from 0.59 to 0.614). Despite this, fine-tuning significantly enhances performance, with AUCs rising close to or above 0.69 in most cases. This improvement is particularly encouraging, as head movement artifacts are common in practical, mobile BCI scenarios. The proposed models, along with the effective fine-tuning strategy, demonstrate a meaningful ability to adapt to such challenging conditions.

Talking and body movement artifacts are comparatively less severe. For both, strong baseline performance is observed (e.g., up to 0.735 AUC for EEGNet under talking noise), with further improvements post fine-tuning (e.g., 0.81735 AUC for EEGNet-Transformer). The consistent boost after FT underscores the critical role of even short calibration sessions in adapting pre-trained models to individual users and environmental noise, reflecting realistic BCI usage.

Across all models, EEGNet-Transformer stands out by delivering the highest or

near-highest scores in each noise category, both before and after fine-tuning. This reaffirms its robustness not only in clean environments but also in noisy, real-world settings, further solidifying its role as a strong candidate for deployable BCIs.

Overall, these findings highlight the practical feasibility of using noisy calibration data to personalize subject-independent models. With only two short calibration sessions, the models adapt effectively, marking a step forward toward truly adaptive and usable BCIs in naturalistic, everyday environments.

6.5.3 Optimal Channels

The evaluation of reduced-channel configurations across four architectures reveals several critical insights about the trade-offs between performance and practicality in EEG-based BCIs. While the default 32-channel setup consistently delivered the highest average performance across all models (AUC: 0.763–0.768), certain reduced-channel configurations achieved competitive results, suggesting that careful channel selection can significantly minimise hardware complexity without substantially compromising decoding accuracy.

The CNN-1 architecture demonstrated significant sensitivity to channel reduction, particularly for subjects with complex activation patterns. For subject P2 (high performer), the 32-channel configuration achieved 0.737 AUC, while even the best reduced configuration (Ch-16b) showed a 4.3% performance drop. However, P9 exhibited remarkable resilience, maintaining 0.87 AUC with Ch-16b (only 4.8% reduction from 32-channel). The steepest declines were observed in frontal-dominant subjects like P3, where Ch-16a (excluding frontal electrodes) led to a 14.7% reduction.

The 32-channel configuration achieved the highest average performance (0.763 AUC), with my custom Ch-16b emerging as the closest competitor at 0.740 AUC (97% of full-channel performance). Notably, Noble's Ch-16a (0.697 AUC) demonstrated that simply matching channel count isn't sufficient - my Ch-16b's parietal-occipital focus provided an 8.3% improvement over Ch-16a despite identical channel

numbers. Performance declined sharply below 12 channels: Ch-12 (0.721 AUC), Ch-10 (0.724 AUC), with the most dramatic drops occurring at minimal configurations - Ch-4a (0.594 AUC) and Ch-4b (0.623 AUC) representing just 78-82% of full-channel performance. The 8-channel configurations revealed my Ch-8b (0.702 AUC) outperformed Sahay's Ch-8a (0.665 AUC) by 5.6%, validating the temporal-parietal electrode selection strategy.

Table 6.9: Channel set comparison: ROC-AUC scores for CNN-1.

Sub	**CNN-1 – Optimal Channel Identification**										
	Ch-32	**Ch-16a**	**Ch-16b**	**Ch-12**	**Ch-10**	**Ch-8a**	**Ch-8b**	**Ch-6a**	**Ch-6b**	**Ch-4a**	**Ch-4b**
1	0.595	0.590	0.627	0.634	0.596	0.559	0.618	0.574	0.555	0.517	0.572
2	0.737	0.701	0.737	0.706	0.708	0.637	0.714	0.628	0.606	0.663	0.547
3	0.678	0.578	0.621	0.671	0.602	0.540	0.671	0.594	0.543	0.499	0.555
4	0.801	0.671	0.766	0.731	0.704	0.690	0.700	0.637	0.586	0.634	0.620
5	0.852	0.776	0.830	0.816	0.811	0.687	0.811	0.702	0.746	0.603	0.704
6	0.817	0.760	0.791	0.731	0.749	0.718	0.729	0.723	0.743	0.711	0.704
7	0.747	0.711	0.672	0.690	0.684	0.619	0.653	0.671	0.623	0.577	0.572
8	0.702	0.724	0.715	0.716	0.690	0.644	0.684	0.599	0.571	0.659	0.590
9	0.830	0.758	0.870	0.859	0.826	0.784	0.859	0.812	0.772	0.714	0.729
10	0.808	0.710	0.732	0.669	0.737	0.607	0.488	0.565	0.594	0.479	0.545
11	0.876	0.812	0.824	0.803	0.786	0.797	0.770	0.755	0.758	0.672	0.697
12	0.849	0.806	0.804	0.786	0.828	0.654	0.759	0.741	0.694	0.691	0.694
13	0.713	0.594	0.705	0.666	0.727	0.688	0.667	0.514	0.573	0.580	0.531
14	0.831	0.731	0.811	0.812	0.838	0.686	0.792	0.586	0.696	0.528	0.705
15	0.804	0.779	0.791	0.791	0.793	0.725	0.757	0.699	0.719	0.678	0.687
16	0.548	0.512	0.505	0.483	0.483	0.544	0.510	0.517	0.613	0.449	0.560
17	0.732	0.556	0.644	0.651	0.657	0.612	0.659	0.559	0.595	0.440	0.646
18	0.733	0.625	0.726	0.681	0.677	0.664	0.680	0.582	0.692	0.532	0.638
19	0.844	0.760	0.806	0.763	0.784	0.695	0.751	0.582	0.662	0.504	0.638
20	0.798	0.792	0.833	0.764	0.799	0.756	0.765	0.484	0.604	0.738	0.519
Avg	**0.763**	**0.697**	**0.740**	**0.721**	**0.724**	**0.665**	**0.702**	**0.626**	**0.647**	**0.594**	**0.623**

Despite its simpler architecture, EEGNet showed surprising efficiency in some cases. Subject P19 retained 95.3% of full performance using Ch-16b (0.818 vs. 0.859 AUC). However, temporal-lobe-dependent subjects like P7 suffered dramatically with channel reduction (22.4% drop with Ch-4a). Interestingly, Noble's 6a configuration worked exceptionally well for P6 (0.721 AUC) while failing for P10 (0.565), highlighting subject-specific optimal configurations.

The full 32-channel setup set the benchmark at 0.767 AUC, with Ch-16b again proving most effective among reduced configurations (0.747 AUC, 97.4% retention). Interestingly, while Ch-16a (0.695 AUC) underperformed, the 12-channel setup (0.734 AUC) maintained 95.7% performance, suggesting EEGNet can effectively utilize focused channel sets. Below this threshold, degradation accelerated:

Ch-10 (0.726 AUC), Ch-8b (0.709 AUC), reaching minimums at Ch-4a (0.601 AUC) - a 21.6% drop from full configuration. The 6-channel results showed Noble's Ch-6a (0.658 AUC) slightly outperformed my Ch-6b (0.651 AUC), suggesting occipital coverage may be more critical than parietal for EEGNet.

Table 6.10: Channel set comparison: ROC-AUC scores for EEGNet.

Sub	**EEGNet – Optimal Channel Identification**										
	Ch-32	**Ch-16a**	**Ch-16b**	**Ch-12**	**Ch-10**	**Ch-8a**	**Ch-8b**	**Ch-6a**	**Ch-6b**	**Ch-4a**	**Ch-4b**
1	0.632	0.619	0.639	0.647	0.645	0.568	0.615	0.581	0.531	0.491	0.556
2	0.773	0.648	0.747	0.753	0.735	0.644	0.688	0.604	0.580	0.670	0.628
3	0.679	0.635	0.646	0.674	0.673	0.621	0.698	0.626	0.577	0.547	0.527
4	0.785	0.703	0.761	0.741	0.741	0.668	0.689	0.672	0.622	0.632	0.554
5	0.845	0.782	0.818	0.793	0.805	0.807	0.785	0.619	0.761	0.661	0.679
6	0.794	0.758	0.778	0.744	0.754	0.703	0.719	0.721	0.747	0.699	0.649
7	0.778	0.692	0.730	0.711	0.690	0.603	0.712	0.661	0.658	0.621	0.640
8	0.671	0.659	0.713	0.734	0.721	0.668	0.661	0.661	0.609	0.641	0.639
9	0.845	0.754	0.845	0.832	0.816	0.779	0.851	0.774	0.695	0.695	0.673
10	0.856	0.709	0.736	0.652	0.744	0.644	0.582	0.687	0.644	0.523	0.727
11	0.848	0.801	0.798	0.823	0.786	0.782	0.796	0.759	0.713	0.686	0.658
12	0.834	0.805	0.819	0.819	0.796	0.694	0.724	0.770	0.719	0.699	0.718
13	0.709	0.601	0.710	0.655	0.675	0.630	0.632	0.568	0.588	0.550	0.575
14	0.819	0.768	0.844	0.842	0.827	0.653	0.772	0.647	0.696	0.535	0.744
15	0.795	0.785	0.797	0.785	0.760	0.705	0.750	0.728	0.720	0.675	0.719
16	0.571	0.472	0.561	0.506	0.495	0.541	0.557	0.527	0.573	0.431	0.618
17	0.740	0.593	0.641	0.659	0.622	0.581	0.666	0.563	0.624	0.464	0.664
18	0.735	0.618	0.745	0.725	0.693	0.612	0.737	0.642	0.676	0.601	0.621
19	0.859	0.779	0.818	0.795	0.754	0.731	0.808	0.715	0.663	0.533	0.718
20	0.761	0.714	0.804	0.792	0.788	0.692	0.734	0.629	0.622	0.658	0.557
Avg	**0.767**	**0.695**	**0.747**	**0.734**	**0.726**	**0.666**	**0.709**	**0.658**	**0.651**	**0.601**	**0.643**

The EEG Conformer model exhibited the most stable performance across channel reductions, with only 2.4% average degradation from 32 to 16 channels. Subject P14 showed exceptional performance retention (0.862 AUC with Ch-16b vs 0.835 32-channel), likely due to the model's ability to attend to critical parietal regions. However, frontal-lobe-dependent tasks (P13) still suffered 15.3% performance loss with Ch-12. The 8-channel configurations revealed an interesting contrast - while Ch-8a (centered on midline electrodes) worked well for P15 (0.706 AUC), my temporal-parietal Ch-8b better served P18 (0.712 vs 0.664).

With the most stable channel scaling, EEG Conformer's 32-channel baseline (0.764 AUC) saw minimal degradation to Ch-16b (0.746 AUC, 97.6% retention). The model maintained exceptional performance down to 12 channels (0.720 AUC, 94.2%), outperforming EEGNet by 3.5% at this level. The 8-channel configurations revealed Ch-8b (0.697 AUC) was superior to Ch-8a (0.670 AUC), particularly

Table 6.11: Channel set comparison: ROC-AUC scores for EEG Conformer.

Sub	EEG Conformer – Optimal Channel Identification										
	Ch-32	**Ch-16a**	**Ch-16b**	**Ch-12**	**Ch-10**	**Ch-8a**	**Ch-8b**	**Ch-6a**	**Ch-6b**	**Ch-4a**	**Ch-4b**
1	0.613	0.614	0.625	0.613	0.597	0.560	0.600	0.544	0.563	0.536	0.596
2	0.859	0.725	0.840	0.786	0.692	0.695	0.708	0.611	0.687	0.618	0.678
3	0.635	0.647	0.622	0.603	0.635	0.592	0.623	0.612	0.514	0.560	0.566
4	0.835	0.685	0.720	0.751	0.752	0.697	0.722	0.662	0.588	0.582	0.576
5	0.857	0.735	0.837	0.790	0.792	0.642	0.779	0.552	0.668	0.602	0.604
6	0.817	0.768	0.777	0.668	0.753	0.746	0.688	0.640	0.679	0.593	0.683
7	0.749	0.527	0.756	0.686	0.686	0.653	0.641	0.609	0.586	0.571	0.630
8	0.645	0.646	0.697	0.666	0.664	0.612	0.685	0.644	0.660	0.593	0.592
9	0.674	0.695	0.772	0.797	0.689	0.692	0.741	0.557	0.771	0.599	0.741
10	0.855	0.685	0.732	0.580	0.631	0.711	0.490	0.678	0.626	0.548	0.680
11	0.843	0.774	0.807	0.729	0.779	0.699	0.736	0.713	0.635	0.703	0.668
12	0.889	0.822	0.877	0.867	0.848	0.679	0.819	0.731	0.770	0.693	0.764
13	0.720	0.651	0.779	0.750	0.714	0.680	0.708	0.579	0.611	0.552	0.628
14	0.751	0.771	0.862	0.852	0.834	0.696	0.782	0.646	0.706	0.565	0.763
15	0.783	0.696	0.759	0.771	0.744	0.706	0.755	0.646	0.701	0.616	0.719
16	0.576	0.600	0.547	0.535	0.526	0.580	0.524	0.570	0.648	0.504	0.576
17	0.700	0.584	0.613	0.658	0.609	0.640	0.622	0.634	0.614	0.523	0.602
18	0.757	0.691	0.800	0.739	0.745	0.708	0.712	0.669	0.687	0.584	0.635
19	0.846	0.738	0.668	0.753	0.661	0.677	0.794	0.679	0.703	0.554	0.664
20	0.868	0.733	0.835	0.814	0.808	0.730	0.817	0.663	0.718	0.723	0.673
Avg	**0.764**	**0.689**	**0.746**	**0.720**	**0.708**	**0.670**	**0.697**	**0.632**	**0.657**	**0.591**	**0.652**

for temporal-lobe tasks (7.1% improvement). At extreme reductions, Ch-4b (0.652 AUC) surprisingly outperformed Ch-4a (0.591 AUC) by 10.3%, suggesting the model leverages temporal-parietal connections more effectively than pure occipital coverage.

The EEGNet-Transformer demonstrated the most graceful degradation, maintaining >90% of 32-channel performance down to 12 channels for 75% of subjects. Subject P9's performance actually improved with Ch-16b (0.886 vs 0.863), suggesting the transformer's attention mechanism can better utilize focused channel sets. However, subjects with diffuse activation patterns (P16) still required more channels (minimum 6 for acceptable performance). The 4-channel configurations showed the largest variance, with P12 maintaining 0.737 AUC (Ch-4b) while P17 dropped to 0.453.

Overall, this hybrid architecture showed remarkable resilience, with 32-channel (0.768 AUC) to Ch-16b (0.749 AUC) representing just a 2.5% drop - the smallest degradation among all models. Performance remained strong at Ch-12 (0.727 AUC, 94.7%) and even Ch-8b (0.712 AUC, 92.7%), demonstrating the transformer attention mechanism's ability to compensate for spatial sparsity. The 4-channel

Table 6.12: Channel set comparison: ROC-AUC scores for EEGNet-Transformer.

Sub	EEGNet-Transformer – Optimal Channel Identification										
	Ch-32	**Ch-16a**	**Ch-16b**	**Ch-12**	**Ch-10**	**Ch-8a**	**Ch-8b**	**Ch-6a**	**Ch-6b**	**Ch-4a**	**Ch-4b**
1	0.634	0.638	0.632	0.635	0.625	0.583	0.653	0.563	0.548	0.507	0.590
2	0.731	0.631	0.738	0.740	0.723	0.634	0.713	0.505	0.552	0.656	0.587
3	0.705	0.625	0.660	0.660	0.685	0.619	0.709	0.594	0.539	0.573	0.540
4	0.811	0.696	0.734	0.725	0.731	0.656	0.698	0.629	0.580	0.652	0.614
5	0.852	0.818	0.821	0.801	0.803	0.752	0.783	0.675	0.743	0.642	0.687
6	0.807	0.767	0.758	0.724	0.735	0.736	0.730	0.712	0.730	0.685	0.674
7	0.758	0.675	0.725	0.749	0.743	0.655	0.679	0.678	0.680	0.598	0.615
8	0.693	0.680	0.726	0.704	0.689	0.641	0.699	0.663	0.610	0.629	0.540
9	0.863	0.762	0.886	0.832	0.839	0.783	0.822	0.803	0.781	0.751	0.654
10	0.861	0.700	0.739	0.639	0.740	0.597	0.518	0.684	0.638	0.514	0.667
11	0.852	0.819	0.817	0.769	0.774	0.754	0.772	0.771	0.711	0.683	0.710
12	0.837	0.824	0.834	0.824	0.811	0.625	0.756	0.731	0.734	0.678	0.737
13	0.706	0.645	0.711	0.678	0.686	0.619	0.663	0.551	0.564	0.596	0.594
14	0.816	0.734	0.855	0.831	0.829	0.698	0.802	0.655	0.741	0.531	0.739
15	0.787	0.783	0.788	0.786	0.764	0.718	0.749	0.727	0.718	0.671	0.730
16	0.544	0.515	0.503	0.495	0.511	0.536	0.517	0.561	0.582	0.476	0.578
17	0.704	0.516	0.666	0.622	0.617	0.607	0.681	0.542	0.577	0.453	0.665
18	0.749	0.670	0.764	0.719	0.709	0.664	0.710	0.620	0.715	0.567	0.665
19	0.847	0.753	0.818	0.775	0.754	0.743	0.789	0.670	0.624	0.539	0.659
20	0.795	0.728	0.805	0.830	0.804	0.728	0.804	0.647	0.672	0.660	0.601
Avg	**0.768**	**0.699**	**0.749**	**0.727**	**0.729**	**0.667**	**0.712**	**0.649**	**0.652**	**0.603**	**0.642**

configurations exhibited the largest model-specific variance: Ch-4b (0.642 AUC) outperformed Ch-4a (0.603 AUC) by 6.5%, contrasting with CNN-1's preference for Ch-4a. This suggests the attention mechanism redistributes focus differently than convolutional filters.

- **Cross-Model Insights and Practical Recommendations**

Notably, the 16-channel configurations (i.e., Ch-16b) emerged as particularly robust, nearly matching the 32-channel performance (average AUC drop: ≤ 0.023 across models). For instance, in EEGNet-Transformer, Ch-16b achieved an average AUC of 0.749 compared to 0.768 for 32 channels, demonstrating that halving the number of channels can retain 97% of the original performance. This trend held for most subjects, with exceptions like P10 and P16, where performance drops were more pronounced, likely due to their reliance on discarded frontal channels. The success of Ch-16b over literature-based Ch-16a (AUC: +0.05 on average) underscores the value of task-specific channel optimisation.

Further reduction to 12–10 channels introduced only marginal additional declines (average AUC: 0.721–0.734), suggesting these configurations may offer a "sweet

spot" for applications demanding minimal setup time. Transformer-based models, particularly EEG Conformer, maintained an AUC of 0.72 with Ch-12, indicating that transformer-based models can effectively leverage sparse spatial sampling. However, performance variability increased among subjects, highlighting that individual neuroanatomical differences may necessitate adaptive channel selection.

Below 10 channels, performance degradation became more pronounced, particularly for traditional CNNs (CNN-1, EEGNet), which struggled with very sparse configurations (e.g., EEGNet's AUC dropped by 0.106 from 32 to 4 channels). In contrast, transformer-based models (EEG Conformer, EEGNet-Transformer) exhibited greater resilience to channel reduction, especially with occipito-parietal coverage (e.g., Ch-8b preserved 93% of 32-channel performance in EEGNet-Transformer). This aligns with their ability to capture long-range dependencies, compensating for lost spatial resolution.

The most extreme reductions (Ch-4a/b) showed significant variability across subjects. While some (e.g., P2, P15) retained reasonable performance (AUC: 0.65–0.73), others (e.g., P16, P17) dropped below 0.45 AUC. This contrast suggests that ultra-sparse configurations may be feasible only for specific user subgroups or paradigms, necessitating pre-screening.

This analysis demonstrates that although 32-channel configurations offer optimal performance, a well-designed 16-channel subset, specifically Ch-16b can retain >95% of classification accuracy across diverse models, offering an effective trade-off between usability and decoding power. Transformer-based architectures, particularly EEGNet-Transformer and EEG Conformer, exhibit strong potential for low-channel scenarios, maintaining high performance even at $\leq$8 channels where traditional CNNs fail. However, three critical limitations must be acknowledged:

1. **Subject-Specific Variability:** Optimal channel configuration varies across individuals. For example, temporal lobe dominance in one subject doesn't guarantee similar patterns in others, necessitating personalized calibration.

2. **Task Dependence:** Channel sets optimised for P300 paradigms may be un-

suitable for other BCI tasks like motor imagery or SSVEP, etc., which engage different neural circuits.

3. **Non-Linear Spatial Effects:** The removal of seemingly redundant electrodes can unpredictably disrupt model performance, particularly in subjects with diffuse or atypical neural activation.

Figure 6.8 presents a bar chart comparing the average ROC-AUC scores across various EEG channel configurations and four different models. The visualisation highlights that strategically selected reduced-channel sets can retain competitive classification performance. This supports the feasibility of more practical and wearable BCI implementations without incurring substantial accuracy loss.

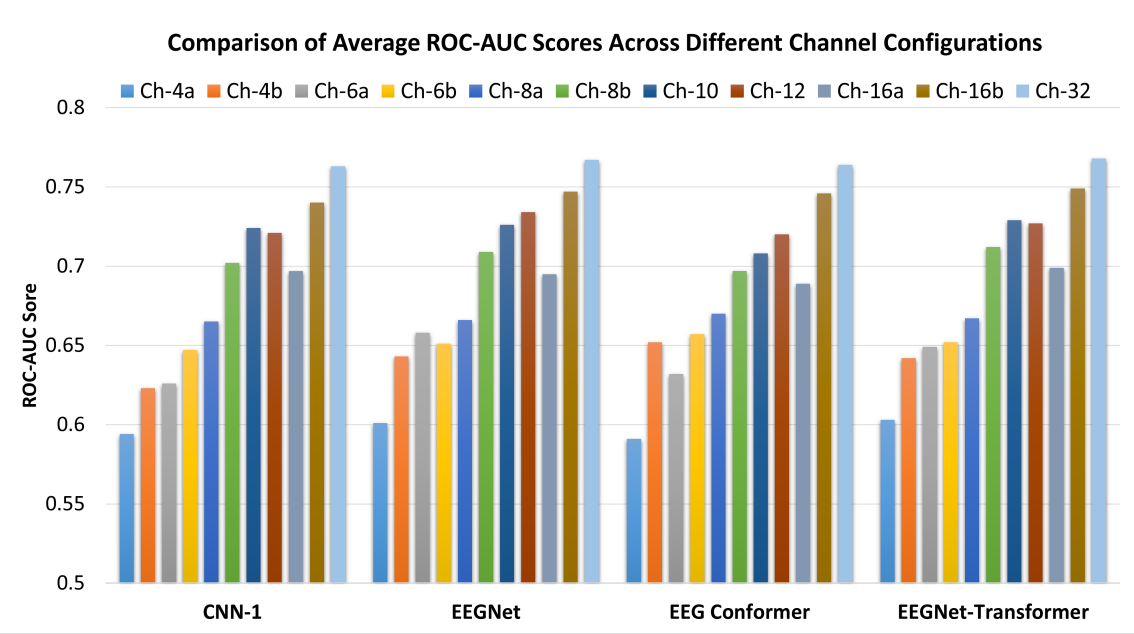


Figure 6.8: Bar chart showing the average ROC-AUC performance for different channel subsets.

To support the analysis on optimal channel selection and ensure statistically robust, model-independent conclusions, a comprehensive statistical evaluation was conducted. The ROC-AUC scores across all four models were pooled together, yielding 80 samples for each channel configuration (excluding the default 32-channel configuration). A Friedman test indicated a significant difference in ROC-AUC scores among the ten channel configurations, $\chi^2(9) = 401.51$, $p < 0.001$. To further inves-

tigate these differences, post-hoc Wilcoxon signed-rank tests with Holm-Bonferroni correction were conducted. The pairwise significance results are summarised visually using a heatmap in the -10log (p-value) scale, where more intense colors reflect stronger statistical significance (see Figure 6.9).

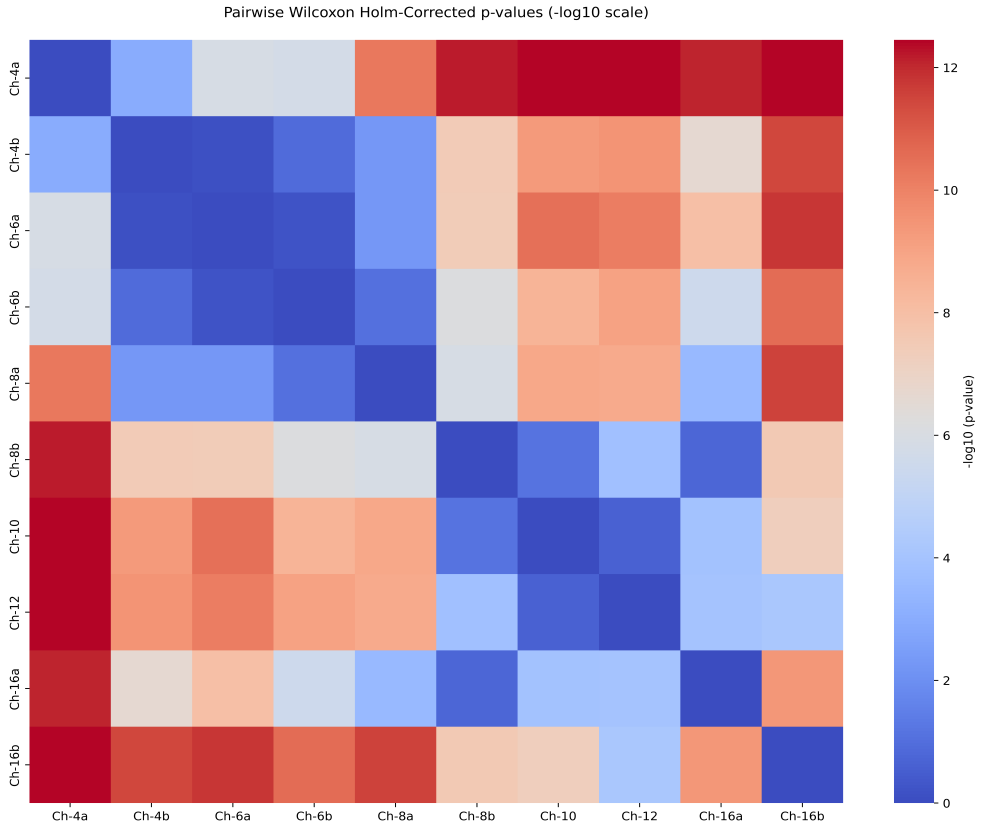


Figure 6.9: Heatmap of post-hoc Wilcoxon signed-rank test results comparing ROC-AUC scores across all pairwise combinations of the ten channel configurations (The default 32 channel configuration excluded).

The results revealed that Ch-4a [199] consistently performed significantly worse than all other configurations, with significant differences observed in every pairwise comparison. This strongly suggests that Ch-4a is suboptimal for classification performance. While Ch-4b showed modest improvement over Ch-4a, it still performed significantly worse than several higher-performing configurations, particularly Ch-8b, Ch-10, Ch-12, and Ch-16b, though it was not significantly different from Ch-6a and Ch-6b, both of which showed moderate classification performance.

The average performers, Ch-6a [199], Ch-6b, and Ch-8a [200] were not significantly different from each other. However, both were significantly outperformed

by Ch-10, Ch-12, and Ch-16b. Among all configurations, Ch-16b emerged as the top performer, showing significantly higher ROC-AUC scores than nearly all other channel groups, including its counterpart Ch-16a [199], highlighting its potential as the most effective configuration for this classification task.

6.5.4 Effect of Display Modality

The results across the two display modalities are presented in Tables 6.13 – 6.20, which report the average ROC-AUC scores for the four selected models under a Leave-One-Subject-Out cross-validation scheme. For clarity, the tables are divided by modality: Tables 6.13 – 6.16 correspond to the Monitor display condition, while Tables 6.17 – 6.20 present the results for the Head-Mounted Display condition.

Overall, the findings suggest that both display methods yield comparable and promising classification performance, with no drastic drop or gain in overall average ROC-AUC across models. Specifically, for the HMD dataset, transformer-based models (i.e., EEGNet-Transformer and EEG Conformer) performed best, achieving average ROC-AUC scores of 0.7693 and 0.7706, respectively. EEGNet and CNN-1 also performed competitively, both slightly trailing behind with scores of 0.7554 and 0.7562. In contrast, the monitor-based dataset showed similar trends, where EEGNet-Transformer again performed best (0.7636), followed closely by EEGNet (0.7606) and EEG Conformer (0.7523), with CNN-1 at 0.7396.

EEGNet-Transformer consistently performed well across both display modalities, achieving the highest or near-highest average ROC-AUC scores in each case. The other models also delivered competitive results, with only slight variations in performance. These differences, however, are not substantial, indicating that all models remained robust regardless of the display modality used. This suggests that HMDs can be effectively used as an alternative to traditional monitors in RSVP-based BCI tasks without compromising classification accuracy, thereby supporting their potential utility in portable and immersive real-world BCI applications.

To further extend this analysis and better simulate real-world BCI usage, noisy

data conditions, specifically, body movement, head movement, and talking, were introduced alongside the clean (lab-controlled) data. The goal was to assess how each modality withstands common sources of EEG contamination. Intensionally-Contaminated RSVP results (Tables 6.13 – 6.20) reveal that noise significantly impacts classification performance across all models, with the greatest drop observed under head movement conditions. For instance, in the HMD setup, the average AUC for EEGNet-Transformer decreased from 0.7693 (traditional RSVP) to 0.6067 (head movement), a drop of roughly 16.2%. In comparison, the monitor dataset showed a smaller reduction from 0.7636 to 0.6067, around 15.7% for the same condition. While the absolute differences are small, the HMD setup generally exhibited a slightly larger performance degradation under all three noise types.

These findings suggest that while both display modalities demonstrate strong potential for real-world BCI applications, monitor-based systems may offer slightly more stable performance under noisy conditions. The HMD setup, though effective, appears somewhat more sensitive to movement-related artifacts, particularly during head movement, which can affect classification accuracy. Nonetheless, the overall performance of HMD-based systems remains comparable, highlighting their promise for enabling portable and immersive BCI solutions. With thoughtful system design and user-centric considerations, HMDs can serve as a compelling alternative to traditional monitors, supporting the development of more flexible and accessible BCI technologies suited for dynamic, real-world environments.

6.6 Practical BCI Design Recommendations

As brain-computer interfaces move from controlled laboratory experiments toward real-world adoption, there is increasing interest in their integration into diverse application domains such as assistive technologies, neurorehabilitation, cognitive monitoring, gaming, and immersive human-computer interaction. For these applications, practical deployment requires systems that are not only accurate and fast but also resilient to noise, minimally intrusive, and easy to set up and use in every-

Table 6.13: Results of CNN-1 on the AMBER Dataset (Monitor Display).

Subject	**AMBER dataset – CNN-1**			
	RSVP	**IC-RSVP**		
		Body movement	**Head movement**	**Talking**
P1	0.599	0.583	0.578	0.676
P2	0.762	0.758	0.646	0.729
P3	0.627	0.636	0.518	0.702
P4	0.834	0.839	0.593	0.829
P5	0.827	0.789	0.518	0.786
P6	0.827	0.706	0.596	0.704
P7	0.698	0.699	0.639	0.656
P8	0.598	0.662	0.621	0.674
P9	0.853	0.817	0.753	0.752
P10	0.771	0.790	0.552	0.562
Average	**0.740**	**0.728**	**0.601**	**0.707**

Table 6.14: Results of EEGNet on the AMBER Dataset (Monitor Display).

Subject	**AMBER dataset – EEGNet**			
	RSVP	**IC-RSVP**		
		Body movement	**Head movement**	**Talking**
P1	0.605	0.607	0.525	0.703
P2	0.732	0.789	0.723	0.736
P3	0.749	0.746	0.557	0.769
P4	0.825	0.802	0.570	0.806
P5	0.860	0.805	0.568	0.855
P6	0.803	0.676	0.616	0.695
P7	0.698	0.668	0.604	0.641
P8	0.636	0.596	0.612	0.693
P9	0.831	0.832	0.748	0.781
P10	0.867	0.818	0.638	0.728
Average	**0.761**	**0.734**	**0.616**	**0.741**

Table 6.15: Results of EEG Conformer on the AMBER Dataset (Monitor Display).

Subject	AMBER dataset – EEG Conformer			
	RSVP	IC-RSVP		
		Body movement	Head movement	Talking
P1	0.583	0.578	0.543	0.668
P2	0.787	0.746	0.695	0.780
P3	0.680	0.708	0.544	0.749
P4	0.838	0.791	0.588	0.823
P5	0.868	0.810	0.570	0.833
P6	0.817	0.665	0.620	0.696
P7	0.618	0.649	0.585	0.704
P8	0.659	0.542	0.513	0.616
P9	0.823	0.822	0.715	0.776
P10	0.850	0.742	0.571	0.540
Average	**0.752**	**0.705**	**0.594**	**0.719**

Table 6.16: Results of EEGNet-Transformer on the AMBER Dataset (Monitor Display).

Subject	AMBER dataset – EEGNet-Transformer			
	RSVP	IC-RSVP		
		Body movement	Head movement	Talking
P1	0.627	0.607	0.539	0.704
P2	0.780	0.783	0.658	0.751
P3	0.712	0.666	0.540	0.704
P4	0.834	0.835	0.583	0.821
P5	0.836	0.836	0.556	0.826
P6	0.787	0.734	0.611	0.741
P7	0.760	0.667	0.592	0.694
P8	0.636	0.630	0.648	0.665
P9	0.844	0.797	0.749	0.751
P10	0.820	0.797	0.591	0.606
Average	**0.764**	**0.735**	**0.607**	**0.726**

Table 6.17: Results of CNN-1 on the AMBER 2.0 Dataset (HMD Display).

Subject	**AMBER 2.0 dataset – CNN-1**			
	RSVP	**IC-RSVP**		
		Body movement	**Head movement**	**Talking**
P1	0.849	0.882	0.658	0.889
P2	0.698	0.658	0.580	0.541
P3	0.734	0.681	0.530	0.743
P4	0.832	0.769	0.718	0.853
P5	0.794	0.650	0.603	0.642
P6	0.550	0.576	0.509	0.560
P7	0.736	0.669	0.539	0.605
P8	0.708	0.696	0.636	0.694
P9	0.844	0.811	0.674	0.766
P10	0.817	0.793	0.602	0.784
Average	**0.756**	**0.719**	**0.605**	**0.708**

Table 6.18: Results of EEGNet on the AMBER 2.0 Dataset (HMD Display).

Subject	**AMBER 2.0 dataset – EEGNet**			
	RSVP	**IC-RSVP**		
		Body movement	**Head movement**	**Talking**
P1	0.819	0.873	0.656	0.837
P2	0.780	0.716	0.557	0.608
P3	0.742	0.674	0.534	0.759
P4	0.823	0.773	0.692	0.857
P5	0.792	0.643	0.596	0.668
P6	0.583	0.637	0.520	0.605
P7	0.737	0.652	0.559	0.600
P8	0.657	0.702	0.637	0.733
P9	0.837	0.821	0.680	0.712
P10	0.784	0.750	0.645	0.721
Average	**0.755**	**0.724**	**0.608**	**0.710**

Table 6.19: Results of EEG Conformer on the AMBER 2.0 Dataset (HMD Display).

Subject	AMBER 2.0 dataset – EEG Conformer			
	RSVP	IC-RSVP		
		Body movement	Head movement	Talking
P1	0.838	0.798	0.567	0.832
P2	0.871	0.757	0.602	0.722
P3	0.756	0.661	0.530	0.747
P4	0.851	0.767	0.691	0.773
P5	0.795	0.598	0.605	0.660
P6	0.547	0.565	0.521	0.555
P7	0.760	0.577	0.545	0.680
P8	0.678	0.661	0.606	0.710
P9	0.767	0.757	0.611	0.672
P10	0.843	0.763	0.599	0.712
Average	**0.771**	**0.690**	**0.588**	**0.706**

Table 6.20: Results of EEGNet-Transformer on the AMBER 2.0 Dataset (HMD Display).

Subject	AMBER 2.0 dataset – EEGNet-Transformer			
	RSVP	IC-RSVP		
		Body movement	Head movement	Talking
P1	0.823	0.851	0.639	0.854
P2	0.840	0.779	0.581	0.756
P3	0.725	0.675	0.514	0.764
P4	0.854	0.807	0.705	0.830
P5	0.768	0.609	0.609	0.662
P6	0.547	0.596	0.518	0.576
P7	0.735	0.625	0.554	0.606
P8	0.728	0.714	0.653	0.706
P9	0.845	0.816	0.696	0.757
P10	0.831	0.797	0.647	0.785
Average	**0.770**	**0.727**	**0.612**	**0.730**

day environments. This work offers several insights and guidelines to support that transition.

While the analysis primarily focused on RSVP-based P300 signals, many of the conclusions, particularly regarding model robustness, data efficiency, and hardware considerations, are likely to generalise across similar ERP-based paradigms. Among the evaluated models, EEGNet-Transformer consistently emerged as the most effective for real-world noisy conditions, offering strong performance in both clean and artifact-contaminated data. Its transformer-based attention mechanisms appear particularly well-suited to handling signal variability and noise, making it a compelling option for subject-independent applications in practical settings.

Another clear finding is that more training data improves subject-independent model performance, enhancing generalisability across users. However, when large-scale data is unavailable, transformer-based models still demonstrate acceptable robustness, highlighting their adaptability for low-data scenarios where broad user coverage is required.

From a hardware standpoint, channel reduction is both feasible and beneficial for practical use. The experiments showed that a custom 16-channel configuration strikes a strong balance between signal quality and usability, preserving most of the performance of full 32-channel setups. If reducing complexity is a priority, 10–12 channels can still offer decent performance, though some trade-offs may occur. Importantly, optimal channel sets can vary across individuals and tasks, so personalisation or adaptive channel selection strategies should be considered in deployment.

Subject-specific calibration through fine-tuning also proves valuable. Even with noisy EEG data, fine-tuning using just two task-specific sessions significantly improved performance across models. This reduces the user burden and increases feasibility for real-world use, where long calibration times are impractical.

Finally, the findings show that both traditional monitor-based and HMD setups can support effective RSVP-based BCIs. HMDs, while slightly more sensitive to motion artifacts, offer the advantage of portability and immersive interaction with-

out a major performance trade-off. This flexibility broadens the potential use cases for BCIs, supporting mobile, ecologically valid applications beyond the lab.

6.7 Summary

- **Training Size**

To address research question **(RQ.4a)**, which explores the impact of training data volume on subject-independent P300 classification performance, the number of training subjects was systematically varied under a Leave-One-Subject-Out framework. Starting with the full set of 19 subjects, the training pool was progressively reduced to 15, 10, and 5 subjects. For each reduced setting, subjects were randomly selected and the process was repeated five times to ensure that the results were not biased by any specific subject combination.

The analysis revealed a consistent performance degradation across all models as the number of training subjects decreased. This trend underscores the importance of data diversity and quantity in enhancing generalisability across unseen subjects. Despite the decline in performance with reduced data, transformer-based models, especially EEGNet-Transformer, demonstrated notable robustness. EEGNet-Transformer consistently outperformed other models in both high-data and low-data scenarios, highlighting its ability to generalise well even when training data is scarce.

These findings confirm the scalability of transformer-based architectures and their potential for developing reliable BCI systems in settings where large-scale data collection may not be feasible. This is particularly valuable for real-world applications where access to large datasets is often limited.

- **FineTuning**

Addressing research question **(RQ.4b)**, this work systematically evaluates subject-specific calibration strategies for EEG-based BCIs, addressing two critical questions: which fine-tuning approach is most effective, and how much calibration data suffices

for real-world deployment. Random data splitting (idealized but impractical) was compared with session-based fine-tuning (real-world feasible) across four neural architectures. While random splits yielded superior results, session-based calibration using just two sessions (50% of available data) achieved comparable performance, establishing a practical benchmark for clinical implementation. This finding is particularly valuable as it demonstrates that extended calibration periods may be unnecessary, reducing user burden without compromising accuracy.

Extending the analysis to noisy real-world conditions, the two-session strategy was validated across three noise categories: body movements, head movements, and talking. Fine-tuning consistently improved performance in different conditions. This confirms that fine-tuning not only personalizes the model to neural patterns but also enhances robustness against real-world disturbances.

Notably, the EEGNet-Transformer proved particularly resilient, maintaining high performance in both clean and noisy settings. Its ability to leverage just two noisy calibration sessions for meaningful adaptation underscores its suitability for practical BCI deployment. These results highlight a critical insight: even imperfect, noise-contaminated calibration data can effectively tailor models to individual users, compensating for both individual variability and environmental interference.

- **Channels**

Addressing research question **(RQ.4c)**, this chapter presents a systematic investigation of optimal EEG channel configurations for practical BCI applications, evaluating performance across four neural network architectures (CNN-1, EEGNet, EEG Conformer, and EEGNet-Transformer) using progressively sparser channel sets from the standard 32-channel configuration down to minimal 4-channel setups. The analysis reveals that carefully designed 16-channel configurations, particularly my custom Ch-16b set, can preserve over 95% of the full-channel performance (0.740-0.749 AUC versus baseline 0.763-0.768 AUC), demonstrating that significant channel reduction is possible without substantial accuracy loss. Transformer-based architectures show particular robustness to spatial sparsity, maintaining more than 90%

accuracy even with only 8 channels, while traditional CNNs require at least 12 channels for comparable performance. Below this 12-channel threshold, the relationship between channel count and accuracy becomes increasingly nonlinear, with disproportionate performance degradation occurring for certain subject-model combinations.

The findings carry important practical implications while revealing fundamental limitations of channel reduction approaches. While my custom configurations consistently outperformed literature-derived sets at equivalent channel counts, the study uncovered substantial subject-specific variability in optimal channel configurations, with different individuals showing distinct spatial activation patterns. Furthermore, the effectiveness of specific channel sets varies significantly across different BCI paradigms, meaning configurations optimised for P300 detection may perform poorly for other tasks like motor imagery. The research also identified hard limits to spatial undersampling, with all models showing greater than 25% accuracy loss when reduced below 6 channels for certain subjects. These results suggest a systematic implementation approach: the Ch-16b configuration serves as an excellent default starting point that balances performance and practicality, while transformer-based systems may consider further reduction to 8-12 channels after careful subject-specific validation.

- **HMD vs Monitor**

Addressing research question **RQ.4d**, this study investigated the impact of display modality, specifically, traditional monitor versus head-mounted display, on the performance of RSVP-based EEG classification in BCI systems. Two datasets were used: the AMBER dataset, which employed a standard monitor, and the AMBER 2.0 dataset, which utilized an HMD for RSVP presentation. Both datasets included EEG recordings from 10 subjects each, and identical experimental protocols were followed, apart from the display modality. A Leave-One-Subject-Out cross-validation strategy was applied to evaluate the generalisability of four classification models: CNN-1, EEGNet, EEG Conformer, and EEGNet-Transformer.

Results showed that both display modalities yielded promising classification performance, with EEGNet-Transformer consistently achieving high ROC-AUC scores in both setups. Although slight variations were observed, for instance, EEGNet performed marginally better on the monitor data, while CNN-1 and EEG Conformer slightly favored the HMD, these differences were not substantial. This indicates that both display modalities are viable for RSVP-based BCIs, with HMDs offering additional portability and immersive benefits for real-world deployment.

To examine practical feasibility, the models were also tested on noisy EEG data collected under body movement, head movement, and talking conditions. Performance declined across all noise types, with head movement introducing the most severe degradation. Notably, the HMD setup showed slightly greater sensitivity to noise, particularly from head motion, likely due to its physical design. However, the performance remained acceptable, further supporting the potential of HMDs for portable BCI systems.

Chapter 7

Conclusion and Future Work

This thesis has been motivated by the need to bridge the gap between high-performing BCI systems developed under ideal laboratory conditions and their effective deployment in real-world environments. Specifically, it investigates the development of noise-resilient EEG-based BCIs through the lens of single-trial classification, subject-independent generalisation, minimal calibration requirements, and robustness to behavioral and environmental noise. The findings have broad implications for improving the usability, reliability, and practical viability of BCI systems in naturalistic contexts, where artifacts are unavoidable, yet system performance must remain stable and effective.

To address this research problem, a comprehensive approach was adopted that encompasses experimental design, systematic data collection under varying conditions, extensive machine learning evaluations, and critical analysis of robustness across multiple factors. The work is grounded in the real-world challenges facing BCI deployment and aims to deliver both scientific insight and practical recommendations for next-generation BCI development.

The major contributions of this thesis are as follows:

1. A systematic protocol was developed to collect EEG data under varied environmental conditions and display modalities (AMBER and AMBER 2.0), enabling controlled comparisons of laboratory and real-world settings.

2. Noise was found to significantly degrade EEG signal quality, with head movement causing the strongest disruptions among the tested artifact conditions.

3. Across a wide range of models, a hybrid Multiband EEGNet-Transformer achieved the best performance in both clean and noisy conditions, outperforming traditional and baseline deep learning approaches.

4. Session-based fine-tuning demonstrated that using as little as two sessions of subject-specific data notably improved performance, enhancing real-world adaptability of subject-independent models.

5. A direct relationship between training data volume and performance was revealed, with larger datasets and greater subject diversity improving generalisation across individuals.

6. A reduced EEG channel set of 16 electrodes preserved about 95% of performance compared to the full 32-channel configuration, supporting the development of lightweight, wearable BCI systems.

7. The use of head-mounted displays showed no significant differences from monitor-based presentation, indicating their potential as a practical and portable solution for real-world BCI applications.

A chapter-wise recap of the thesis is provided below to summarise the key focus and contributions of each chapter.

7.1 Systematic EEG Data Collection – Chapter 3

As part of this thesis's core contributions (Chapter 3), the AMBER and AMBER 2.0 datasets were systematically designed and collected to address a key gap in BCI research: the lack of ecologically valid, noise-inclusive EEG datasets. AMBER, based on a traditional monitor display, provides a controlled setup with synchronized video for multimodal analysis, while AMBER 2.0 advances toward real-world use by

employing head-mounted displays and introducing behavioral artifacts to simulate practical conditions.

Together, these datasets enable cross-condition evaluation, support the development of robust machine learning approaches, and serve as benchmarks for understanding the impact of environmental context, task modality, and user behavior on BCI performance. They provide essential resources for advancing deployable, noise-resilient BCIs.

7.2 Impact of Noise – Chapter 4

The work presented in Chapter 4 addressed the foundational challenges of deploying RSVP-based P300 BCI systems in real-world conditions, focusing on the impact of behavioral artifacts and the generalisability of models trained under clean versus noisy settings. Specifically, it investigated **Research Question 1 (RQ.1)**, the effect of real-world behaviors such as body movement, head movement, and talking on EEG signals, and **Research Question 2 (RQ.2)**, whether models trained on noisy EEG can generalise effectively to artifact-contaminated conditions. These questions aligned with the broader objective of systematically examining the practical viability of traditional BCIs beyond controlled laboratory environments.

Noise was found to significantly degrade EEG signal quality, with head movement causing the strongest disruptions among the tested artifact conditions. Across experiments, models trained and tested on clean data consistently outperformed those involving noisy or mixed-quality inputs. Notably, while training on both clean and noisy data sometimes improved robustness when tested on similarly noisy data, it did not yield reliable performance gains overall. These findings underscore the dual challenge: (1) noise in testing data degrades model predictions, and (2) the inclusion of noise in training data does not guarantee improved generalisation. Furthermore, high rates of trial rejection, especially in head movement conditions, demonstrated that traditional denoising strategies may result in substantial data loss, making them impractical for real-world deployment.

7.3 Subject Independent Benchmarking – Chapter 5

One of the central aims of this thesis, aligned with the objective "to evaluate the effectiveness of different machine learning approaches for subject-independent and noise-resilient P300 detection" (presented in Chapter 5), was to explore and advance subject-independent classification strategies, particularly under noisy, real-world conditions. While most BCI studies focus on subject-specific classification, these approaches are impractical for real-world use due to their heavy calibration requirements. To enable scalable, plug-and-play BCIs, this work systematically evaluated a broad range of machine learning and deep learning approaches using a rigorous LOSO framework. Specifically, this chapter responds to **Research Question 3 (RQ.3)**: Which advanced machine learning models are most effective in achieving subject-independent P300 detection robust to real-world noise?

A wide spectrum of models was evaluated and the findings revealed that models incorporating spatial filtering, temporal attention, and frequency-specific processing, most notably the EEGNet-Transformer with multiband processing, appear to deliver better performance across both clean and noisy test conditions. This suggests that carefully designed hybrid architectures are more effective than either traditional methods or deep models used in isolation.

However, it was also found that more complex models, especially those with multiband and MoE enhancements, come with increased computational costs, raising practical considerations for real-time or embedded deployment.

7.4 Practical Refinements and Design Considerations for Real-World BCIs – Chapter 6

The work presented in Chapter 6 explores essential practical considerations for the deployment of BCIs in real-world environments, addressing **Research Question 4**

(RQ.4): What are the key practical refinements and design considerations needed to transition BCI systems from lab-controlled environments to effective real-world applications?

The objective was to identify practical refinements and constraints that guide the real-world deployment of BCI systems with minimal performance loss under noisy, constrained, and variable conditions. To address this broader research question, the following four critical factors were systematically investigated:

- What is the impact of the amount of training data from different subjects on the performance of subject-independent BCI models? **(RQ.4a)**

To address RQ.4a, I evaluated model generalisability in a LOSO setting by varying the number of training subjects (19, 15, 10, and 5). The results showed a clear decline in ROC-AUC performance as the training set size decreased, highlighting the critical role of training data volume. Importantly, it is not only the quantity of data but also the diversity of subjects that matters: including more individuals exposes the model to a broader range of neural patterns and inter-subject variability, which enhances its ability to generalise to unseen users in real-world settings.

- How much subject-specific calibration data is required to effectively adapt subject-independent models for improved performance in the presence of noise? **(RQ.4b)**

To address RQ.4b, two fine-tuning strategies were compared: (i) random splitting of available data, providing an optimistic upper bound, and (ii) session-based adaptation, which better reflects realistic deployment. While random splitting unsurprisingly achieved the highest gains, session-based fine-tuning with only two sessions yielded nearly equivalent improvements.

Importantly, these insights were further validated under noisy conditions, including body movement, head movement, and talking artifacts. In all scenarios, fine-tuning, even on noisy data, consistently improved classification performance.

This underscores the value of modest subject-specific calibration in boosting both personalisation and noise robustness.

- What is the optimal number and combination of EEG channels required to maintain high classification accuracy while minimising complexity in real-world BCI applications? **(RQ.4c)**

To address RQ.4c, I systematically evaluated a broad range of channel configurations, including both those reported in the literature and newly proposed combinations. The results showed that the Ch-16b configuration offered the best trade-off, retaining over 95% of the performance of the full 32-channel setup while halving the number of electrodes. This demonstrates the feasibility of reducing hardware complexity without substantially compromising accuracy. However, it is important to note that channel selection is highly task-dependent: while Ch-16b proves effective for RSVP-P300, the same configuration may not generalize to other BCI paradigms such as motor imagery or SSVEP.

- How does the use of a head-mounted display (HMD) for RSVP presentation, compared to a traditional monitor, affect the EEG signals and BCI performance in both clean and noisy conditions? **(RQ.4d)**

To address RQ.4d, a comparative analysis was performed using two matched datasets: AMBER (monitor display) and AMBER 2.0 (HMD display). The results revealed that both modalities supported reliable P300-based BCI classification, with models achieving broadly comparable performance in both clean and noisy conditions. These findings demonstrate that HMD-based RSVP paradigms are not only feasible but also hold strong potential for advancing mobile and ecologically valid BCI applications, offering greater flexibility for real-world deployment without sacrificing performance.

7.5 Future Work

The following future work directions are proposed to advance the development of subject-independent, noise-resilient, and practically deployable BCI systems.

7.5.1 Datasets and Paradigms

Future work should expand to larger and more diverse datasets that incorporate richer real-world noise types and additional BCI paradigms beyond RSVP-P300. This will enable broader validation and ensure that insights generalize across tasks such as motor imagery or SSVEP, strengthening the ecological validity of the findings.

7.5.2 Noise Handling

Developing adaptive noise-aware architectures remains a key priority. Instead of relying on trial rejection or handcrafted preprocessing, lightweight denoising mechanisms embedded directly within models could enhance robustness while preserving data volume. Exploring multimodal sensor fusion and artifact-specific handling strategies (e.g., motion vs. speech artifacts) may further improve resilience under naturalistic conditions.

7.5.3 Subject-Independent Adaptation

Improving generalisation across users requires adaptive models capable of handling inter-subject and cross-condition variability. Promising avenues include smarter calibration protocols, such as dynamic session lengths tailored to user performance, as well as hybrid approaches combining lightweight fine-tuning with unsupervised domain adaptation. Leveraging subject similarity metrics and EEG-specific augmentation could also boost performance under data-limited scenarios.

7.5.4 Real-World Deployment

For practical BCI use, models must balance accuracy with efficiency. Future work should explore model compression techniques and dynamic channel selection tailored to task and user neuroanatomy. HMD-based paradigms further require attention to artifact mitigation, user comfort, and real-time feasibility. Ultimately, combining algorithmic advances with hardware co-design will be critical for enabling mobile, plug-and-play BCI systems.

7.6 Closing Statement

This thesis has explored the challenges and opportunities involved in transitioning RSVP-based P300 BCI systems from controlled laboratory environments to real-world, noise-prone applications. Through systematic investigations into the impact of behavioral artifacts, subject-independent classification, model robustness, channel optimisation, and display modalities, the work has laid a strong foundation for building more practical and generalisable BCI systems. The findings not only reveal the extent to which noise and individual variability affect system performance but also underscore the importance of architecture choice, training strategies, and careful channel selection in mitigating these effects. Importantly, this research offers concrete design recommendations, such as the use of transformer-based models, inclusion of task-relevant noisy data, and the adoption of compact yet effective electrode layouts, for the development of noise-resilient and portable BCIs.

While the primary focus was on RSVP-based P300 detection, the insights and methodologies developed here hold broader relevance for other BCI paradigms. The challenges identified, such as data loss during denoising, limited subject generalisability, and training under low-SNR conditions, point toward future avenues benefiting from advances in self-supervised learning, adaptive calibration, and robust representation learning. Ultimately, this thesis moves the field closer to practical, real-world BCIs that are both technically sound and usable in diverse settings.

Bibliography

[1] J. R. Wolpaw, "Brain-computer interfaces (bcis) for communication and control," in *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, 2007, pp. 1–2.

[2] J. R. Wolpaw, "Brain–computer interfaces," in *Handbook of clinical neurology*, vol. 110, Elsevier, 2013, pp. 67–74.

[3] W. O. Tatum IV, *Handbook of EEG interpretation*. Springer Publishing Company, 2021.

[4] M. Schmoigl-Tonis, C. Schranz, and G. R. Müller-Putz, "Methods for motion artifact reduction in online brain-computer interface experiments: A systematic review," *Frontiers in Human Neuroscience*, vol. 17, 2023.

[5] F. Lotte, A. Van Langhenhove, F. Lamarche, *et al.*, "Exploring large virtual environments by thoughts using a brain–computer interface based on motor imagery and high-level commands," *Presence: teleoperators and virtual environments*, vol. 19, no. 1, pp. 54–70, 2010.

[6] M. Welter, A. Bouneau, F. Lotte, and T. E. Ward, "Towards curating personalized art exhibitions in virtual reality with multimodal brain-computer-interfaces," in *VSAC 2022-Visual Science of Art*, 2022.

[7] M. Welter, J. C. Martínez, E. Redmond, J. Baum, T. Ward, and F. Lotte, "Eeg single-trial decoding of visual art preference," in *Graz BCI Conference*, 2024.

[8] M. Welter, M. Ahsan, T. Ward, and F. Lotte, "Challenges towards automated art presentation brain-computer-interfaces," in *Visual Science of Art Conference*, vol. 13, 2024.

[9] Z. Wang, G. Healy, A. F. Smeaton, and T. E. Ward, "Spatial filtering pipeline evaluation of cortically coupled computer vision system for rapid serial visual presentation," *Brain-Computer Interfaces*, vol. 5, no. 4, pp. 132–145, 2018.

[10] G. Healy, T. E. Ward, C. Gurrin, and A. F. Smeaton, "Overview of ntcir-13 nails task," in *The 13th NTCIR Conference on Evaluation of Information Access Technologies*, Tokyo, Japan, 2017.

[11] N. Tiwari, D. R. Edla, S. Dodia, and A. Bablani, "Brain computer interface: A comprehensive survey," *Biologically inspired cognitive architectures*, vol. 26, pp. 118–129, 2018.

[12] D. Yadav, S. Yadav, and K. Veer, "A comprehensive assessment of brain computer interfaces: Recent trends and challenges," *Journal of Neuroscience Methods*, vol. 346, p. 108 918, 2020.

[13] J. J. Shih, D. J. Krusienski, and J. R. Wolpaw, "Brain-computer interfaces in medicine," in *Mayo clinic proceedings*, Elsevier, vol. 87, 2012, pp. 268–279.

[14] S. K. Loo and S. Makeig, "Clinical utility of eeg in attention-deficit/hyperactivity disorder: A research update," *Neurotherapeutics*, vol. 9, no. 3, pp. 569–587, 2012.

[15] N. Ferrazoli, C. Donadon, A. Rezende, P. H. Skarzynski, and M. D. Sanfins, "The application of p300-long-latency auditory-evoked potential in parkinson disease," *International archives of otorhinolaryngology*, vol. 26, no. 01, pp. 158–166, 2022.

[16] M. Tao, J. Sun, S. Liu, *et al.*, "An event-related potential study of p300 in preschool children with attention deficit hyperactivity disorder," *Frontiers in Pediatrics*, vol. 12, p. 1 461 921, 2024.

[17] M. Mohamed, N. Mohamed, and J. G. Kim, "P300 latency with memory performance: A promising biomarker for preclinical stages of alzheimer's disease," *Biosensors*, vol. 14, no. 12, p. 616, 2024.

[18] V. Fateeva, A. Kushnir, A. Grechko, and L. Mayorova, "Rehabilitation of patients with post-stroke cognitive impairments using a p300-based brain–computer interface: Results of a randomized controlled trial," *Neuroscience and Behavioral Physiology*, vol. 54, no. 4, pp. 575–580, 2024.

[19] W. H. Elashmawi, A. Ayman, M. Antoun, *et al.*, "A comprehensive review on brain–computer interface (bci)-based machine and deep learning algorithms for stroke rehabilitation," *Applied Sciences*, vol. 14, no. 14, p. 6347, 2024.

[20] R. A. Ramadan and A. B. Altamimi, "Unraveling the potential of brain-computer interface technology in medical diagnostics and rehabilitation: A comprehensive literature review," *Health and Technology*, vol. 14, no. 2, pp. 263–276, 2024.

[21] V. Khorev, S. Kurkin, A. Badarin, *et al.*, "Review on the use of brain computer interface rehabilitation methods for treating mental and neurological conditions," *Journal of Integrative Neuroscience*, vol. 23, no. 7, p. 125, 2024.

[22] L. Judijanto, A. Y. Vandika, and R. Toalib, "Brain-computer interfaces in assistive technologies: A bibliometric review," *The Eastasouth Journal of Information System and Computer Science*, vol. 2, no. 03, pp. 272–287, 2025.

[23] V. Galiotta, V. Caracci, J. Toppi, *et al.*, "P300-based brain-computer interface for communication in assistive technology centers: Influence of users' profile on bci access," *JOURNAL OF NEURAL ENGINEERING*, 2024.

[24] S. Zare and Y. Sun, "Eeg motor imagery classification using integrated transformer-cnn for assistive technology control," in *2024 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, IEEE, 2024, pp. 189–190.

[25] B. Sivakumar, R. Hariharan, and V. Mohanavel, "Advanced wheelchair and robotic assistance technologies utilizing autonomous robotics for individuals with disabilities," in *2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, IEEE, 2024, pp. 1–7.

[26] M. Albán-Escobar, P. Navarrete-Arroyo, D. R. De la Cruz-Guevara, and J. Tobar-Quevedo, "Assistance device based on ssvep-bci online to control a 6-dof robotic arm," *Sensors*, vol. 24, no. 6, p. 1922, 2024.

[27] D. Millett, "Hans berger: From psychic energy to the eeg," *Perspectives in biology and medicine*, vol. 44, no. 4, pp. 522–542, 2001.

[28] M. Tudor, L. Tudor, and K. I. Tudor, "Hans berger (1873-1941)–the history of electroencephalography," *Acta medica Croatica: casopis Hravatske akademije medicinskih znanosti*, vol. 59, no. 4, pp. 307–313, 2005.

[29] J. J. Vidal, "Toward direct brain-computer communication," *Annual review of Biophysics and Bioengineering*, vol. 2, no. 1, pp. 157–180, 1973.

[30] J. J. Vidal, "Real-time detection of brain events in eeg," *Proceedings of the IEEE*, vol. 65, no. 5, pp. 633–641, 1977.

[31] J. P. Donoghue, "Connecting cortex to machines: Recent advances in brain interfaces," *Nature neuroscience*, vol. 5, no. Suppl 11, pp. 1085–1088, 2002.

[32] M. A. Nicolelis, "Brain–machine interfaces to restore motor function and probe neural circuits," *Nature Reviews Neuroscience*, vol. 4, no. 5, pp. 417–422, 2003.

[33] A. Kawala-Sterniuk, N. Browarska, A. Al-Bakri, *et al.*, "Summary of over fifty years with brain-computer interfaces—a review," *Brain sciences*, vol. 11, no. 1, p. 43, 2021.

[34] A. Kübler, "The history of bci: From a vision for the future to real support for personhood in people with locked-in syndrome," *Neuroethics*, vol. 13, no. 2, pp. 163–180, 2020.

[35] M. Zhuang, Q. Wu, F. Wan, and Y. Hu, "State-of-the-art non-invasive brain–computer interface for neural rehabilitation: A review," *Journal of Neurorestoratology*, vol. 8, no. 1, pp. 12–25, 2020.

[36] J. d. R. Millán and J. Carmena, "Invasive or noninvasive: Understanding brain-machine interface technology," Institute of Electrical and Electronics Engineers, 2010.

[37] Y. Zhang, "Invasive bci and noninvasive bci with vr/ar technology," in *International conference on artificial intelligence, virtual reality, and visualization (AIVRV 2021)*, SPIE, vol. 12153, 2021, pp. 186–192.

[38] L. R. Hochberg, D. Bacher, B. Jarosiewicz, *et al.*, "Reach and grasp by people with tetraplegia using a neurally controlled robotic arm," *Nature*, vol. 485, no. 7398, pp. 372–375, 2012.

[39] Y. Zhou, T. Yu, W. Gao, *et al.*, "Shared three-dimensional robotic arm control based on asynchronous bci and computer vision," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 3163–3175, 2023.

[40] A. Keutayeva, C. J. Nwachukwu, M. Alaran, Z. Otarbay, and B. Abibullaev, "Neurotechnology in gaming: A systematic review of visual evoked potential-based brain-computer interfaces," *IEEE Access*, 2025.

[41] A. Alexopoulou, P. Pergantis, C. Koutsojannis, V. Triantafillou, and A. Drigas, "Non-invasive bci-vr applied protocols as intervention paradigms on school-aged subjects with asd: A systematic review," *Sensors*, vol. 25, no. 5, p. 1342, 2025.

[42] P. Mi, L. Yan, Y. Cheng, *et al.*, "Driver cognitive architecture based on eeg signals: A review," *IEEE Sensors Journal*, 2024.

[43] M.-V. Drăgoi, I. Nisipeanu, A. Frimu, *et al.*, "Real-time home automation system using bci technology," *Biomimetics*, vol. 9, no. 10, p. 594, 2024.

[44] J. L. Collinger, B. Wodlinger, J. E. Downey, *et al.*, "High-performance neuroprosthetic control by an individual with tetraplegia," *The Lancet*, vol. 381, no. 9866, pp. 557–564, 2013.

[45] M. Islam, A. Vashishat, and M. Kumar, "Advancements beyond limb loss: Exploring the intersection of ai and bci in prosthetic evaluation," *Current Pharmaceutical Design*, vol. 30, no. 35, pp. 2749–2752, 2024.

[46] N. Robinson, R. Mane, T. Chouhan, and C. Guan, "Emerging trends in bci-robotics for motor control and rehabilitation," *Current Opinion in Biomedical Engineering*, vol. 20, p. 100 354, 2021.

[47] X. Zhang, J. Li, R. Zhang, and T. Liu, "A brain-controlled and user-centered intelligent wheelchair: A feasibility study," *Sensors*, vol. 24, no. 10, p. 3000, 2024.

[48] M. A. Awais, M. Z. Yusoff, N. Yahya, S. Z. Ahmed, and M. U. Qamar, "Brain controlled wheelchair: A smart prototype," in *Journal of Physics: Conference Series*, IOP Publishing, vol. 1529, 2020, p. 042 075.

[49] F. Wang, Y. Wen, J. Bi, H. Li, and J. Sun, "A portable ssvep-bci system for rehabilitation exoskeleton in augmented reality environment," *Biomedical Signal Processing and Control*, vol. 83, p. 104 664, 2023.

[50] A. Colucci, M. Vermehren, A. Cavallo, *et al.*, *Neurorehabilitation and neural repair*, vol. 36, no. 12, pp. 747–756, 2022.

[51] T. W. Picton *et al.*, "The p300 wave of the human event-related potential," *Journal of clinical neurophysiology*, vol. 9, pp. 456–456, 1992.

[52] L. L. Oganesian and M. M. Shanechi, "Brain–computer interfaces for neuropsychiatric disorders," *Nature Reviews Bioengineering*, vol. 2, no. 8, pp. 653–670, 2024.

[53] M. Z. Raza, M. Omais, H. M. E. Arshad, M. Maqsood, and A. A. Nadeem, "Effectiveness of brain-computer interface (bci)-based attention training game system for symptom reduction, behavioral enhancement, and brain function

modulation in children with adhd: A systematic review and single-arm meta-analysis," *NeuroRegulation*, vol. 12, no. 1, pp. 51–51, 2025.

[54] H. Givian, "The application of brain–computer interface in alzheimer's disease studies based on machine learning algorithms," in *Artificial Intelligence Applications for Brain–Computer Interfaces*, Elsevier, 2025, pp. 121–143.

[55] N. Huynh and G. Deshpande, "A review of the applications of generative adversarial networks to structural and functional mri based diagnostic classification of brain disorders," *Frontiers in Neuroscience*, vol. 18, p. 1 333 712, 2024.

[56] P. Hui, Y. Jiang, J. Wang, *et al.*, "Exploring the application and challenges of fnirs technology in early detection of parkinson's disease," *Frontiers in Aging Neuroscience*, vol. 16, p. 1 354 147, 2024.

[57] M. Bamdad, H. Zarshenas, and M. A. Auais, "Application of bci systems in neurorehabilitation: A scoping review," *Disability and Rehabilitation: Assistive Technology*, vol. 10, no. 5, pp. 355–364, 2015.

[58] A.-K. R. J. Salman, A.-D. M. A. Luti, *et al.*, "Brain-computer interface technologies and neurofeedback," *Western European Journal of Modern Experiments and Scientific Methods*, vol. 2, no. 7, pp. 67–74, 2024.

[59] R. K. Kanna, S. V. Athawale, M. Y. Naniwadekar, C. Choudhari, N. R. Talhar, and S. Dhengre, "Anxiety controlling application using eeg neurofeedback system.," *EAI Endorsed Transactions on Pervasive Health & Technology*, vol. 10, no. 1, 2024.

[60] F. Krause, D. E. Linden, and E. J. Hermans, "Getting stress-related disorders under control: The untapped potential of neurofeedback," *Trends in neurosciences*, 2024.

[61] T. Suhail, R. Subasree, V. AP, *et al.*, "On the feasibility of an online brain-computer interface-based neurofeedback game for enhancing attention and

working memory in stroke and mild cognitive impairment patients," *Biomedical Physics & Engineering Express*, vol. 11, no. 2, p. 025 049, 2025.

[62] F. Atilla, M. Postma, and M. Alimardani, "Gamification of motor imagery brain-computer interface training protocols: A systematic review," *Computers in Human Behavior Reports*, p. 100 508, 2024.

[63] P. Arpaia, A. Esposito, E. Galasso, *et al.*, "Endless runner game in virtual reality controlled by a self-paced brain-computer interface based on eeg and motor imagery," in *International Conference on Extended Reality*, Springer, 2024, pp. 208–225.

[64] S. D. ROSCA, M. MARCU, M. LEBA, D. I. DINEA, and D. C. I. DINEA, "The impact of meditation in a video game based on bci," *Annals of the University of Petrosani Electrical Engineering*, vol. 26, 2024.

[65] G. Prapas, P. Angelidis, P. Sarigiannidis, S. Bibi, and M. G. Tsipouras, "Connecting the brain with augmented reality: A systematic review of bci-ar systems," *Applied Sciences*, vol. 14, no. 21, p. 9855, 2024.

[66] M. Al-Mohammadi, A. Al-Nuaimi, and L. B. Farah, "Smart home based on bci for disabled people: A state-of-the-art review," in *AIP Conference Proceedings*, AIP Publishing, vol. 3232, 2024.

[67] M. F. K. Khondakar, M. H. Sarowar, M. H. Chowdhury, *et al.*, "A systematic review on eeg-based neuromarketing: Recent trends and analyzing techniques," *Brain Informatics*, vol. 11, no. 1, p. 17, 2024.

[68] S. F. B. Ahmed, M. Hasan, M. T. I. Opu, *et al.*, "Recent trends in eeg-based p300, neuromarketing, and e-sports brain-computer interface applications: A review," *Advanced Electroencephalography Analytical Methods*, pp. 87–110, 2024.

[69] M. Aldayel, N. Alsedairy, A. Al-Nafjan, and S. Alsenan, "Systematic review of brain-computer interface based user authentication system: Trends, challenges, and directions," *IEEE Access*, 2024.

[70] M. Chanda, O. Mazumder, S. Paul, and T. Rahman, "Brain–computer interfaces for security and authentication," *Artificial Intelligence Applications for Brain–Computer Interfaces*, pp. 205–225, 2025.

[71] J. Ju and H. Li, "A survey of eeg-based driver state and behavior detection for intelligent vehicles," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2024.

[72] E. van Weelden, C. W. van Beek, M. Alimardani, *et al.*, "A passive brain-computer interface for predicting pilot workload in virtual reality flight training," in *2024 IEEE 4th International Conference on Human-Machine Systems (ICHMS)*, IEEE, 2024, pp. 1–6.

[73] Y. Li and J. He, "A review of strategies to detect fatigue and sleep problems in aviation: Insights from artificial intelligence," *Archives of Computational Methods in Engineering*, vol. 31, no. 8, pp. 4655–4672, 2024.

[74] L. Angrisani, P. Arpaia, A. Esposito, *et al.*, "Passive and active brain-computer interfaces for rehabilitation in health 4.0," *Measurement: Sensors*, vol. 18, p. 100 246, 2021.

[75] J. H. Teo, N. S. Ahmad, and P. Goh, "Visual stimuli-based dynamic commands with intelligent control for reactive bci applications," *IEEE Sensors Journal*, vol. 22, no. 2, pp. 1435–1448, 2021.

[76] H. Wang, F. Yan, T. Xu, *et al.*, "Brain-controlled wheelchair review: From wet electrode to dry electrode, from single modal to hybrid modal, from synchronous to asynchronous," *IEEE Access*, vol. 9, pp. 55 920–55 938, 2021.

[77] H. Yuan, Y. Li, J. Yang, *et al.*, "State of the art of non-invasive electrode materials for brain–computer interface," *Micromachines*, vol. 12, no. 12, p. 1521, 2021.

[78] J. Wolpaw and D. McFarland, "Development of an eeg-based brain-computer interface (bci)," *Rehabil. Eng. Soc. N. Am*, vol. 15, pp. 645–648, 1995.

[79] D. McFarland, W. Sarnacki, and J. Wolpaw, "Eeg-based brain-computer interface (bci): Multiple selections with one dimensional control," in *Soc. Neurosci. Abstr*, vol. 23, 1998, p. 656.

[80] D. Flotzinger, J. Kalcher, and J. Wolpaw, *Off-line Classification of EEG from the" New York Brain-computer Interface (BCI)"*. Technische Universität Graz/Österreichische Computer Gesellschaft ..., 1993.

[81] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.

[82] G. Pandarinathan, S. Mishra, A. M. Nedumaran, P. Padmanabhan, and B. Gulyás, "The potential of cognitive neuroimaging: A way forward to the mind-machine interface," *Journal of Imaging*, vol. 4, no. 5, p. 70, 2018.

[83] S. Saha, K. A. Mamun, K. Ahmed, *et al.*, "Progress in brain computer interface: Challenges and opportunities," *Frontiers in systems neuroscience*, vol. 15, p. 578 875, 2021.

[84] A. L. Fred, S. N. Kumar, A. Kumar Haridhas, *et al.*, "A brief introduction to magnetoencephalography (meg) and its clinical applications," *Brain sciences*, vol. 12, no. 6, p. 788, 2022.

[85] W.-L. Chen, J. Wagner, N. Heugel, *et al.*, "Functional near-infrared spectroscopy and its clinical application in the field of neuroscience: Advances and future directions," *Frontiers in neuroscience*, vol. 14, p. 724, 2020.

[86] B. R. Buchbinder, "Functional magnetic resonance imaging," *Handbook of clinical neurology*, vol. 135, pp. 61–92, 2016.

[87] K. J. Miller, D. Hermes, and N. P. Staff, "The current state of electrocorticography-based brain–computer interfaces," *Neurosurgical focus*, vol. 49, no. 1, E2, 2020.

[88] Y. Dong, S. Wang, Q. Huang, R. W. Berg, G. Li, and J. He, "Neural decoding for intracortical brain–computer interfaces," *Cyborg and Bionic Systems*, vol. 4, p. 0044, 2023.

[89] L. F. Haas, "Hans berger (1873–1941), richard caton (1842–1926), and electroencephalography," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 74, no. 1, pp. 9–9, 2003.

[90] A. Coenen and O. Zayachkivska, "Adolf beck: A pioneer in electroencephalography in between richard caton and hans berger," *Advances in cognitive psychology*, vol. 9, no. 4, p. 216, 2013.

[91] W. Kułak and W. Sobaniec, "Historia odkrycia eeg," *Neurologia dziecieca*, vol. 15, no. 29, pp. 53–56, 2006.

[92] S. Sadiya, T. Alhanai, and M. M. Ghassemi, "Artifact detection and correction in eeg data: A review," in *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*, IEEE, 2021, pp. 495–498.

[93] J. J. Newson and T. C. Thiagarajan, "Eeg frequency bands in psychiatric disorders: A review of resting state studies," *Frontiers in human neuroscience*, vol. 12, p. 521, 2019.

[94] P. Tai, P. Ding, F. Wang, *et al.*, "Brain-computer interface paradigms and neural coding," *Frontiers in neuroscience*, vol. 17, p. 1 345 961, 2024.

[95] D. J. Creel, "Visually evoked potentials," *Handbook of clinical neurology*, vol. 160, pp. 501–522, 2019.

[96] R. F. Burkard, J. J. Eggermont, and M. Don, *Auditory evoked potentials: basic principles and clinical application*. Lippincott Williams & Wilkins, 2007.

[97] D. Regan, "Steady-state evoked potentials," *Journal of the optical society of America*, vol. 67, no. 11, pp. 1475–1489, 1977.

[98] A. Schielke and B. Krekelberg, "Steady state visual evoked potentials in schizophrenia: A review," *Frontiers in Neuroscience*, vol. 16, p. 988 077, 2022.

[99] H. Zhang, J. Xie, Q. Tao, *et al.*, "The effect of motion frequency and sound source frequency on steady-state auditory motion evoked potential," *Hearing Research*, vol. 439, p. 108 897, 2023.

[100] M. Borirakarawin, N. Siribunyaphat, and Y. Punsawad, "Steady-state somatosensory evoked potential-based brain-computer interface using a vibrotactile stimulus by mixing vibration frequencies," in *2024 16th Biomedical Engineering International Conference (BMEiCON)*, IEEE, 2024, pp. 1–4.

[101] M. Miyajima, A. Toyomaki, I. Kusumi, and T. Koyama, "Comparison of neural responses between exogenous and endogenous rule shifting in cued switching task; an erps study," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 34, 2012.

[102] M. Lotze and U. Halsband, "Motor imagery," *Journal of Physiology-paris*, vol. 99, no. 4-6, pp. 386–395, 2006.

[103] S. Rimbert, D. Trocellier, and F. Lotte, "Impact of the baseline temporal selection on the erd/ers analysis for motor imagery-based bci," in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2023, pp. 1–4.

[104] T. Ma, H. Li, L. Deng, *et al.*, "The hybrid bci system for movement control by combining motor imagery and moving onset visual evoked potential," *Journal of neural engineering*, vol. 14, no. 2, p. 026 015, 2017.

[105] X. Chi, C. Wan, C. Wang, Y. Zhang, X. Chen, and H. Cui, "A novel hybrid brain-computer interface combining motor imagery and intermodulation steady-state visual evoked potential," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 1525–1535, 2022.

[106] L.-M. Zhao, X. Yan, and B.-L. Lu, "Plug-and-play domain adaptation for cross-subject eeg-based emotion recognition," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, 2021, pp. 863–870.

[107] A. Jayanthi and E. C. Djamal, "Hybrid cnn-rnn in motor imagery identification of brain-computer interface," in *2021 8th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*, IEEE, 2021, pp. 1–6.

[108] M. A. Pfeffer, S. S. H. Ling, and J. K. W. Wong, "Exploring the frontier: Transformer-based models in eeg signal analysis for brain-computer interfaces," *Computers in Biology and Medicine*, p. 108 705, 2024.

[109] M. A. Awais, T. Ward, P. Redmond, and G. Healy, "From lab to life: Assessing the impact of real-world interactions on the operation of rapid serial visual presentation-based brain-computer interfaces," *Journal of Neural Engineering*, vol. 21, no. 4, p. 046 011, 2024.

[110] W. Gao, W. Huang, M. Li, *et al.*, "Eliminating or shortening the calibration for a p300 brain–computer interface based on a convolutional neural network and big electroencephalography data: An online study," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 1754–1763, 2023.

[111] E. C. Gordon and A. K. Seth, "Ethical considerations for the use of brain–computer interfaces for cognitive enhancement," *PLoS biology*, vol. 22, no. 10, e3002899, 2024.

[112] W. S. Pritchard, "Psychophysiology of p300.," *Psychological bulletin*, vol. 89, no. 3, p. 506, 1981.

[113] S. J. Luck, *An introduction to the event-related potential technique*. MIT press, 2014.

[114] K. Värbu, N. Muhammad, and Y. Muhammad, "Past, present, and future of eeg-based bci applications," *Sensors*, vol. 22, no. 9, p. 3331, 2022.

[115] J. Polich, "Clinical application of the p300 event-related brain potential," *Physical Medicine and Rehabilitation Clinics*, vol. 15, no. 1, pp. 133–161, 2004.

[116] S. K. Mudgal, S. K. Sharma, J. Chaturvedi, and A. Sharma, "Brain computer interface advancement in neurosciences: Applications and issues," *Interdisciplinary Neurosurgery*, vol. 20, p. 100 694, 2020.

[117] R. Bettencourt, M. Castelo-Branco, E. Gonçalves, U. J. Nunes, and G. Pires, "Comparing several p300-based visuo-auditory brain-computer interfaces for a completely locked-in als patient: A longitudinal case study," *Applied Sciences*, vol. 14, no. 8, p. 3464, 2024.

[118] A. Bablani and D. Tripathi, "A review on methods applied on p300-based lie detectors," *Advances in Machine Learning and Data Science: Recent Achievements and Research Directives*, pp. 251–257, 2018.

[119] M. Zabcikova, Z. Koudelkova, and R. Jasek, "Eeg-based lie detection using erp p300 in response to known and unknown faces: An overview," in *2022 26th International Conference on Circuits, Systems, Communications and Computers (CSCC)*, IEEE, 2022, pp. 11–15.

[120] J. Gao, X. Min, Q. Kang, *et al.*, "Effective connectivity in cortical networks during deception: A lie detection study based on eeg," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 8, pp. 3755–3766, 2022.

[121] H. Wang, W. Chang, and C. Zhang, "Functional brain network and multichannel analysis for the p300-based brain computer interface system of lying detection," *Expert Systems with Applications*, vol. 53, pp. 117–128, 2016.

[122] M. Maleki, N. Manshouri, and T. Kayikcioglu, "Brain-computer interface systems for smart homes-a review study," *Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering)*, vol. 14, no. 2, pp. 144–155, 2021.

[123] E. Muglerab, M. Benschc, S. Haldera, *et al.*, "Control of an internet browser using the p300 event-related potential," *International Journal of Bioelectromagnetism*, vol. 10, no. 1, pp. 56–63, 2008.

[124] Z. Tang, X. Wang, J. Wu, Y. Ping, X. Guo, and Z. Cui, "A bci painting system using a hybrid control approach based on ssvep and p300," *Computers in Biology and Medicine*, vol. 150, p. 106 118, 2022.

[125] J. I. Münßinger, S. Halder, S. C. Kleih, *et al.*, "Brain painting: First evaluation of a new brain–computer interface application with als-patients and healthy volunteers," *Frontiers in neuroscience*, vol. 4, p. 182, 2010.

[126] D. Boland, M. Quek, M. Tangermann, J. Williamson, and R. Murray-Smith, "Using simulated input into brain-computer interfaces for user-centred design," *International Journal of Bioelectromagnetism*, vol. 13, no. 2, pp. 86–87, 2011.

[127] A. Finke, A. Lenhardt, and H. Ritter, "The mindgame: A p300-based brain–computer interface game," *Neural Networks*, vol. 22, no. 9, pp. 1329–1333, 2009.

[128] C. Mühl, H. Gürkök, D. P.-O. Bos, *et al.*, "Bacteria hunt: A multimodal, multiparadigm bci game," in *5th International Summer Workshop on Multimodal Interfaces, eNTERFACE 2009*, University of Genua, 2010, pp. 41–62.

[129] D. De Venuto, V. F. Annese, G. Mezzina, M. Ruta, and E. Di Sciascio, "Brain-computer interface using p300: A gaming approach for neurocognitive impairment diagnosis," in *2016 IEEE International High Level Design Validation and Test Workshop (HLDVT)*, IEEE, 2016, pp. 93–99.

[130] K. Yamauchi, B. Chaleunxay, K. Lin, *et al.*, "Utility of the p300 eeg biomarker in the early identification of alzheimer's disease: A systematic review (p1-3.005)," in *Neurology*, Lippincott Williams & Wilkins Hagerstown, MD, vol. 104, 2025, p. 4485.

[131] Y. Zhang, T. Yang, Y. He, *et al.*, "Value of p300 amplitude in the diagnosis of untreated first-episode schizophrenia and psychosis risk syndrome in children and adolescents," *BMC psychiatry*, vol. 23, no. 1, p. 743, 2023.

[132] S. Arheix-Parras, B. Glize, D. Guehl, and G. Python, "Electrophysiological changes in patients with post-stroke aphasia: A systematic review," *Brain topography*, vol. 36, no. 2, pp. 135–171, 2023.

[133] S. C. Kleih, L. Gottschalt, E. Teichlein, and F. X. Weilbach, "Toward a p300 based brain-computer interface for aphasia rehabilitation after stroke: Presentation of theoretical considerations and a pilot feasibility study," *Frontiers in human neuroscience*, vol. 10, p. 547, 2016.

[134] S. Noble, E. Woods, T. Ward, and J. Ringwood, "Accelerating p300-based neurofeedback training for attention enhancement using iterative learning control: A randomised controlled trial," *Journal of Neural Engineering*, vol. 21, no. 2, p. 026 006, 2024.

[135] R. Manor, L. Mishali, and A. B. Geva, "Multimodal neural network for rapid serial visual presentation brain computer interface," *Frontiers in computational neuroscience*, vol. 10, p. 130, 2016.

[136] P. Sajda, E. Pohlmeyer, J. Wang, *et al.*, "In a blink of an eye and a switch of a transistor: Cortically coupled computer vision," *Proceedings of the IEEE*, vol. 98, no. 3, pp. 462–478, 2010.

[137] H. Cecotti, M. P. Eckstein, and B. Giesbrecht, "Effects of performing two visual tasks on single-trial detection of event-related potentials," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, 2012, pp. 1723–1726.

[138] H. Cecotti, A. J. Ries, M. P. Eckstein, and B. Giesbrecht, "Multiclass classification of single-trial evoked eeg responses," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, 2012, pp. 1719–1722.

[139] J. Wang, E. Pohlmeyer, B. Hanna, Y.-G. Jiang, P. Sajda, and S.-F. Chang, "Brain state decoding for rapid image retrieval," in *Proceedings of the 17th ACM international conference on Multimedia*, 2009, pp. 945–954.

[140] M. Weiden, D. Khosla, and M. Keegan, "Electroencephalographic detection of visual saliency of motion towards a practical brain-computer interface for

video analysis," in *Proceedings of the 14th ACM international conference on Multimodal interaction*, 2012, pp. 601–606.

[141] D. Rosenthal, P. DeGuzman, L. C. Parra, and P. Sajda, "Evoked neural responses to events in video," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 3, pp. 358–365, 2014.

[142] S. Lees, N. Dayan, H. Cecotti, *et al.*, "A review of rapid serial visual presentation-based brain–computer interfaces," *Journal of neural engineering*, vol. 15, no. 2, p. 021 001, 2018.

[143] A. D. Gerson, L. C. Parra, and P. Sajda, "Cortical origins of response time variability during rapid discrimination of visual objects," *Neuroimage*, vol. 28, no. 2, pp. 342–353, 2005.

[144] S. Mathan, D. Erdogmus, Y. Huang, *et al.*, "Rapid image analysis using neural signals," in *CHI'08 Extended Abstracts on Human Factors in Computing Systems*, 2008, pp. 3309–3314.

[145] Y. Huang, D. Erdogmus, S. Mathan, and M. Pavel, "A fusion approach for image triage using single trial erp detection," in *2007 3rd International IEEE/EMBS Conference on Neural Engineering*, IEEE, 2007, pp. 473–476.

[146] Y. Zhou, W. Wang, L. Yan, and B. Yang, "Research on the relationship between fatigue and p300 potential in multi-stage rsvp small target detection," in *Proceedings of the 2021 10th International Conference on Computing and Pattern Recognition*, 2021, pp. 92–98.

[147] C. F. Blanco-Díaz, C. D. Guerrero-Méndez, T. Bastos-Filho, S. Jaramillo-Isaza, and A. F. Ruiz-Olaya, "Effects of the concentration level, eye fatigue and coffee consumption on the performance of a bci system based on visual erp-p300," *Journal of Neuroscience Methods*, vol. 382, p. 109 722, 2022.

[148] R. Ron-Angevin, M. T. Medina-Juliá1, Á. Fernández-Rodríguez, *et al.*, "Performance analysis with different types of visual stimuli in a bci-based speller

under an rsvp paradigm," *Frontiers in computational neuroscience*, vol. 14, p. 587 702, 2021.

[149] A. Riccio, L. Simione, F. Schettini, *et al.*, "Attention and working memory influence on p300-based bci performance in people with amyotrophic lateral sclerosis," in *Proceedings of the 6th International Brain-Computer Interface Conference*, Graz University of Technology Publishing House, 2014.

[150] C. N. Gupta, R. Palaniappan, and R. Paramesran, "Exploiting the p300 paradigm for cognitive biometrics," *International Journal of Cognitive Biometrics*, vol. 1, no. 1, pp. 26–38, 2012.

[151] M. T. Mittelstadt, "Deep learning for rsvp-ssvep game," M.S. thesis, The University of Texas at San Antonio, 2019.

[152] J. T. Rexwinkle, G. Lieberman, M. Jaswa, and B. J. Lance, "Development of a game with a purpose for acquisition of brain-computer interface data," *arXiv preprint arXiv:1910.00106*, 2019.

[153] W. Malik and N. Muhktar, "Rapid serial visual presentation (rsvp) technique analysis for young adult's strength memory," *Malaysian Journal of Industrial Technology*, vol. 8, no. 2, pp. 23–36, 2024.

[154] D. Galvin-McLaughlin, D. Klee, T. Memmott, *et al.*, "Methodology and preliminary data on feasibility of a neurofeedback protocol to improve visual attention to letters in mild alzheimer's disease," *Contemporary Clinical Trials Communications*, vol. 28, p. 100 950, 2022.

[155] Y. Zhou, B. Yang, L. Yan, and W. Wang, "Evaluation of fatigue based on reaction time and fnirs of rsvp small target detection," in *Intelligent Life System Modelling, Image Processing and Analysis: 7th International Conference on Life System Modeling and Simulation, LSMS 2021 and 7th International Conference on Intelligent Computing for Sustainable Energy and Environment, ICSEE 2021, Hangzhou, China, October 22–24, 2021, Proceedings, Part I 7*, Springer, 2021, pp. 84–93.

[156] C. Hope, A. Sterr, P. Elangovan, *et al.*, "High throughput screening for mammography using a human-computer interface with rapid serial visual presentation (rsvp)," in *Medical Imaging 2013: Image Perception, Observer Performance, and Technology Assessment*, SPIE, vol. 8673, 2013, pp. 9–16.

[157] M. Fried-Oken, A. Mooney, B. Peters, and B. Oken, "A clinical screening protocol for the rsvp keyboard brain–computer interface," *Disability and Rehabilitation: Assistive Technology*, vol. 10, no. 1, pp. 11–18, 2015.

[158] Y. Sheng, S. Liu, W. Wang, *et al.*, "A study on rsvp paradigm based on brain computer interface across subjects," in *2018 9th International Conference on Awareness Science and Technology (iCAST)*, IEEE, 2018, pp. 42–46.

[159] Y. Cui, S. Xie, X. Xie, *et al.*, "Lder: A classification framework based on erp enhancement in rsvp task," *Journal of Neural Engineering*, vol. 20, no. 3, p. 036 029, 2023.

[160] S. Zhang, Y. Wang, L. Zhang, and X. Gao, "A benchmark dataset for rsvp-based brain–computer interfaces," *Frontiers in neuroscience*, vol. 14, p. 568 000, 2020.

[161] A. M. Mijani, M. B. Shamsollahi, and M. S. Hassani, "A novel dual and triple shifted rsvp paradigm for p300 speller," *Journal of neuroscience methods*, vol. 328, p. 108 420, 2019.

[162] X. Xiao, M. Xu, and D. Ming, "A comparison of classification methods for recognizing single-trial erp in rsvp-based brain-computer interfaces," in *2019 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, IEEE, 2019, pp. 1–4.

[163] S. U. Rahman, N. O'Connor, J. Lemley, and G. Healy, "Using pre-stimulus eeg to predict driver reaction time to road events," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2022, pp. 4036–4039.

[164] H. Cecotti, J. Sato-Reinhold, J. L. Sy, J. C. Elliott, M. P. Eckstein, and B. Giesbrecht, "Impact of target probability on single-trial eeg target detection in a difficult rapid serial visual presentation task," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, 2011, pp. 6381–6384.

[165] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: A compact convolutional neural network for eeg-based brain–computer interfaces," *Journal of neural engineering*, vol. 15, no. 5, p. 056 013, 2018.

[166] H. Cecotti and A. Graser, "Convolutional neural networks for p300 detection with application to brain-computer interfaces," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 3, pp. 433–445, 2010.

[167] M. Uma, S. Prabhu, M. Subramaniyam, and S. N. Min, "Analysis of effect of rsvp speller bci paradigm along with cnn to analysis p300 signals," in *International Conference on Human-Computer Interaction*, Springer, 2021, pp. 84–96.

[168] B. Zang, Y. Lin, Z. Liu, and X. Gao, "A deep learning method for single-trial eeg classification in rsvp task based on spatiotemporal features of erps," *Journal of Neural Engineering*, vol. 18, no. 4, p. 0460c8, 2021.

[169] Z. Yuan, Q. Zhou, B. Wang, *et al.*, "Psaeegnet: Pyramid squeeze attention mechanism-based cnn for single-trial eeg classification in rsvp task," *Frontiers in Human Neuroscience*, vol. 18, p. 1 385 360, 2024.

[170] H. Wang, Z. Wang, Y. Sun, Z. Yuan, T. Xu, and J. Li, "A cascade xdawn eegnet structure for unified visual-evoked related potential detection," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2024.

[171] A. M. Aljlaly, I. F. Elshami, and A. Almijbari, "Long short-term memory employed for classifying p300-based bci," in *2024 IEEE 4th International*

Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA), IEEE, 2024, pp. 666–671.

[172] M. Xu, Y. Chen, Y. Wang, D. Wang, Z. Liu, and L. Zhang, "Bwgan-gp: An eeg data generation method for class imbalance problem in rsvp tasks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 251–263, 2022.

[173] R. Ranjan, B. C. Sahana, and A. K. Bhandari, "Ocular artifact elimination from electroencephalography signals: A systematic review," *Biocybernetics and Biomedical Engineering*, vol. 41, no. 3, pp. 960–996, 2021.

[174] G. Huang, Z. Hu, L. Zhang, L. Li, Z. Liang, and Z. Zhang, "Removal of eye-blinking artifacts by ica in cross-modal long-term eeg recording," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2020, pp. 217–220.

[175] K. Won, M. Kwon, M. Ahn, and S. C. Jun, "Eeg dataset for rsvp and p300 speller brain-computer interfaces," *Scientific Data*, vol. 9, no. 1, p. 388, 2022.

[176] L. Acqualagna and B. Blankertz, "Gaze-independent bci-spelling using rapid serial visual presentation (rsvp)," *Clinical Neurophysiology*, vol. 124, no. 5, pp. 901–908, 2013.

[177] S. Xue, B. Jin, J. Jiang, *et al.*, "A multi-subject and multi-session eeg dataset for modelling human visual object recognition," *Scientific Data*, vol. 12, no. 1, p. 663, 2025.

[178] B. Abibullaev, K. Kunanbayev, and A. Zollanvari, "Subject-independent classification of p300 event-related potentials using a small number of training subjects," *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 5, pp. 843–854, 2022.

[179] J. Qu, F. Wang, Z. Xia, *et al.*, "A novel three-dimensional p300 speller based on stereo visual stimuli," *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 4, pp. 392–399, 2018.

[180] D. Ryan, G. Townsend, N. Gates, K. Colwell, and E. Sellers, "Evaluating brain-computer interface performance using color in the p300 checkerboard speller," *Clinical Neurophysiology*, vol. 128, no. 10, pp. 2050–2057, 2017.

[181] Y. Li, H. Liu, and S. Wang, "Exploiting eeg channel correlations in p300 speller paradigm for brain-computer interface," *IEICE TRANSACTIONS on Information and Systems*, vol. 99, no. 6, pp. 1653–1662, 2016.

[182] A. Farahat, C. Reichert, C. M. Sweeney-Reed, and H. Hinrichs, "Convolutional neural networks for decoding of covert attention focus and saliency maps for eeg feature visualization," *Journal of neural engineering*, vol. 16, no. 6, p. 066 010, 2019.

[183] S. Kundu and S. Ari, "Fusion of convolutional neural networks for p300 based character recognition," in *2019 International Conference on Information Technology (ICIT)*, IEEE, 2019, pp. 155–159.

[184] R. Maddula, J. Stivers, M. Mousavi, S. Ravindran, and V. de Sa, "Deep recurrent convolutional neural networks for classifying p300 bci signals.," *GBCIC*, vol. 201, pp. 18–22, 2017.

[185] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from eeg with deep recurrent-convolutional neural networks. arxiv 2015," *arXiv preprint arXiv:1511.06448*, 2015.

[186] R. Joshi, P. Goel, M. Sur, and H. A. Murthy, "Single trial p300 classification using convolutional lstm and deep learning ensembles method," in *Intelligent Human Computer Interaction: 10th International Conference, IHCI 2018, Allahabad, India, December 7–9, 2018, Proceedings 10*, Springer, 2018, pp. 3–15.

[187] S. An, S. Kim, P. Chikontwe, and S. H. Park, "Dual attention relation network with fine-tuning for few-shot eeg motor imagery classification," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[188] W. Cui, W. Jeong, P. Thölke, *et al.*, "Neuro-gpt: Towards a foundation model for eeg," in *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2024, pp. 1–5.

[189] M. Śliwowski, M. Martin, A. Souloumiac, P. Blanchart, and T. Aksenova, "Impact of dataset size and long-term ecog-based bci usage on deep learning decoders performance," *Frontiers in Human Neuroscience*, vol. 17, p. 1 111 645, 2023.

[190] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: A systematic review," *Journal of neural engineering*, vol. 16, no. 5, p. 051 001, 2019.

[191] C. Rommel, J. Paillard, T. Moreau, and A. Gramfort, "Data augmentation for learning predictive models on eeg: A systematic comparison," *Journal of Neural Engineering*, vol. 19, no. 6, p. 066 020, 2022.

[192] B. Rivet, H. Cecotti, E. Maby, and J. Mattout, "Impact of spatial filters during sensor selection in a visual p300 brain-computer interface," *Brain topography*, vol. 25, pp. 55–63, 2012.

[193] H. Cecotti, B. Rivet, M. Congedo, *et al.*, "A robust sensor-selection method for p300 brain–computer interfaces," *Journal of neural engineering*, vol. 8, no. 1, p. 016 001, 2011.

[194] T. Yu, Z. Yu, Z. Gu, and Y. Li, "Grouped automatic relevance determination and its application in channel selection for p300 bcis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 6, pp. 1068–1077, 2015.

[195] S. A. Ludwig and J. Kong, "Investigation of different classifiers and channel configurations of a mobile p300-based brain–computer interface," *Medical & biological engineering & computing*, vol. 55, pp. 2143–2154, 2017.

[196] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "Toward enhanced p300 speller performance," *Journal of neuroscience methods*, vol. 167, no. 1, pp. 15–21, 2008.

[197] W. Speier, A. Deshpande, and N. Pouratian, "A method for optimizing eeg electrode number and configuration for signal acquisition in p300 speller systems," *Clinical Neurophysiology*, vol. 126, no. 6, pp. 1171–1177, 2015.

[198] M. T. McCann, D. E. Thompson, Z. H. Syed, and J. E. Huggins, "Electrode subset selection methods for an eeg-based p300 brain-computer interface," *Disability and Rehabilitation: Assistive Technology*, vol. 10, no. 3, pp. 216–220, 2015.

[199] S.-C. Noble, T. Ward, and J. V. Ringwood, "Comparing the effect of different electrode subsets on p300 speller performance," in *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE, 2023, pp. 1110–1115.

[200] R. Sahay and C. G. Brinton, "Robust subject-independent p300 waveform classification via signal pre-processing and deep learning," *IEEE Access*, vol. 9, pp. 87 579–87 591, 2021.

[201] M. Rashid, N. Sulaiman, A. PP Abdul Majeed, *et al.*, "Current status, challenges, and possible solutions of eeg-based brain-computer interface: A comprehensive review," *Frontiers in neurorobotics*, p. 25, 2020.

[202] C. Rashmi and C. Shantala, "Eeg artifacts detection and removal techniques for brain computer interface applications: A systematic review," *International Journal of Advanced Technology and Engineering Exploration*, vol. 9, no. 88, p. 354, 2022.

[203] D. Gorjan, K. Gramann, K. De Pauw, and U. Marusic, "Removal of movement-induced eeg artifacts: Current state of the art and guidelines," *Journal of neural engineering*, vol. 19, no. 1, p. 011 004, 2022.

[204] O. Dimigen, "Optimizing the ica-based removal of ocular eeg artifacts from free viewing experiments," *NeuroImage*, vol. 207, p. 116 117, 2020.

[205] X. Gao, S. Zhang, K. Liu, *et al.*, "An adaptive joint cca-ica method for ocular artifact removal and its application to emotion classification," *Journal of Neuroscience Methods*, vol. 390, p. 109 841, 2023.

[206] J. Liu, S. Ramakrishnan, S. Laxminarayan, *et al.*, "Effects of signal artefacts on electroencephalography spectral power during sleep: Quantifying the effectiveness of automated artefact-rejection algorithms," *Journal of Sleep Research*, vol. 27, no. 1, pp. 98–102, 2018.

[207] W. Shi, Y. Li, N. Cai, R. Chen, W. Cao, and J. Li, "Removal of ocular and muscular artifacts from multi-channel eeg using improved spatial-frequency filtering," *IEEE Journal of Biomedical and Health Informatics*, 2024.

[208] S. Tortora, S. Ghidoni, C. Chisari, S. Micera, and F. Artoni, "Deep learning-based bci for gait decoding from eeg with lstm recurrent neural network," *Journal of neural engineering*, vol. 17, no. 4, p. 046 011, 2020.

[209] A. I. Sburlea, L. Montesano, R. C. de la Cuerda, I. M. Alguacil Diego, J. C. Miangolarra-Page, and J. Minguez, "Detecting intention to walk in stroke patients from pre-movement eeg correlates," *Journal of neuroengineering and rehabilitation*, vol. 12, no. 1, pp. 1–12, 2015.

[210] A. Cimmino, A. Ciaramella, G. Dezio, and P. J. Salma, "Non-linear pca neural network for eeg noise reduction in brain-computer interface," *Progresses in Artificial Intelligence and Neural Systems*, pp. 405–413, 2021.

[211] E. Javed, I. Faye, A. S. Malik, and J. M. Abdullah, "Removal of bcg artefact from concurrent fmri-eeg recordings based on emd and pca," *Journal of neuroscience methods*, vol. 291, pp. 150–165, 2017.

[212] N. Bajaj, J. R. Carrión, F. Bellotti, R. Berta, and A. De Gloria, "Automatic and tunable algorithm for eeg artifact removal using wavelet decomposition

with applications in predictive modeling during auditory tasks," *Biomedical Signal Processing and Control*, vol. 55, p. 101 624, 2020.

[213] A. Narmada and M. Shukla, "A novel adaptive artifacts wavelet denoising for eeg artifacts removal using deep learning with meta-heuristic approach," *Multimedia Tools and Applications*, pp. 1–39, 2023.

[214] E. M. ter Braack, B. de Jonge, and M. J. Van Putten, "Reduction of tms induced artifacts in eeg using principal component analysis," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 21, no. 3, pp. 376–382, 2013.

[215] E. Brophy, P. Redmond, A. Fleury, M. De Vos, G. Boylan, and T. Ward, "Denoising eeg signals for real-world bci applications using gans," *Frontiers in Neuroergonomics*, vol. 2, p. 805 573, 2022.

[216] S. Gandhi, T. Oates, T. Mohsenin, and D. Hairston, "Denoising time series data using asymmetric generative adversarial networks," in *Advances in Knowledge Discovery and Data Mining: 22nd Pacific-Asia Conference, PAKDD 2018, Melbourne, VIC, Australia, June 3-6, 2018, Proceedings, Part III 22*, Springer, 2018, pp. 285–296.

[217] J. F. Hwaidi and T. M. Chen, "A noise removal approach from eeg recordings based on variational autoencoders," in *2021 13th International Conference on Computer and Automation Engineering (ICCAE)*, IEEE, 2021, pp. 19–23.

[218] H. Zhang, M. Zhao, C. Wei, D. Mantini, Z. Li, and Q. Liu, "Eegdenoisenet: A benchmark dataset for deep learning solutions of eeg denoising," *Journal of Neural Engineering*, vol. 18, no. 5, p. 056 057, 2021.

[219] V. H. G. Gomes de Novais and A. Galindo Leal, "Advancing eeg classification with transformer architecture: A case study on p300 potentials for assistive devices," in *Intelligent Systems Conference*, Springer, 2024, pp. 611–627.

[220] G. Wang, W. Liu, Y. He, C. Xu, L. Ma, and H. Li, "Eegpt: Pretrained transformer for universal and reliable representation of eeg signals," *Advances in Neural Information Processing Systems*, vol. 37, pp. 39 249–39 280, 2024.

[221] Y. Ding, Y. Li, H. Sun, *et al.*, "Eeg-deformer: A dense convolutional transformer for brain-computer interfaces," *IEEE Journal of Biomedical and Health Informatics*, 2024.

[222] Y. Song, Q. Zheng, B. Liu, and X. Gao, "Eeg conformer: Convolutional transformer for eeg decoding and visualization," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 710–719, 2023.

[223] D.-H. Shih, F.-I. Chung, T.-W. Wu, S.-Y. Huang, and M.-H. Shih, "Advanced trans-eegnet deep learning model for hypoxic-ischemic encephalopathy severity grading," *Mathematics*, vol. 12, no. 24, p. 3915, 2024.

[224] W. Zhao, X. Jiang, B. Zhang, S. Xiao, and S. Weng, "Ctnet: A convolutional transformer network for eeg-based motor imagery classification," *Scientific Reports*, vol. 14, no. 1, p. 20 237, 2024.

[225] E. Vafaei and M. Hosseini, "Transformers in eeg analysis: A review of architectures and applications in motor imagery, seizure, and emotion classification," *Sensors*, vol. 25, no. 5, p. 1293, 2025.

[226] S. Cong, H. Wang, Y. Zhou, Z. Wang, X. Yao, and C. Yang, "Comprehensive review of transformer-based models in neuroscience, neurology, and psychiatry," *Brain-X*, vol. 2, no. 2, e57, 2024.

[227] P.-L. Li and J.-J. Yuan, "Utilizing the transformer architecture combined with eegnet to achieve real-time manipulation of eeg in the metaverse.," in *2024 International Conference on System Science and Engineering (ICSSE)*, IEEE, 2024, pp. 1–8.

[228] Y. Cui, X. Shen, D. Zhang, and C. Yang, "A contrastive learning based convolutional neural network for erp brain-computer interfaces," *arXiv preprint arXiv:2407.04738*, 2024.

[229] L. Hu, W. Gao, Z. Lu, *et al.*, "Subject-independent wearable p300 brain-computer interface based on convolutional neural network and metric learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2024.

[230] X. Shen, X. Liu, X. Hu, D. Zhang, and S. Song, "Contrastive learning of subject-invariant eeg representations for cross-subject emotion recognition," *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 2496–2511, 2022.

[231] E. Shi, S. Yu, Y. Kang, *et al.*, "Meet: A multi-band eeg transformer for brain states decoding," *IEEE Transactions on Biomedical Engineering*, vol. 71, no. 5, pp. 1442–1453, 2023.

[232] Y. Shang, X. Gao, and A. An, "Multi-band spatial feature extraction and classification for motor imaging eeg signals based on osfbcsp-gao-svm model: Eeg signal processing," *Medical & Biological Engineering & Computing*, vol. 61, no. 6, pp. 1581–1602, 2023.

[233] L. Tang and M. Zhao, "Epileptic seizure detection in neonatal eeg using a multi-band graph neural network model," *Applied Sciences*, vol. 14, no. 21, p. 9712, 2024.

[234] R. Yaghoobi Karimu and S. Azadi, "Diagnosing the adhd using a mixture of expert fuzzy models," *International Journal of Fuzzy Systems*, vol. 20, pp. 1282–1296, 2018.

[235] L. Yang, D. Liu, Q. Zhang, *et al.*, "Eeg emotion recognition via identity based multi-gate mixture-of-experts network," in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2022, pp. 2498–2505.

[236] Z. Du, R. Peng, W. Liu, W. Li, and D. Wu, "Mixture of experts for eeg-based seizure subtype classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 4781–4789, 2023.

[237] E. Sankar and V. Dimitri, "Mixture of experts models in deep learning and their techniques applications and challenges," *Authorea Preprints*,

[238] M. A. Awais, M. Z. Yusoff, D. M. Khan, N. Yahya, N. Kamel, and M. Ebrahim, "Effective connectivity for decoding electroencephalographic motor imagery using a probabilistic neural network," *Sensors*, vol. 21, no. 19, p. 6570, 2021.

[239] A. R. Marathe, A. J. Ries, V. J. Lawhern, *et al.*, "The effect of target and non-target similarity on neural classification performance: A boost from confidence," *Frontiers in neuroscience*, vol. 9, p. 270, 2015.

[240] N. Bigdely-Shamlo, A. Vankov, R. R. Ramirez, and S. Makeig, "Brain activity-based image classification from rapid serial visual presentation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 16, no. 5, pp. 432–441, 2008.

[241] E. A. Pohlmeyer, J. Wang, D. C. Jangraw, B. Lou, S.-F. Chang, and P. Sajda, "Closing the loop in cortically-coupled computer vision: A brain–computer interface for searching image databases," *Journal of neural engineering*, vol. 8, no. 3, p. 036 025, 2011.

[242] J. A. Urigüen and B. Garcia-Zapirain, "Eeg artifact removal—state-of-the-art and guidelines," *Journal of neural engineering*, vol. 12, no. 3, p. 031 001, 2015.

[243] X. Jiang, G.-B. Bian, and Z. Tian, "Removal of artifacts from eeg signals: A review," *Sensors*, vol. 19, no. 5, p. 987, 2019.

[244] Z. Wang, G. Healy, A. F. Smeaton, and T. E. Ward, "An investigation of triggering approaches for the rapid serial visual presentation paradigm in brain computer interfacing," in *2016 27th Irish Signals and Systems Conference (ISSC)*, IEEE, 2016, pp. 1–6.

[245] G. Healy, Z. Wang, C. Gurrin, T. E. Ward, and A. F. Smeaton, "An eeg image-search dataset: A first-of-its-kind in ir/iir. nails: Neurally augmented image labelling strategies," 2017.

[246] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2002.

[247] P. Redmond, A. Fleury, and T. Ward, "An open source multi-modal data-acquisition platform for experimental investigation of blended control of scale vehicles," in *2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*, IEEE, 2022, pp. 673–678.

[248] M. A. Awais, P. Redmond, T. E. Ward, and G. Healy, "Amber: Advancing multimodal brain-computer interfaces for enhanced robustness-a dataset for naturalistic settings," *Frontiers in Neuroergonomics*, vol. 4, p. 1 216 440, 2023.

[249] M. A. Awais, T. E. Ward, and G. Healy, *Amber 2.0: A dataset for naturalistic settings with hmd-based rsvp tasks*, Zenodo, Jan. 2025. DOI: 10.5281/zenodo.14750342. [Online]. Available: https://doi.org/10.5281/zenodo.14750342.

[250] K. A. Ludwig, R. M. Miriani, N. B. Langhals, M. D. Joseph, D. J. Anderson, and D. R. Kipke, "Using a common average reference to improve cortical neuron recordings from microelectrode arrays," *Journal of neurophysiology*, vol. 101, no. 3, pp. 1679–1689, 2009.

[251] E. A. Katyal and R. Singla, "Eeg-based hybrid qwerty mental speller with high information transfer rate," *Medical & Biological Engineering & Computing*, vol. 59, pp. 633–661, 2021.

[252] J. Pan, X. Chen, N. Ban, J. He, J. Chen, and H. Huang, "Advances in p300 brain–computer interface spellers: Toward paradigm design and performance evaluation," *Frontiers in Human Neuroscience*, vol. 16, p. 1 077 717, 2022.

[253] H. Chang, Y. Zong, W. Zheng, *et al.*, "Eeg-based major depressive disorder recognition by selecting discriminative features via stochastic search," *Journal of Neural Engineering*, vol. 20, no. 2, p. 026 021, 2023.

[254] H. Cecotti and A. J. Ries, "Best practice for single-trial detection of event-related potentials: Application to brain-computer interfaces," *International Journal of Psychophysiology*, vol. 111, pp. 156–169, 2017.

[255] J. Han, S. Y. Kim, J. Lee, and W. H. Lee, "Brain age prediction: A comparison between machine learning models using brain morphometric data," *Sensors*, vol. 22, no. 20, p. 8077, 2022.

[256] Y. Liu, L. Sun, C. Du, and X. Wang, "Near-infrared prediction of edible oil frying times based on bayesian ridge regression," *Optik*, vol. 218, p. 164 950, 2020.

[257] B. Ren and Q. Zhou, "Assessing passengers' motion sickness levels based on cerebral blood oxygen signals and simulation of actual ride sensation," *Diagnostics*, vol. 13, no. 8, p. 1403, 2023.

[258] Q. Shi, M. Abdel-Aty, and J. Lee, "A bayesian ridge regression analysis of congestion's impact on urban expressway safety," *Accident Analysis & Prevention*, vol. 88, pp. 124–137, 2016.

[259] D. J. MacKay, "Bayesian interpolation," *Neural computation*, vol. 4, no. 3, pp. 415–447, 1992.

[260] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *Journal of machine learning research*, vol. 1, no. Jun, pp. 211–244, 2001.

[261] K. Wang, S. Qiu, W. Wei, *et al.*, "Vigilance estimating in ssvep-based bci using multimodal signals," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2021, pp. 5974–5978.

[262] P.-J. Kindermans, M. Tangermann, K.-R. Müller, and B. Schrauwen, "Integrating dynamic stopping, transfer learning and language models in an adaptive zero-training erp speller," *Journal of neural engineering*, vol. 11, no. 3, p. 035 005, 2014.

[263] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[264] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[265] S. Lees, P. McCullagh, P. Payne, L. Maguire, F. Lotte, and D. Coyle, "Speed of rapid serial visual presentation of pictures, numbers and words affects event-related potential-based detection accuracy," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 1, pp. 113–122, 2019.

[266] S. Liu, W. Wang, Y. Sheng, L. Zhang, M. Xu, and D. Ming, "Improving the cross-subject performance of the erp-based brain–computer interface using rapid serial visual presentation and correlation analysis rank," *Frontiers in human neuroscience*, vol. 14, p. 296, 2020.

[267] Y. Cui, S. Xie, X. Xie, X. Zhang, and X. Liu, "Dynamic probability integration for electroencephalography-based rapid serial visual presentation performance enhancement: Application in nighttime vehicle detection," *Frontiers in Computational Neuroscience*, vol. 16, p. 1 006 361, 2022.

[268] D. V. Moretti, F. Babiloni, F. Carducci, *et al.*, "Computerized processing of eeg–eog–emg artifacts for multi-centric studies in eeg oscillations and event-related potentials," *International Journal of Psychophysiology*, vol. 47, no. 3, pp. 199–216, 2003.

[269] A. Kabbara, N. Forde, C. Maumet, and M. Hassan, "Successful reproduction of a large eeg study across software packages," *Neuroimage: Reports*, vol. 3, no. 2, p. 100 169, 2023.

[270] H. M. K. Yip, L. Y. Cheung, V. S. Ngan, Y. K. Wong, and A. C.-N. Wong, "The effect of task on object processing revealed by eeg decoding," *European Journal of Neuroscience*, vol. 55, no. 5, pp. 1174–1199, 2022.

[271] F. M. Barbey, F. R. Farina, A. R. Buick, *et al.*, "Neuroscience from the comfort of your home: Repeated, self-administered wireless dry eeg measures brain function with high fidelity," *Frontiers in Digital Health*, vol. 4, p. 944 753, 2022.

[272] N. Kasthuri, R. Ramyea, V. Arunprasshath, S. Abhineeth, and S. Bharathraj, "Eeg conformer model based epileptic seizure prediction using deep learning," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, IEEE, 2024, pp. 1–7.

[273] B. V. Phanikrishna, P. Pławiak, and A. J. Prakash, "A brief review on eeg signal pre-processing techniques for real-time brain-computer interface applications," *Authorea Preprints*, 2021.

[274] S. Mu and S. Lin, "A comprehensive survey of mixture-of-experts: Algorithms, theory, and applications," *arXiv preprint arXiv:2503.07137*, 2025.

[275] X.-H. Liu, W.-B. Jiang, W.-L. Zheng, and B.-L. Lu, "Moge: Mixture of graph experts for cross-subject emotion recognition via decomposing eeg," in *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2024, pp. 3515–3520.

[276] H. Deng, M. Li, J. Li, M. Guo, and G. Xu, "A robust multi-branch multi-attention-mechanism eegnet for motor imagery bci decoding," *Journal of Neuroscience Methods*, vol. 405, p. 110 108, 2024.

[277] H. Zhang, Z. Wang, Y. Yu, H. Yin, C. Chen, and H. Wang, "An improved eegnet for single-trial eeg classification in rapid serial visual presentation task," *Brain Science Advances*, vol. 8, no. 2, pp. 111–126, 2022.

[278] Y. Zhu, Y. Li, J. Lu, and P. Li, "Eegnet with ensemble learning to improve the cross-session classification of ssvep based bci from ear-eeg," *IEEE Access*, vol. 9, pp. 15 295–15 303, 2021.

[279] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive mixtures of local experts," *Neural computation*, vol. 3, no. 1, pp. 79–87, 1991.

[280] W. Fedus, J. Dean, and B. Zoph, "A review of sparse expert models in deep learning," *arXiv preprint arXiv:2209.01667*, 2022.

[281] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[282] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Advances in neural information processing systems*, vol. 27, 2014.

[283] T. Lin, Y. Wang, X. Liu, and X. Qiu, "A survey of transformers," *AI open*, vol. 3, pp. 111–132, 2022.

[284] Y. Tay, M. Dehghani, D. Bahri, and D. Metzler, "Efficient transformers: A survey," *ACM Computing Surveys*, vol. 55, no. 6, pp. 1–28, 2022.

[285] P. H. Le-Khac, G. Healy, and A. F. Smeaton, "Contrastive representation learning: A framework and review," *Ieee Access*, vol. 8, pp. 193 907–193 934, 2020.

[286] H. Hu, X. Wang, Y. Zhang, Q. Chen, and Q. Guan, "A comprehensive survey on contrastive learning," *Neurocomputing*, p. 128 645, 2024.

[287] W. Li, H. Li, X. Sun, *et al.*, "Self-supervised contrastive learning for eeg-based cross-subject motor imagery recognition," *Journal of Neural Engineering*, vol. 21, no. 2, p. 026 038, 2024.

[288] C.-S. Chen and C.-S. Wei, "Mind's eye: Image recognition by eeg via multimodal similarity-keeping contrastive learning," *arXiv preprint arXiv:2406.16910*, 2024.

Appendix A

Appendix A: Plain Language Statement and Informed Consent

DUBLIN CITY UNIVERSITY

Plain Language Statement

Purpose of the Study

My name is Muhammad Ahsan Awais and in this study I and my colleagues from the School of Computing at Dublin City University, Dr. Graham Healy, Prof. Tomás Ward and Peter Redmond, are conducting research in the area of EEG (Electroencephalography) denoising by using real-time sensor streams like video captured from participants.

We are researching how sensor sources like real-time video can be used to improve the signal quality characteristics of EEG measurements. Noise in EEG signals, often as a result of physical movements, has an impact on the quality of EEG measurements, and in turn creates a significant barrier to its adoption for Brain-computer Interfacing (BCI) applications.

By using the data (EEG, video, etc) collected during this experiment, we aim to create automated EEG denoising techniques that will enable EEG to be effectively used in the future in situations where noise is inherent.

The study is concerned with recording different types of artifacts along with

clean EEG data, and then developing robust algorithms for SNR improvement and validating them in real-world scenarios.

This experiment (from start to finish) will require approximately 120 minutes of your time.

What will your involvement in the research study require?

1. Being connected to an EEG (Electroencephalography) system for recording, and similarly having your movements recorded by video.

2. Completing a number of short tasks during which you can take breaks (self-paced), including:

    - Eye open baselines (e.g., stare at the screen)
    - Closed eye baselines (e.g., eyes closed, sitting still)
    - Producing artifacts in the EEG through instructed behaviour, e.g., (move head left to right vs up and down, eye movement, eye blinks, facial expressions, lip movement, body movement, etc.)
    - Completing (c) while attending to a computer screen or HMD (head mounted display) to detect image targets, e.g., "count the number of images you see with dogs in the following sequence of images".

3. Allowing for EEG (Electroencephalography) to be disconnected*

4. Be debriefed and allowed to ask questions e.g., on the experiment's purpose.

The experiment will be broken into 50 short blocks (each lasting approximately 60 seconds).

* Conductive gel is used to form a link between the electrodes and your scalp. The experience is very comfortable, but it does mean that you will have water-soluble gel in your hair at the end of the experiment. This gel washes out easily with warm water.

Who has approved this study?

This study has been reviewed and received ethical approval from DCU Research Ethics Committee DCUREC/2021/175).

Do you have to take part?

No, you are under no obligation whatsoever to take part in this research. Participation is entirely voluntary. However, we hope that you will agree to take part and give us some of your time to participate in this experiment. It is entirely up to you to decide whether or not you would like to take part. If you decide to do so, you will be asked to sign a consent form and given a copy and the information sheet for your own records. If you decide to take part, you are still free to withdraw at any time without giving a reason and/or to withdraw your information up until such time as the research findings are anonymized and published. A decision to withdraw at any time, or a decision not to take part, will not affect your relationships with the school of computing, DCU, or have any other negative consequences for you.

What information will be collected?

Your brain activity (EEG) will be recorded. Video of your face and eyes will also be recorded during the task. Will your participation in the study be kept confidential? Yes, all data that is collected about you during the course of the research will be kept confidential. No names will be identified at any time. All hard copy information will be held in a locked cabinet at the researchers' place of work, electronic information will be encrypted and held securely on DCU PC or servers and will be accessed only by myself and my supervisors, Dr Graham Healy and Prof Tomas Ward. No information will be distributed to any other unauthorised individual or third party. If you so wish, the data that you provide can also be made available to you at your own discretion.

'It must be recognised that, in some circumstances, confidentiality of research data and records may be overridden by courts in the event of litigation or in the course of investigation by lawful authority. In such circumstances the University will take all reasonable steps within law to ensure that confidentiality is maintained

to the greatest possible extent.'

What will happen to the information that you give?

All the information you provide will be kept at DCU in such a way that it will not be possible to identify you. On completion of the research, the data will be retained on the DCU server. We will make the de-identified data available to support other researchers in carrying out research in this area (and not to be burdened with the data collection process). This means your data will be made indefinitely available for others to analyse and derive results on and carry out activities such as publishing research that will use this data. This data will be de-identified hence only your movements (extracted from video) and brain signals, will be made indefinitely available to approved researchers. Your name or other personally identifiable information will not be made available so as for you to remain anonymous to anybody who uses this dataset.

What will happen to the results?

The research will be discussed at internal group meetings and written up and presented at national and international conferences. It will also be published in scientific journals. A copy of the research findings will be made available to you upon request.

What are the possible disadvantages of taking part?

In order to record your brain activity (EEG) a gel needs to be applied to certain locations on your scalp. This can be removed easily with a brush or paper towel. To make sure that you are not allergic to this gel, a patch test will be carried out on your forearm at least 24 hours before the experiment. If you have an allergic reaction, you cannot participate in the experiment and might want to contact your GP (general practitioner).

The cap used to measure the EEG is meant to fit tightly and, therefore, might be a little uncomfortable after a while. Breaks can be taken, or the experiment can be stopped, should this be the case. I don't envisage any other negative consequences for you in taking part.

What if there is a problem?

You may contact my supervisors, Dr. Graham Healy (graham.healy@dcu.ie) or Prof Tomás Ward (tomas.ward@dcu.ie) if you feel the research has not been carried out as described above.

Any further queries?

If you need any further information, you can contact me: Muhammad Ahsan Awais (muhammad.awais2@mail.dcu.ie)

If you agree to take part in the study, please complete and sign the consent form attached.

If participants have concerns about this study and wish to contact an independent person, please contact:

The Secretary, Dublin City University Research Ethics Committee, c/o Research and Innovation Support, Dublin City University, Dublin 9. Tel 01-7008000, e-mail rec@dcu.ie

DUBLIN CITY UNIVERSITY

Informed Consent Form

I ..... agree to participate in Muhammad Ahsan Awais's research study titled "Multimodal Brain-computer Interface Denoising" for the School of Computing DCU, as described to me in the "Plain Language Statement".

Please complete the following ([x] or [/]):

• I have read the Plain Language Statement. [ ]

• The purpose and nature of the study have been explained to me verbally and in writing. I've been able to ask questions, which were answered satisfactorily. [ ]

• I am participating voluntarily. [ ]

• I give permission for my brain activity (EEG) to be recorded. [ ]

• I give permission for a video of my face during the entire experiment to be recorded. [ ]

• I understand that I can withdraw from the study, without repercussions, at any time, whether that is before it starts or while I am participating. [ ]

• I understand that I can withdraw permission to use the data right up to its anonymisation. [ ]

• It has been explained to me how my data will be managed and that I may access it on request. [ ]

• I understand the limits of confidentiality as described in the information sheet. [ ]

• I understand that my data, in an anonymous format, may be used in further research projects and any subsequent publications if I give permission below. [ ]

• I agree that my data to be used for further research projects and subsequent publications. [ ]

• I am aware that my de-identified data will be made publically available for researchers to analyze and derive results and carry out activities such as publishing research that will use this data. [ ]

Signature:

I have read and understood the information in this form. My questions and concerns have been answered by the researchers, and I have a copy of this consent form. Therefore, I consent to take part in this research.

Participants' Signature: ..... Date: .....

Name in Block Capitals: .....

Researchers' Signature: ..... Date: .....

Name in Block Capitals: .....

Appendix B

Appendix B: Subject Independent Classification - Model Hyperparameter and Extended Evaluation Metrics

B.1 Hyperparameter

- **CNN-1**

  L1_filter: 20

  L2_filter: 25

  L2 kernel_length: 20

  L2-strides: 5

  L3-dense: 50

  batch_size: 32

  learning_rate: 0.0005

- **EEGNet**

  L1_filter: 4

  L2_filter: 16

kernel_length: 32

dropout: 0.3

depth_multiplier: 2

batch_size: 32

learning_rate: 0.001

- **MoE-EEGNet**

  L1_filter: 4

  L2_filter: 16

  kernel_length: 32

  dropout: 0.3

  depth_multiplier: 2

  batch_size: 32

  learning_rate: 0.001

  num_experts: 4

  expert_dim: 128

- **Multiband EEGNet**

  L1_filter: 4

  L2_filter: 16

  kernel_length: 32

  dropout: 0.3

  depth_multiplier: 2

  batch_size: 32

  learning_rate: 0.001

- **Multiband MoE EEGNet**

  L1_filter: 4

  L2_filter: 16

  kernel_length: 32

dropout: 0.3

depth_multiplier: 2

batch_size: 64

learning_rate: 0.001

num_experts: 4

expert_dim: 64

- **EEGTransformer**

  num_heads: 8

  num_layers: 2

  dim_feedforward: 128

  dropout: 0.3

  batch_size: 128

  learning_rate: 0.0001

- **MoE Transformer**

  num_heads: 4

  dim_feedforward: 128

  dropout: 0.3

  batch_size: 64

  learning_rate: 0.001

  num_experts: 6

  expert_dim: 64

- **EEG Conformer**

  filter_time_length=25

  pool_time_length=4

  pool_time_stride=2

drop_prob=0.5

att_depth=2

att_heads=8

att_drop_prob=0.4

learning_rate: 0.001

batch_size: 64

- **CNN1-Transformer**

  L1_filter: 20

  L2_filter: 25

  L2 kernel_length: 20

  L2-strides: 5

  L3-dense: 50

  batch_size: 32

  learning_rate: 0.0005

  num_heads: 4

  dim_feedforward: 128

- **EEGNet-Transformer**

  L1_filter: 4

  L2_filter: 8

  kernel_length: 8

  dropout: 0.3

  depth_multiplier: 2

  batch_size: 64

  learning_rate: 0.001

  num_heads: 2

  num_layers: 2

  dim_feedforward: 128

- **Multiband EEGNet-Transformer**

  L1_filter: 4

  L2_filter: 16

  kernel_length: 32

  dropout: 0.3

  depth_multiplier: 2

  batch_size: 64

  learning_rate: 0.001

  num_heads: 4

  num_layers: 2

  dim_feedforward: 128

- **EEGNet-MoE Transformer**

  L1_filter: 4

  L2_filter: 8

  kernel_length: 32

  dropout: 0.3

  depth_multiplier: 2

  batch_size: 128

  learning_rate: 0.0001

  num_heads: 8

  num_layers: 2

  dim_feedforward: 128

  num_experts: 4

  expert_dim: 128

- **Multiband EEGNet-MoE Transformer**

  L1_filter: 4

  L2_filter: 16

  kernel_length: 32

  dropout: 0.3

depth_multiplier: 2

batch_size: 64

learning_rate: 0.001

num_heads: 4

num_layers: 2

dim_feedforward: 128

num_experts: 4

expert_dim: 64

B.2 Subject Independent Classification Results

Table B.1: Bayesian Ridge Regression: Evaluation metrics for classification model trained on traditional (clean) RSVP data.

Bayesian Ridge Regression : Train RSVP																								
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.535	0.561	0.453	0.537	0.596	0.599	0.557	0.632	0.135	0.138	0.118	0.146	0.674	0.646	0.688	0.75	0.225	0.227	0.201	0.245	0.622	0.55	0.507	0.649
P2	0.603	0.661	0.642	0.61	0.664	0.676	0.653	0.691	0.166	0.184	0.171	0.177	0.74	0.694	0.667	0.792	0.271	0.291	0.272	0.289	0.646	0.642	0.609	0.691
P3	0.645	0.682	0.56	0.797	0.65	0.638	0.558	0.662	0.17	0.174	0.123	0.244	0.656	0.583	0.556	0.493	0.27	0.268	0.202	0.326	0.672	0.665	0.539	0.739
P4	0.677	0.693	0.722	0.782	0.733	0.74	0.574	0.749	0.209	0.218	0.152	0.273	0.802	0.799	0.389	0.708	0.332	0.342	0.219	0.394	0.744	0.711	0.521	0.748
P5	0.792	0.748	0.733	0.765	0.784	0.749	0.586	0.786	0.295	0.248	0.162	0.273	0.774	0.75	0.403	0.812	0.427	0.373	0.232	0.409	0.844	0.824	0.564	0.849
P6	0.745	0.616	0.554	0.776	0.732	0.648	0.579	0.666	0.24	0.163	0.131	0.23	0.715	0.688	0.611	0.528	0.359	0.264	0.215	0.32	0.773	0.653	0.581	0.661
P7	0.689	0.735	0.703	0.724	0.696	0.686	0.625	0.702	0.2	0.216	0.175	0.217	0.705	0.625	0.528	0.674	0.312	0.321	0.263	0.328	0.776	0.722	0.607	0.719
P8	0.663	0.447	0.691	0.624	0.646	0.603	0.652	0.655	0.173	0.13	0.183	0.168	0.625	0.799	0.604	0.694	0.27	0.224	0.281	0.27	0.602	0.579	0.566	0.614
P9	0.762	0.705	0.766	0.785	0.76	0.756	0.676	0.699	0.262	0.228	0.228	0.254	0.757	0.819	0.562	0.59	0.389	0.357	0.325	0.355	0.799	0.793	0.707	0.738
P10	0.825	0.757	0.746	0.681	0.758	0.723	0.581	0.631	0.322	0.244	0.164	0.171	0.674	0.681	0.375	0.569	0.435	0.359	0.228	0.263	0.866	0.788	0.549	0.661
P11	0.864	0.821	0.606	0.784	0.761	0.78	0.63	0.787	0.389	0.324	0.155	0.289	0.632	0.729	0.66	0.792	0.481	0.449	0.251	0.423	0.763	0.743	0.623	0.807
P12	0.794	0.709	0.654	0.789	0.781	0.706	0.583	0.713	0.296	0.212	0.143	0.263	0.764	0.701	0.493	0.618	0.426	0.325	0.222	0.369	0.833	0.762	0.553	0.764
P13	0.787	0.685	0.578	0.74	0.666	0.628	0.547	0.674	0.239	0.171	0.12	0.212	0.514	0.556	0.507	0.59	0.326	0.261	0.194	0.312	0.671	0.649	0.55	0.659
P14	0.78	0.714	0.695	0.763	0.754	0.693	0.621	0.723	0.273	0.209	0.17	0.248	0.722	0.667	0.528	0.674	0.396	0.318	0.257	0.363	0.806	0.751	0.669	0.757
P15	0.832	0.566	0.672	0.731	0.743	0.626	0.599	0.703	0.325	0.148	0.154	0.221	0.632	0.701	0.507	0.667	0.429	0.244	0.236	0.332	0.777	0.629	0.546	0.673
P16	0.584	0.706	0.631	0.581	0.544	0.571	0.52	0.582	0.119	0.146	0.111	0.134	0.493	0.403	0.382	0.583	0.192	0.215	0.172	0.218	0.512	0.531	0.503	0.555
P17	0.549	0.65	0.707	0.36	0.663	0.605	0.544	0.58	0.157	0.153	0.13	0.12	0.806	0.549	0.34	0.854	0.263	0.239	0.188	0.211	0.668	0.541	0.512	0.471
P18	0.69	0.732	0.548	0.694	0.676	0.657	0.607	0.669	0.193	0.2	0.139	0.191	0.66	0.562	0.681	0.639	0.298	0.296	0.231	0.294	0.716	0.733	0.614	0.695
P19	0.826	0.81	0.606	0.758	0.761	0.731	0.661	0.671	0.323	0.292	0.166	0.221	0.681	0.632	0.729	0.562	0.438	0.399	0.27	0.318	0.786	0.718	0.594	0.669
P20	0.69	0.813	0.815	0.616	0.729	0.714	0.582	0.685	0.213	0.288	0.203	0.176	0.778	0.59	0.292	0.771	0.334	0.387	0.239	0.286	0.762	0.715	0.559	0.665
Avg	**0.712**	**0.691**	**0.654**	**0.695**	**0.705**	**0.676**	**0.597**	**0.683**	**0.235**	**0.204**	**0.155**	**0.211**	**0.690**	**0.659**	**0.525**	**0.668**	**0.344**	**0.308**	**0.235**	**0.316**	**0.732**	**0.685**	**0.574**	**0.689**

Table B.2: CNN-1: Evaluation metrics for classification model trained on traditional (clean) RSVP data.

CNN-1 : Train RSVP																								
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.577	0.642	0.57	0.676	0.582	0.585	0.57	0.625	0.133	0.143	0.126	0.167	0.587	0.514	0.611	0.562	0.217	0.223	0.209	0.258	0.595	0.6	0.555	0.661
P2	0.7	0.578	0.677	0.715	0.679	0.688	0.677	0.672	0.197	0.17	0.221	0.2	0.653	0.826	0.583	0.618	0.303	0.281	0.321	0.302	0.737	0.741	0.722	0.718
P3	0.666	0.644	0.518	0.669	0.632	0.617	0.518	0.644	0.168	0.157	0.106	0.173	0.59	0.583	0.618	0.611	0.261	0.247	0.18	0.27	0.678	0.635	0.49	0.677
P4	0.769	0.697	0.571	0.722	0.73	0.76	0.571	0.771	0.255	0.226	0.131	0.241	0.681	0.84	0.535	0.833	0.371	0.356	0.211	0.374	0.801	0.832	0.568	0.815
P5	0.809	0.771	0.57	0.689	0.785	0.758	0.57	0.781	0.312	0.268	0.157	0.23	0.753	0.743	0.347	0.896	0.442	0.393	0.216	0.365	0.852	0.814	0.538	0.842
P6	0.759	0.788	0.617	0.714	0.747	0.654	0.617	0.718	0.255	0.233	0.157	0.218	0.733	0.486	0.576	0.722	0.378	0.315	0.247	0.335	0.817	0.701	0.62	0.752
P7	0.663	0.669	0.629	0.736	0.692	0.674	0.629	0.671	0.19	0.185	0.153	0.209	0.729	0.681	0.674	0.59	0.302	0.291	0.249	0.309	0.747	0.695	0.633	0.691
P8	0.604	0.673	0.648	0.713	0.652	0.605	0.648	0.646	0.162	0.157	0.173	0.188	0.712	0.521	0.632	0.562	0.264	0.242	0.272	0.281	0.702	0.619	0.685	0.681
P9	0.75	0.739	0.663	0.69	0.759	0.75	0.663	0.716	0.254	0.243	0.186	0.208	0.771	0.764	0.632	0.75	0.382	0.369	0.288	0.326	0.83	0.815	0.725	0.774
P10	0.76	0.758	0.576	0.706	0.742	0.723	0.576	0.559	0.253	0.244	0.135	0.139	0.719	0.681	0.528	0.375	0.375	0.36	0.215	0.203	0.808	0.773	0.578	0.561
P11	0.815	0.843	0.643	0.826	0.805	0.808	0.643	0.796	0.326	0.364	0.164	0.336	0.792	0.764	0.66	0.757	0.462	0.493	0.262	0.466	0.876	0.863	0.672	0.863
P12	0.873	0.801	0.592	0.733	0.784	0.726	0.592	0.737	0.416	0.28	0.14	0.235	0.674	0.632	0.583	0.743	0.515	0.388	0.225	0.357	0.849	0.796	0.586	0.798
P13	0.762	0.678	0.556	0.608	0.689	0.615	0.556	0.689	0.232	0.163	0.129	0.176	0.597	0.535	0.444	0.792	0.334	0.25	0.2	0.288	0.713	0.644	0.544	0.751
P14	0.813	0.691	0.686	0.812	0.767	0.705	0.686	0.716	0.31	0.205	0.212	0.288	0.708	0.722	0.632	0.597	0.431	0.319	0.318	0.388	0.831	0.76	0.734	0.747
P15	0.778	0.636	0.573	0.706	0.738	0.668	0.573	0.673	0.265	0.175	0.148	0.197	0.688	0.708	0.41	0.632	0.383	0.28	0.217	0.301	0.804	0.677	0.58	0.707
P16	0.454	0.615	0.541	0.496	0.552	0.595	0.541	0.581	0.116	0.143	0.116	0.127	0.674	0.569	0.549	0.688	0.198	0.228	0.191	0.214	0.548	0.621	0.523	0.598
P17	0.647	0.705	0.549	0.678	0.687	0.617	0.549	0.602	0.184	0.171	0.115	0.157	0.736	0.507	0.688	0.507	0.294	0.256	0.197	0.24	0.732	0.645	0.543	0.609
P18	0.709	0.748	0.632	0.728	0.684	0.693	0.632	0.676	0.203	0.226	0.163	0.208	0.653	0.625	0.611	0.611	0.31	0.331	0.258	0.31	0.733	0.722	0.648	0.732
P19	0.815	0.688	0.661	0.736	0.771	0.755	0.661	0.644	0.314	0.221	0.166	0.196	0.715	0.84	0.729	0.528	0.436	0.35	0.27	0.286	0.844	0.828	0.679	0.683
P20	0.76	0.728	0.593	0.691	0.749	0.688	0.593	0.72	0.256	0.213	0.161	0.21	0.736	0.639	0.444	0.757	0.38	0.319	0.236	0.329	0.798	0.762	0.62	0.753
Avg	**0.724**	**0.705**	**0.603**	**0.702**	**0.711**	**0.684**	**0.603**	**0.682**	**0.240**	**0.209**	**0.153**	**0.205**	**0.695**	**0.659**	**0.574**	**0.657**	**0.352**	**0.315**	**0.239**	**0.310**	**0.763**	**0.727**	**0.612**	**0.721**

Table B.3: EEGNet: Evaluation metrics for classification model trained on traditional (clean) RSVP data.

EEGNet : Train RSVP																								
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.551	0.396	0.328	0.672	0.61	0.569	0.546	0.679	0.141	0.119	0.111	0.188	0.684	0.785	0.819	0.688	0.233	0.206	0.196	0.295	0.632	0.61	0.529	0.733
P2	0.691	0.695	0.714	0.657	0.713	0.729	0.696	0.695	0.207	0.215	0.21	0.19	0.74	0.771	0.674	0.743	0.324	0.336	0.32	0.302	0.773	0.756	0.679	0.765
P3	0.608	0.658	0.628	0.693	0.648	0.649	0.544	0.601	0.162	0.173	0.122	0.16	0.698	0.639	0.438	0.486	0.262	0.272	0.191	0.241	0.679	0.725	0.534	0.766
P4	0.75	0.742	0.488	0.606	0.722	0.699	0.571	0.682	0.239	0.225	0.123	0.173	0.688	0.646	0.674	0.778	0.355	0.334	0.208	0.283	0.785	0.829	0.568	0.806
P5	0.835	0.768	0.399	0.699	0.776	0.717	0.586	0.774	0.342	0.249	0.123	0.232	0.701	0.653	0.819	0.868	0.46	0.36	0.214	0.366	0.845	0.762	0.533	0.796
P6	0.727	0.673	0.634	0.626	0.728	0.667	0.559	0.681	0.229	0.184	0.13	0.177	0.729	0.66	0.465	0.75	0.348	0.287	0.203	0.286	0.794	0.672	0.546	0.71
P7	0.785	0.658	0.626	0.608	0.726	0.674	0.626	0.659	0.266	0.182	0.157	0.166	0.653	0.694	0.625	0.722	0.378	0.289	0.251	0.269	0.778	0.719	0.63	0.705
P8	0.55	0.54	0.455	0.699	0.633	0.602	0.614	0.66	0.148	0.137	0.134	0.189	0.736	0.681	0.812	0.611	0.247	0.228	0.23	0.289	0.671	0.65	0.679	0.717
P9	0.777	0.758	0.694	0.697	0.782	0.752	0.673	0.745	0.281	0.256	0.193	0.221	0.788	0.743	0.646	0.806	0.415	0.381	0.297	0.347	0.845	0.845	0.73	0.816
P10	0.842	0.745	0.656	0.683	0.772	0.726	0.611	0.676	0.352	0.238	0.157	0.19	0.684	0.701	0.556	0.667	0.465	0.355	0.244	0.296	0.856	0.742	0.601	0.602
P11	0.848	0.831	0.763	0.763	0.786	0.807	0.674	0.779	0.366	0.347	0.226	0.269	0.708	0.778	0.562	0.799	0.482	0.48	0.322	0.403	0.848	0.841	0.723	0.862
P12	0.858	0.633	0.51	0.662	0.773	0.716	0.583	0.726	0.381	0.19	0.128	0.202	0.667	0.819	0.674	0.806	0.485	0.309	0.216	0.323	0.834	0.753	0.555	0.774
P13	0.651	0.626	0.451	0.661	0.664	0.592	0.544	0.701	0.177	0.143	0.114	0.193	0.681	0.549	0.66	0.75	0.281	0.227	0.194	0.307	0.709	0.643	0.492	0.688
P14	0.777	0.742	0.705	0.683	0.74	0.687	0.654	0.747	0.265	0.22	0.188	0.216	0.694	0.618	0.59	0.826	0.384	0.324	0.286	0.343	0.819	0.769	0.705	0.801
P15	0.727	0.685	0.83	0.747	0.747	0.646	0.566	0.677	0.236	0.178	0.201	0.218	0.771	0.597	0.236	0.59	0.361	0.275	0.217	0.318	0.795	0.665	0.57	0.72
P16	0.667	0.544	0.646	0.598	0.559	0.58	0.541	0.591	0.133	0.13	0.122	0.139	0.424	0.625	0.41	0.583	0.203	0.215	0.188	0.225	0.571	0.611	0.525	0.617
P17	0.693	0.451	0.534	0.61	0.703	0.602	0.534	0.601	0.204	0.13	0.113	0.145	0.715	0.792	0.535	0.59	0.318	0.224	0.187	0.233	0.74	0.66	0.554	0.605
P18	0.762	0.826	0.686	0.728	0.689	0.693	0.628	0.723	0.232	0.293	0.171	0.227	0.597	0.528	0.556	0.715	0.335	0.377	0.261	0.345	0.735	0.756	0.668	0.775
P19	0.824	0.784	0.705	0.66	0.797	0.794	0.709	0.719	0.333	0.291	0.211	0.199	0.764	0.806	0.715	0.792	0.464	0.427	0.326	0.318	0.859	0.801	0.708	0.72
P20	0.769	0.611	0.609	0.754	0.714	0.698	0.61	0.691	0.248	0.179	0.148	0.228	0.646	0.806	0.611	0.611	0.358	0.293	0.238	0.332	0.761	0.745	0.6	0.715
Avg	**0.735**	**0.668**	**0.603**	**0.675**	**0.714**	**0.680**	**0.603**	**0.690**	**0.247**	**0.204**	**0.154**	**0.196**	**0.688**	**0.694**	**0.604**	**0.709**	**0.358**	**0.310**	**0.239**	**0.306**	**0.767**	**0.728**	**0.606**	**0.735**

Table B.4: MoE-EEGNet: Evaluation metrics for classification model trained on traditional (clean) RSVP data.

MoE-EEGNet : Train RSVP																								
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.532	0.594	Accuracy	0.711	0.598	0.593	0.552	0.648	0.135	0.139	0.132	0.188	0.681	0.59	0.382	0.569	0.225	0.225	0.196	0.283	0.635	0.615	0.543	0.681
P2	0.78	0.769	0.688	0.688	0.728	0.748	0.689	0.728	0.263	0.263	0.235	0.212	0.663	0.722	0.59	0.778	0.376	0.385	0.337	0.333	0.795	0.809	0.713	0.772
P3	0.709	0.594	0.767	0.544	0.644	0.632	0.579	0.636	0.185	0.154	0.129	0.148	0.562	0.681	0.632	0.75	0.279	0.251	0.214	0.248	0.679	0.673	0.557	0.676
P4	0.631	0.66	0.536	0.735	0.696	0.7	0.575	0.683	0.183	0.192	0.126	0.214	0.778	0.75	0.653	0.618	0.297	0.306	0.211	0.318	0.757	0.764	0.574	0.724
P5	0.829	0.748	0.513	0.731	0.78	0.746	0.581	0.773	0.335	0.247	0.132	0.247	0.719	0.743	0.604	0.826	0.457	0.371	0.216	0.38	0.851	0.812	0.588	0.831
P6	0.723	0.722	0.562	0.726	0.746	0.62	0.6	0.678	0.233	0.178	0.139	0.207	0.774	0.493	0.639	0.618	0.358	0.262	0.229	0.311	0.814	0.645	0.618	0.721
P7	0.745	0.682	0.569	0.586	0.664	0.657	0.635	0.647	0.21	0.182	0.165	0.158	0.562	0.625	0.618	0.722	0.306	0.282	0.261	0.259	0.726	0.692	0.639	0.681
P8	0.515	0.635	0.649	0.531	0.627	0.584	0.675	0.628	0.143	0.141	0.18	0.144	0.767	0.521	0.708	0.75	0.24	0.222	0.287	0.242	0.671	0.589	0.688	0.647
P9	0.752	0.758	0.648	0.663	0.766	0.758	0.719	0.723	0.257	0.258	0.22	0.201	0.785	0.757	0.722	0.799	0.387	0.385	0.337	0.322	0.836	0.834	0.768	0.782
P10	0.792	0.801	0.716	0.542	0.752	0.751	0.572	0.545	0.283	0.291	0.136	0.117	0.701	0.688	0.486	0.549	0.403	0.409	0.213	0.193	0.823	0.8	0.577	0.538
P11	0.765	0.74	0.64	0.756	0.743	0.779	0.664	0.778	0.257	0.254	0.176	0.264	0.715	0.826	0.688	0.806	0.378	0.389	0.28	0.397	0.823	0.846	0.712	0.838
P12	0.794	0.722	0.646	0.774	0.722	0.66	0.569	0.665	0.272	0.198	0.125	0.228	0.632	0.583	0.625	0.528	0.38	0.296	0.208	0.319	0.783	0.718	0.552	0.705
P13	0.785	0.546	0.524	0.634	0.662	0.621	0.528	0.679	0.235	0.144	0.139	0.178	0.507	0.715	0.181	0.736	0.321	0.24	0.157	0.287	0.685	0.645	0.509	0.738
P14	0.678	0.841	0.806	0.806	0.747	0.733	0.664	0.747	0.215	0.335	0.176	0.294	0.833	0.597	0.688	0.674	0.341	0.429	0.28	0.409	0.822	0.794	0.7	0.802
P15	0.75	0.778	0.646	0.619	0.738	0.648	0.647	0.684	0.245	0.222	0.165	0.176	0.722	0.486	0.674	0.764	0.366	0.304	0.265	0.286	0.817	0.676	0.651	0.727
P16	0.438	0.609	0.626	0.533	0.543	0.579	0.514	0.561	0.113	0.136	0.104	0.123	0.674	0.542	0.743	0.597	0.193	0.217	0.182	0.204	0.531	0.588	0.486	0.567
P17	0.6	0.625	0.331	0.658	0.664	0.606	0.533	0.548	0.166	0.149	0.108	0.127	0.743	0.583	0.806	0.41	0.271	0.237	0.19	0.193	0.701	0.632	0.519	0.526
P18	0.735	0.744	0.315	0.756	0.652	0.654	0.61	0.673	0.2	0.205	0.15	0.22	0.549	0.542	0.597	0.569	0.293	0.298	0.24	0.318	0.695	0.684	0.626	0.725
P19	0.774	0.743	0.621	0.675	0.732	0.734	0.668	0.674	0.259	0.24	0.198	0.187	0.681	0.722	0.611	0.674	0.375	0.36	0.299	0.293	0.799	0.8	0.696	0.723
P20	0.669	0.681	0.713	0.812	0.72	0.724	0.595	0.695	0.202	0.208	0.143	0.277	0.785	0.778	0.569	0.549	0.321	0.328	0.228	0.368	0.783	0.783	0.605	0.759
Avg	**0.700**	**0.700**	**0.615**	**0.674**	**0.696**	**0.676**	**0.608**	**0.670**	**0.220**	**0.207**	**0.154**	**0.196**	**0.692**	**0.647**	**0.611**	**0.664**	**0.328**	**0.310**	**0.242**	**0.298**	**0.751**	**0.720**	**0.616**	**0.708**

Table B.5: Multiband EEGNet: Evaluation metrics for classification model trained on traditional (clean) RSVP data.

Multiband EEGNet : Train RSVP																								
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.643	0.572	0.526	0.735	0.621	0.605	0.567	0.671	0.158	0.141	0.124	0.209	0.594	0.646	0.618	0.59	0.25	0.232	0.207	0.309	0.658	0.607	0.575	0.712
P2	0.699	0.673	0.695	0.635	0.685	0.695	0.698	0.723	0.199	0.194	0.203	0.193	0.667	0.722	0.701	0.833	0.307	0.306	0.315	0.314	0.76	0.74	0.763	0.802
P3	0.698	0.785	0.448	0.719	0.675	0.633	0.539	0.668	0.195	0.218	0.112	0.2	0.646	0.444	0.653	0.604	0.3	0.292	0.191	0.301	0.726	0.68	0.494	0.694
P4	0.786	0.805	0.566	0.814	0.763	0.777	0.589	0.782	0.282	0.305	0.135	0.317	0.733	0.743	0.618	0.743	0.407	0.432	0.222	0.444	0.837	0.836	0.596	0.842
P5	0.857	0.858	0.759	0.687	0.781	0.723	0.579	0.743	0.38	0.362	0.167	0.216	0.688	0.556	0.354	0.812	0.489	0.438	0.227	0.342	0.855	0.795	0.577	0.808
P6	0.718	0.764	0.629	0.769	0.74	0.634	0.606	0.659	0.229	0.205	0.149	0.221	0.767	0.472	0.576	0.521	0.352	0.286	0.237	0.311	0.803	0.663	0.63	0.657
P7	0.728	0.711	0.537	0.637	0.718	0.704	0.598	0.703	0.225	0.212	0.135	0.187	0.705	0.694	0.674	0.785	0.341	0.325	0.226	0.302	0.78	0.724	0.607	0.771
P8	0.645	0.613	0.549	0.685	0.65	0.665	0.614	0.615	0.17	0.169	0.142	0.165	0.656	0.729	0.694	0.528	0.27	0.274	0.236	0.251	0.697	0.691	0.636	0.638
P9	0.747	0.77	0.81	0.662	0.776	0.733	0.684	0.713	0.258	0.257	0.27	0.198	0.812	0.688	0.528	0.778	0.391	0.374	0.357	0.315	0.852	0.78	0.702	0.786
P10	0.804	0.766	0.695	0.672	0.789	0.697	0.574	0.633	0.308	0.238	0.146	0.169	0.771	0.611	0.424	0.583	0.44	0.343	0.217	0.262	0.865	0.719	0.563	0.647
P11	0.803	0.849	0.795	0.807	0.78	0.771	0.593	0.794	0.304	0.362	0.197	0.313	0.75	0.674	0.34	0.778	0.433	0.471	0.249	0.446	0.837	0.843	0.614	0.862
P12	0.822	0.664	0.544	0.806	0.799	0.721	0.592	0.738	0.331	0.201	0.134	0.291	0.771	0.792	0.653	0.653	0.463	0.32	0.222	0.403	0.866	0.793	0.593	0.795
P13	0.653	0.563	0.77	0.74	0.681	0.631	0.545	0.683	0.183	0.149	0.144	0.217	0.715	0.715	0.264	0.611	0.292	0.247	0.187	0.32	0.731	0.662	0.538	0.713
P14	0.742	0.753	0.613	0.728	0.74	0.696	0.674	0.704	0.241	0.23	0.172	0.22	0.736	0.625	0.75	0.674	0.364	0.336	0.279	0.332	0.802	0.773	0.715	0.725
P15	0.81	0.615	0.722	0.731	0.734	0.635	0.583	0.696	0.293	0.158	0.158	0.218	0.639	0.66	0.41	0.653	0.402	0.255	0.228	0.326	0.799	0.676	0.595	0.726
P16	0.643	0.69	0.829	0.626	0.57	0.584	0.535	0.598	0.136	0.15	0.16	0.146	0.479	0.451	0.167	0.562	0.212	0.226	0.163	0.231	0.566	0.589	0.503	0.611
P17	0.776	0.64	0.44	0.492	0.688	0.627	0.559	0.625	0.241	0.16	0.118	0.14	0.576	0.611	0.708	0.792	0.34	0.254	0.202	0.238	0.733	0.645	0.549	0.656
P18	0.722	0.745	0.605	0.724	0.71	0.71	0.626	0.693	0.219	0.231	0.153	0.213	0.694	0.667	0.653	0.653	0.333	0.343	0.248	0.321	0.757	0.745	0.644	0.747
P19	0.835	0.831	0.669	0.772	0.772	0.745	0.65	0.719	0.34	0.324	0.176	0.252	0.694	0.639	0.625	0.653	0.457	0.43	0.274	0.364	0.832	0.815	0.668	0.786
P20	0.828	0.736	0.683	0.797	0.769	0.733	0.592	0.736	0.329	0.235	0.153	0.281	0.694	0.729	0.479	0.66	0.446	0.356	0.232	0.394	0.833	0.8	0.618	0.81
Avg	**0.748**	**0.720**	**0.6442**	**0.712**	**0.722**	**0.686**	**0.600**	**0.695**	**0.251**	**0.225**	**0.157**	**0.218**	**0.689**	**0.643**	**0.544**	**0.673**	**0.364**	**0.327**	**0.236**	**0.326**	**0.779**	**0.729**	**0.609**	**0.739**

Table B.6: Multiband MoE-EEGNet: Evaluation metrics for classification model trained on traditional (clean) RSVP data.

Multiband MoE-EEGNet : Train RSVP																								
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.59	0.699	0.61	0.617	0.62	0.579	0.567	0.645	0.149	0.15	0.131	0.163	0.656	0.431	0.514	0.681	0.243	0.222	0.208	0.262	0.643	0.588	0.567	0.683
P2	0.671	0.645	0.783	0.708	0.674	0.698	0.682	0.748	0.186	0.187	0.243	0.227	0.677	0.764	0.556	0.799	0.292	0.301	0.338	0.353	0.747	0.739	0.727	0.804
P3	0.655	0.766	0.506	0.653	0.682	0.685	0.546	0.675	0.184	0.233	0.116	0.181	0.715	0.583	0.597	0.701	0.293	0.333	0.195	0.288	0.733	0.721	0.514	0.712
P4	0.773	0.714	0.551	0.726	0.757	0.739	0.602	0.746	0.269	0.227	0.138	0.235	0.736	0.771	0.667	0.771	0.394	0.35	0.229	0.36	0.824	0.809	0.609	0.826
P5	0.73	0.726	0.72	0.715	0.76	0.727	0.564	0.746	0.242	0.228	0.145	0.229	0.799	0.729	0.368	0.785	0.372	0.347	0.208	0.355	0.834	0.774	0.546	0.814
P6	0.791	0.643	0.465	0.704	0.748	0.675	0.616	0.666	0.281	0.179	0.135	0.193	0.694	0.715	0.806	0.618	0.4	0.286	0.231	0.295	0.816	0.72	0.635	0.711
P7	0.799	0.688	0.596	0.816	0.717	0.678	0.64	0.737	0.274	0.193	0.157	0.302	0.615	0.667	0.694	0.639	0.379	0.299	0.256	0.41	0.785	0.737	0.645	0.778
P8	0.627	0.617	0.779	0.735	0.645	0.639	0.64	0.618	0.164	0.16	0.218	0.182	0.667	0.667	0.465	0.472	0.263	0.258	0.296	0.263	0.695	0.647	0.656	0.628
P9	0.79	0.76	0.806	0.714	0.788	0.734	0.673	0.718	0.294	0.25	0.26	0.218	0.785	0.701	0.507	0.722	0.428	0.369	0.344	0.335	0.858	0.784	0.704	0.778
P10	0.717	0.74	0.762	0.714	0.752	0.723	0.587	0.619	0.232	0.233	0.174	0.175	0.795	0.701	0.368	0.5	0.359	0.35	0.236	0.259	0.839	0.75	0.599	0.647
P11	0.812	0.87	0.735	0.767	0.766	0.758	0.618	0.769	0.309	0.403	0.182	0.269	0.708	0.618	0.472	0.771	0.43	0.488	0.263	0.399	0.824	0.829	0.648	0.829
P12	0.813	0.628	0.647	0.683	0.754	0.691	0.578	0.728	0.305	0.181	0.14	0.21	0.681	0.771	0.493	0.785	0.422	0.293	0.218	0.331	0.82	0.747	0.583	0.796
P13	0.755	0.556	0.57	0.624	0.706	0.649	0.554	0.698	0.235	0.154	0.122	0.182	0.646	0.764	0.535	0.792	0.345	0.256	0.199	0.296	0.758	0.675	0.559	0.755
P14	0.84	0.818	0.676	0.797	0.766	0.711	0.678	0.721	0.346	0.292	0.189	0.274	0.674	0.576	0.681	0.625	0.458	0.388	0.296	0.381	0.83	0.785	0.724	0.771
P15	0.749	0.675	0.62	0.794	0.74	0.665	0.61	0.688	0.245	0.184	0.15	0.256	0.729	0.653	0.597	0.556	0.367	0.287	0.239	0.35	0.793	0.704	0.635	0.728
P16	0.564	0.582	0.387	0.654	0.569	0.617	0.536	0.598	0.128	0.147	0.11	0.15	0.576	0.66	0.722	0.528	0.209	0.24	0.191	0.234	0.589	0.631	0.527	0.632
P17	0.674	0.603	0.455	0.6	0.689	0.606	0.571	0.645	0.192	0.146	0.122	0.159	0.708	0.611	0.715	0.701	0.303	0.235	0.208	0.26	0.734	0.633	0.573	0.696
P18	0.745	0.809	0.599	0.625	0.71	0.675	0.617	0.677	0.231	0.264	0.149	0.175	0.667	0.507	0.639	0.743	0.343	0.347	0.242	0.284	0.77	0.73	0.651	0.733
P19	0.762	0.707	0.528	0.744	0.754	0.732	0.633	0.691	0.26	0.221	0.146	0.222	0.743	0.764	0.764	0.625	0.385	0.343	0.244	0.328	0.816	0.805	0.683	0.73
P20	0.812	0.681	0.587	0.747	0.779	0.742	0.607	0.715	0.314	0.214	0.144	0.234	0.736	0.819	0.632	0.674	0.44	0.339	0.234	0.348	0.836	0.801	0.632	0.774
Avg	**0.733**	**0.696**	**0.619**	**0.707**	**0.719**	**0.686**	**0.606**	**0.692**	**0.242**	**0.212**	**0.156**	**0.212**	**0.700**	**0.674**	**0.590**	**0.674**	**0.356**	**0.317**	**0.244**	**0.320**	**0.777**	**0.730**	**0.621**	**0.741**

Table B.7: EEGTransformer: Evaluation metrics for classification model trained on traditional (clean) RSVP data.

EEGTransformer : Train RSVP																								
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.517	0.679	0.683	0.662	0.571	0.587	0.534	0.6	0.125	0.15	0.121	0.152	0.639	0.472	0.347	0.521	0.209	0.227	0.18	0.236	0.589	0.589	0.527	0.633
P2	0.72	0.63	0.648	0.489	0.669	0.637	0.579	0.623	0.201	0.162	0.141	0.139	0.604	0.646	0.493	0.792	0.302	0.259	0.219	0.237	0.707	0.663	0.601	0.667
P3	0.592	0.682	0.426	0.726	0.574	0.617	0.53	0.57	0.132	0.165	0.109	0.151	0.552	0.535	0.66	0.375	0.213	0.252	0.187	0.215	0.587	0.626	0.515	0.549
P4	0.701	0.649	0.603	0.602	0.666	0.657	0.57	0.68	0.192	0.174	0.131	0.172	0.622	0.667	0.528	0.778	0.294	0.275	0.21	0.281	0.706	0.71	0.565	0.74
P5	0.761	0.719	0.642	0.729	0.728	0.656	0.601	0.683	0.248	0.195	0.149	0.211	0.688	0.576	0.549	0.625	0.365	0.291	0.235	0.316	0.78	0.702	0.589	0.731
P6	0.727	0.665	0.692	0.539	0.693	0.607	0.585	0.679	0.214	0.156	0.152	0.161	0.649	0.535	0.451	0.854	0.322	0.242	0.227	0.27	0.754	0.624	0.583	0.73
P7	0.746	0.648	0.466	0.626	0.667	0.641	0.573	0.641	0.212	0.167	0.123	0.162	0.569	0.632	0.708	0.66	0.309	0.264	0.21	0.261	0.716	0.656	0.587	0.67
P8	0.47	0.694	0.364	0.467	0.597	0.611	0.579	0.599	0.13	0.165	0.12	0.13	0.757	0.507	0.847	0.764	0.222	0.249	0.21	0.223	0.623	0.627	0.603	0.609
P9	0.762	0.747	0.681	0.669	0.725	0.665	0.582	0.681	0.248	0.212	0.148	0.188	0.677	0.562	0.458	0.694	0.363	0.308	0.223	0.296	0.794	0.711	0.596	0.726
P10	0.814	0.768	0.503	0.481	0.707	0.689	0.561	0.563	0.285	0.236	0.121	0.121	0.573	0.59	0.632	0.667	0.381	0.337	0.203	0.204	0.77	0.737	0.554	0.561
P11	0.778	0.839	0.581	0.826	0.75	0.769	0.637	0.749	0.27	0.345	0.154	0.319	0.715	0.681	0.708	0.653	0.392	0.458	0.252	0.428	0.823	0.817	0.661	0.819
P12	0.812	0.619	0.678	0.758	0.75	0.65	0.58	0.693	0.302	0.164	0.146	0.232	0.674	0.688	0.458	0.611	0.417	0.265	0.221	0.336	0.821	0.685	0.573	0.757
P13	0.754	0.56	0.49	0.645	0.632	0.611	0.519	0.655	0.198	0.142	0.107	0.172	0.479	0.674	0.556	0.667	0.28	0.235	0.179	0.273	0.662	0.63	0.511	0.684
P14	0.638	0.577	0.435	0.676	0.679	0.623	0.587	0.628	0.179	0.148	0.125	0.168	0.729	0.681	0.778	0.569	0.287	0.243	0.216	0.26	0.738	0.628	0.603	0.657
P15	0.675	0.645	0.467	0.478	0.708	0.584	0.574	0.618	0.2	0.142	0.123	0.137	0.75	0.507	0.708	0.792	0.316	0.222	0.21	0.233	0.744	0.603	0.573	0.649
P16	0.376	0.668	0.524	0.687	0.567	0.559	0.554	0.601	0.118	0.134	0.12	0.158	0.806	0.424	0.59	0.493	0.205	0.203	0.199	0.239	0.56	0.57	0.538	0.589
P17	0.715	0.605	0.365	0.631	0.628	0.601	0.542	0.539	0.18	0.144	0.111	0.12	0.521	0.597	0.764	0.424	0.267	0.232	0.194	0.187	0.669	0.601	0.542	0.536
P18	0.603	0.528	0.669	0.767	0.647	0.615	0.575	0.664	0.16	0.14	0.142	0.223	0.701	0.722	0.458	0.535	0.261	0.234	0.217	0.314	0.68	0.644	0.587	0.698
P19	0.676	0.755	0.632	0.801	0.678	0.703	0.617	0.587	0.189	0.234	0.154	0.197	0.681	0.639	0.597	0.319	0.296	0.343	0.245	0.243	0.729	0.767	0.632	0.598
P20	0.709	0.672	0.536	0.61	0.712	0.669	0.606	0.675	0.214	0.184	0.138	0.171	0.715	0.667	0.694	0.757	0.33	0.289	0.23	0.279	0.747	0.704	0.603	0.714
Avg	**0.677**	**0.667**	**0.554**	**0.643**	**0.667**	**0.637**	**0.574**	**0.636**	**0.199**	**0.177**	**0.131**	**0.174**	**0.655**	**0.600**	**0.599**	**0.627**	**0.301**	**0.271**	**0.213**	**0.266**	**0.710**	**0.665**	**0.577**	**0.666**

Table B.8: MoE-Transformer: Evaluation metrics for classification model trained on traditional (clean) RSVP data.

MoE-Transformer : Train RSVP																								
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.608	0.586	0.404	0.544	0.594	0.613	0.573	0.66	0.142	0.146	0.12	0.156	0.576	0.646	0.785	0.806	0.227	0.238	0.208	0.261	0.62	0.625	0.56	0.688
P2	0.733	0.668	0.776	0.664	0.723	0.711	0.672	0.684	0.23	0.199	0.233	0.188	0.712	0.764	0.542	0.708	0.347	0.315	0.326	0.297	0.779	0.765	0.718	0.743
P3	0.668	0.739	0.503	0.765	0.663	0.645	0.542	0.678	0.181	0.198	0.115	0.229	0.656	0.528	0.59	0.569	0.284	0.288	0.192	0.327	0.707	0.686	0.508	0.719
P4	0.678	0.749	0.562	0.759	0.713	0.743	0.571	0.743	0.203	0.247	0.128	0.253	0.757	0.736	0.583	0.722	0.32	0.37	0.21	0.375	0.783	0.811	0.567	0.798
P5	0.832	0.828	0.422	0.78	0.813	0.756	0.577	0.794	0.35	0.324	0.122	0.287	0.788	0.667	0.771	0.812	0.485	0.436	0.21	0.425	0.883	0.826	0.6	0.858
P6	0.814	0.617	0.747	0.697	0.738	0.694	0.588	0.702	0.3	0.179	0.168	0.205	0.642	0.792	0.389	0.708	0.409	0.292	0.235	0.318	0.815	0.732	0.596	0.749
P7	0.726	0.694	0.695	0.677	0.672	0.669	0.63	0.682	0.205	0.191	0.174	0.191	0.604	0.639	0.549	0.688	0.306	0.294	0.265	0.299	0.728	0.694	0.647	0.725
P8	0.601	0.694	0.672	0.729	0.64	0.617	0.645	0.633	0.158	0.168	0.174	0.188	0.688	0.521	0.611	0.514	0.256	0.254	0.271	0.275	0.68	0.644	0.666	0.666
P9	0.816	0.767	0.755	0.741	0.78	0.744	0.703	0.736	0.318	0.259	0.234	0.239	0.736	0.715	0.639	0.729	0.444	0.38	0.343	0.36	0.849	0.816	0.755	0.794
P10	0.838	0.788	0.726	0.489	0.753	0.734	0.573	0.565	0.338	0.272	0.153	0.121	0.646	0.667	0.382	0.66	0.443	0.386	0.218	0.205	0.823	0.782	0.569	0.571
P11	0.744	0.758	0.537	0.695	0.784	0.776	0.638	0.769	0.259	0.264	0.148	0.228	0.833	0.799	0.764	0.861	0.395	0.397	0.248	0.361	0.853	0.837	0.667	0.841
P12	0.784	0.769	0.578	0.726	0.787	0.708	0.602	0.734	0.289	0.246	0.141	0.231	0.792	0.632	0.632	0.743	0.423	0.354	0.23	0.352	0.838	0.763	0.582	0.805
P13	0.788	0.739	0.704	0.74	0.694	0.639	0.549	0.735	0.254	0.195	0.133	0.239	0.576	0.514	0.354	0.729	0.352	0.282	0.193	0.36	0.73	0.662	0.536	0.766
P14	0.76	0.721	0.736	0.801	0.719	0.7	0.659	0.76	0.244	0.215	0.204	0.294	0.667	0.674	0.562	0.708	0.358	0.326	0.299	0.415	0.775	0.747	0.669	0.797
P15	0.755	0.459	0.724	0.665	0.737	0.613	0.547	0.61	0.248	0.134	0.135	0.158	0.715	0.806	0.326	0.542	0.369	0.229	0.191	0.244	0.81	0.652	0.53	0.623
P16	0.674	0.65	0.669	0.664	0.541	0.549	0.526	0.56	0.125	0.127	0.116	0.134	0.375	0.424	0.347	0.431	0.187	0.195	0.174	0.204	0.522	0.54	0.493	0.564
P17	0.774	0.772	0.642	0.565	0.689	0.608	0.554	0.617	0.241	0.193	0.128	0.145	0.583	0.403	0.444	0.681	0.341	0.261	0.199	0.238	0.742	0.612	0.547	0.632
P18	0.762	0.669	0.764	0.718	0.652	0.677	0.576	0.671	0.214	0.187	0.167	0.201	0.514	0.688	0.34	0.611	0.302	0.294	0.224	0.302	0.691	0.713	0.582	0.713
P19	0.826	0.837	0.6	0.721	0.752	0.77	0.62	0.691	0.32	0.343	0.15	0.211	0.66	0.688	0.646	0.653	0.431	0.457	0.244	0.319	0.806	0.833	0.638	0.727
P20	0.79	0.756	0.692	0.674	0.791	0.707	0.601	0.717	0.295	0.237	0.159	0.203	0.792	0.646	0.486	0.771	0.429	0.346	0.24	0.321	0.851	0.768	0.627	0.749
Avg	**0.748**	**0.713**	**0.645**	**0.690**	**0.711**	**0.683**	**0.597**	**0.687**	**0.245**	**0.216**	**0.155**	**0.205**	**0.665**	**0.647**	**0.537**	**0.682**	**0.355**	**0.319**	**0.236**	**0.312**	**0.764**	**0.725**	**0.603**	**0.726**

Table B.9: EEG Conformer: Evaluation metrics for classification model trained on traditional (clean) RSVP data.

EEG Conformer : Train RSVP																								
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.587	0.573	0.549	0.629	0.595	0.581	0.57	0.633	0.139	0.133	0.127	0.16	0.604	0.59	0.597	0.639	0.226	0.217	0.209	0.256	0.613	0.593	0.568	0.654
P2	0.751	0.713	0.76	0.828	0.783	0.751	0.66	0.738	0.262	0.23	0.217	0.317	0.823	0.799	0.535	0.625	0.398	0.358	0.309	0.421	0.859	0.828	0.696	0.815
P3	0.64	0.619	0.542	0.795	0.598	0.646	0.542	0.639	0.148	0.163	0.116	0.229	0.545	0.681	0.542	0.444	0.232	0.263	0.191	0.303	0.635	0.678	0.513	0.681
P4	0.796	0.815	0.549	0.743	0.773	0.761	0.561	0.74	0.294	0.31	0.124	0.242	0.743	0.694	0.576	0.736	0.422	0.428	0.203	0.364	0.835	0.828	0.582	0.811
P5	0.813	0.802	0.537	0.81	0.791	0.745	0.542	0.756	0.318	0.29	0.116	0.303	0.764	0.674	0.549	0.688	0.449	0.405	0.192	0.42	0.857	0.798	0.52	0.806
P6	0.725	0.624	0.667	0.744	0.747	0.628	0.547	0.713	0.234	0.157	0.127	0.232	0.774	0.632	0.396	0.674	0.36	0.252	0.192	0.345	0.817	0.634	0.554	0.77
P7	0.681	0.715	0.653	0.696	0.702	0.669	0.585	0.704	0.2	0.199	0.144	0.206	0.729	0.611	0.5	0.715	0.314	0.3	0.224	0.32	0.749	0.691	0.574	0.768
P8	0.682	0.506	0.767	0.593	0.624	0.584	0.602	0.657	0.168	0.128	0.186	0.162	0.552	0.681	0.396	0.736	0.257	0.216	0.253	0.266	0.645	0.592	0.63	0.678
P9	0.675	0.719	0.778	0.676	0.634	0.693	0.642	0.688	0.171	0.211	0.219	0.193	0.583	0.66	0.472	0.701	0.264	0.32	0.299	0.302	0.674	0.734	0.664	0.751
P10	0.822	0.785	0.504	0.512	0.79	0.711	0.579	0.553	0.328	0.259	0.127	0.119	0.75	0.618	0.674	0.604	0.457	0.366	0.214	0.199	0.855	0.75	0.584	0.563
P11	0.867	0.781	0.806	0.846	0.775	0.77	0.58	0.788	0.399	0.279	0.194	0.363	0.66	0.757	0.299	0.715	0.497	0.408	0.235	0.481	0.843	0.835	0.562	0.86
P12	0.832	0.801	0.592	0.712	0.826	0.698	0.591	0.726	0.353	0.268	0.138	0.221	0.819	0.569	0.59	0.743	0.494	0.364	0.224	0.34	0.889	0.756	0.592	0.767
P13	0.605	0.623	0.613	0.769	0.682	0.602	0.541	0.699	0.173	0.147	0.12	0.242	0.778	0.576	0.451	0.611	0.282	0.234	0.189	0.346	0.72	0.628	0.53	0.759
P14	0.765	0.735	0.713	0.663	0.703	0.696	0.634	0.699	0.24	0.22	0.182	0.193	0.625	0.646	0.535	0.743	0.347	0.328	0.272	0.306	0.751	0.741	0.639	0.739
P15	0.66	0.79	0.506	0.557	0.722	0.655	0.596	0.661	0.2	0.235	0.132	0.158	0.799	0.486	0.708	0.792	0.32	0.317	0.223	0.263	0.783	0.676	0.591	0.714
P16	0.383	0.701	0.678	0.417	0.568	0.572	0.525	0.59	0.118	0.146	0.115	0.125	0.799	0.41	0.333	0.806	0.206	0.215	0.171	0.216	0.576	0.58	0.497	0.609
P17	0.576	0.622	0.632	0.499	0.65	0.62	0.561	0.61	0.157	0.154	0.13	0.136	0.743	0.618	0.472	0.75	0.26	0.246	0.204	0.23	0.7	0.646	0.55	0.64
P18	0.629	0.727	0.635	0.652	0.695	0.685	0.603	0.662	0.182	0.211	0.149	0.176	0.778	0.632	0.562	0.674	0.296	0.317	0.236	0.279	0.757	0.738	0.617	0.708
P19	0.783	0.781	0.528	0.765	0.778	0.749	0.642	0.718	0.285	0.272	0.148	0.247	0.771	0.708	0.785	0.66	0.416	0.393	0.25	0.359	0.846	0.819	0.685	0.77
P20	0.781	0.758	0.451	0.708	0.798	0.742	0.624	0.726	0.289	0.252	0.136	0.219	0.819	0.722	0.84	0.75	0.428	0.373	0.234	0.339	0.868	0.802	0.651	0.789
Avg	**0.702**	**0.709**	**0.623**	**0.680**	**0.711**	**0.677**	**0.586**	**0.685**	**0.232**	**0.213**	**0.147**	**0.212**	**0.722**	**0.638**	**0.540**	**0.690**	**0.346**	**0.316**	**0.226**	**0.317**	**0.764**	**0.717**	**0.590**	**0.733**

Table B.10: CNN1-Transformer: Evaluation metrics for classification model trained on traditional (clean) RSVP data.

CNN1-Transformer : Train RSVP																								
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.604	0.715	0.561	0.592	0.592	0.598	0.552	0.644	0.14	0.164	0.121	0.157	0.576	0.451	0.542	0.708	0.226	0.24	0.198	0.258	0.62	0.617	0.551	0.683
P2	0.695	0.668	0.74	0.703	0.681	0.674	0.658	0.687	0.196	0.185	0.205	0.202	0.663	0.681	0.556	0.667	0.303	0.291	0.299	0.31	0.738	0.729	0.701	0.738
P3	0.657	0.617	0.469	0.787	0.652	0.599	0.507	0.616	0.173	0.145	0.102	0.208	0.646	0.576	0.556	0.403	0.273	0.232	0.173	0.274	0.692	0.638	0.486	0.652
P4	0.757	0.785	0.556	0.749	0.74	0.726	0.587	0.7	0.251	0.266	0.133	0.229	0.719	0.653	0.625	0.639	0.372	0.378	0.22	0.338	0.811	0.794	0.591	0.773
P5	0.796	0.733	0.751	0.71	0.794	0.734	0.541	0.777	0.302	0.234	0.136	0.238	0.792	0.736	0.278	0.861	0.437	0.355	0.183	0.372	0.858	0.796	0.53	0.84
P6	0.725	0.795	0.677	0.777	0.756	0.67	0.586	0.725	0.238	0.247	0.149	0.259	0.795	0.514	0.472	0.66	0.366	0.334	0.226	0.372	0.832	0.714	0.602	0.769
P7	0.769	0.733	0.561	0.642	0.71	0.666	0.62	0.659	0.247	0.205	0.145	0.173	0.635	0.583	0.694	0.681	0.355	0.304	0.24	0.276	0.763	0.708	0.629	0.702
P8	0.495	0.556	0.756	0.66	0.633	0.617	0.639	0.644	0.142	0.144	0.203	0.171	0.806	0.694	0.493	0.625	0.242	0.238	0.287	0.269	0.671	0.616	0.686	0.695
P9	0.793	0.73	0.712	0.701	0.765	0.767	0.679	0.738	0.289	0.244	0.202	0.221	0.729	0.812	0.639	0.785	0.414	0.376	0.307	0.345	0.833	0.837	0.718	0.817
P10	0.793	0.686	0.54	0.556	0.765	0.721	0.584	0.596	0.289	0.208	0.131	0.136	0.729	0.764	0.639	0.646	0.414	0.327	0.217	0.225	0.844	0.779	0.599	0.61
P11	0.833	0.78	0.694	0.795	0.778	0.797	0.648	0.791	0.34	0.289	0.182	0.3	0.708	0.819	0.59	0.785	0.459	0.427	0.279	0.434	0.858	0.871	0.686	0.87
P12	0.774	0.72	0.492	0.748	0.766	0.712	0.582	0.721	0.272	0.219	0.127	0.237	0.757	0.701	0.694	0.688	0.401	0.334	0.215	0.353	0.829	0.786	0.575	0.792
P13	0.693	0.697	0.729	0.742	0.657	0.619	0.529	0.669	0.186	0.17	0.123	0.211	0.611	0.521	0.278	0.576	0.285	0.256	0.17	0.309	0.706	0.632	0.515	0.724
P14	0.831	0.7	0.647	0.809	0.78	0.713	0.668	0.736	0.338	0.211	0.177	0.293	0.715	0.729	0.694	0.646	0.459	0.327	0.282	0.403	0.846	0.756	0.72	0.773
P15	0.822	0.746	0.522	0.717	0.759	0.649	0.617	0.71	0.317	0.203	0.14	0.217	0.681	0.528	0.736	0.701	0.433	0.293	0.235	0.331	0.829	0.7	0.627	0.739
P16	0.43	0.634	0.558	0.559	0.554	0.587	0.532	0.591	0.116	0.142	0.113	0.135	0.708	0.528	0.5	0.632	0.199	0.224	0.184	0.223	0.549	0.611	0.512	0.611
P17	0.669	0.596	0.343	0.562	0.668	0.603	0.552	0.581	0.183	0.143	0.113	0.132	0.667	0.611	0.812	0.604	0.287	0.232	0.198	0.216	0.716	0.621	0.531	0.593
P18	0.713	0.674	0.65	0.723	0.696	0.668	0.623	0.667	0.21	0.184	0.16	0.201	0.674	0.66	0.59	0.597	0.32	0.288	0.252	0.301	0.741	0.709	0.639	0.726
P19	0.76	0.797	0.624	0.657	0.75	0.739	0.668	0.671	0.257	0.282	0.172	0.181	0.736	0.667	0.722	0.688	0.381	0.397	0.278	0.286	0.825	0.808	0.694	0.73
P20	0.667	0.602	0.505	0.747	0.71	0.705	0.592	0.686	0.198	0.179	0.131	0.222	0.764	0.833	0.701	0.611	0.314	0.295	0.221	0.325	0.773	0.771	0.607	0.757
Avg	**0.713**	**0.698**	**0.604**	**0.696**	**0.710**	**0.678**	**0.598**	**0.680**	**0.234**	**0.203**	**0.148**	**0.206**	**0.705**	**0.653**	**0.590**	**0.660**	**0.347**	**0.307**	**0.233**	**0.311**	**0.767**	**0.725**	**0.610**	**0.730**

Table B.11: EEGNet-Transformer: Evaluation metrics for classification model trained on traditional (clean) RSVP data.

EEGNet-Transformer : Train RSVP																								
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.684	0.742	0.644	0.765	0.61	0.588	0.568	0.656	0.162	0.167	0.135	0.217	0.517	0.396	0.472	0.521	0.246	0.235	0.21	0.307	0.634	0.593	0.563	0.688
P2	0.704	0.582	0.817	0.635	0.672	0.684	0.633	0.692	0.196	0.169	0.246	0.183	0.632	0.812	0.403	0.764	0.299	0.28	0.305	0.295	0.731	0.748	0.679	0.733
P3	0.687	0.556	0.447	0.674	0.666	0.633	0.545	0.637	0.187	0.149	0.114	0.172	0.639	0.729	0.667	0.59	0.29	0.247	0.194	0.266	0.705	0.667	0.528	0.673
P4	0.75	0.791	0.578	0.776	0.745	0.757	0.578	0.733	0.248	0.284	0.132	0.261	0.74	0.715	0.576	0.681	0.371	0.406	0.215	0.378	0.811	0.819	0.572	0.792
P5	0.851	0.714	0.647	0.797	0.778	0.761	0.578	0.788	0.368	0.234	0.14	0.3	0.688	0.819	0.493	0.778	0.479	0.364	0.218	0.433	0.852	0.826	0.58	0.853
P6	0.702	0.749	0.679	0.741	0.73	0.657	0.624	0.708	0.218	0.209	0.167	0.228	0.764	0.542	0.556	0.667	0.339	0.301	0.257	0.34	0.807	0.714	0.624	0.765
P7	0.753	0.768	0.602	0.742	0.707	0.689	0.637	0.684	0.234	0.236	0.157	0.218	0.649	0.59	0.681	0.611	0.344	0.337	0.255	0.322	0.758	0.715	0.658	0.728
P8	0.677	0.615	0.541	0.596	0.651	0.613	0.643	0.658	0.178	0.15	0.15	0.163	0.618	0.611	0.771	0.736	0.277	0.241	0.251	0.267	0.693	0.628	0.69	0.706
P9	0.81	0.76	0.581	0.676	0.789	0.709	0.635	0.715	0.314	0.24	0.153	0.203	0.764	0.646	0.701	0.764	0.445	0.35	0.251	0.32	0.863	0.781	0.687	0.774
P10	0.837	0.739	0.644	0.562	0.785	0.741	0.595	0.572	0.348	0.24	0.147	0.128	0.719	0.743	0.535	0.583	0.469	0.363	0.231	0.211	0.861	0.796	0.595	0.586
P11	0.842	0.836	0.664	0.803	0.77	0.798	0.634	0.811	0.35	0.351	0.168	0.315	0.681	0.75	0.597	0.819	0.462	0.478	0.262	0.455	0.852	0.863	0.665	0.882
P12	0.824	0.739	0.654	0.747	0.772	0.731	0.586	0.739	0.325	0.236	0.145	0.244	0.708	0.722	0.5	0.729	0.445	0.356	0.224	0.365	0.837	0.795	0.576	0.805
P13	0.741	0.407	0.688	0.628	0.662	0.593	0.53	0.667	0.207	0.126	0.119	0.173	0.562	0.826	0.333	0.715	0.303	0.218	0.176	0.278	0.706	0.621	0.527	0.729
P14	0.854	0.757	0.617	0.799	0.743	0.714	0.642	0.728	0.362	0.24	0.161	0.28	0.604	0.66	0.674	0.639	0.453	0.352	0.26	0.389	0.816	0.768	0.696	0.765
P15	0.728	0.535	0.695	0.688	0.729	0.659	0.587	0.663	0.23	0.154	0.153	0.187	0.729	0.812	0.451	0.632	0.349	0.259	0.228	0.288	0.787	0.693	0.61	0.7
P16	0.559	0.585	0.733	0.641	0.551	0.606	0.54	0.597	0.121	0.143	0.132	0.147	0.542	0.632	0.299	0.542	0.197	0.233	0.183	0.232	0.544	0.621	0.527	0.615
P17	0.601	0.576	0.378	0.438	0.667	0.622	0.544	0.586	0.167	0.148	0.112	0.125	0.75	0.681	0.75	0.771	0.273	0.243	0.194	0.215	0.704	0.646	0.544	0.606
P18	0.676	0.652	0.673	0.729	0.672	0.668	0.642	0.661	0.187	0.178	0.174	0.201	0.667	0.688	0.604	0.576	0.292	0.283	0.27	0.299	0.749	0.708	0.648	0.707
P19	0.807	0.797	0.615	0.696	0.794	0.77	0.662	0.708	0.313	0.294	0.168	0.207	0.778	0.736	0.722	0.722	0.446	0.421	0.273	0.322	0.847	0.849	0.704	0.753
P20	0.824	0.717	0.617	0.738	0.72	0.723	0.583	0.697	0.304	0.222	0.138	0.222	0.59	0.729	0.542	0.646	0.401	0.34	0.22	0.33	0.795	0.781	0.604	0.749
Avg	**0.745**	**0.680**	**0.625**	**0.693**	**0.710**	**0.685**	**0.599**	**0.685**	**0.250**	**0.208**	**0.150**	**0.208**	**0.667**	**0.691**	**0.566**	**0.674**	**0.359**	**0.315**	**0.233**	**0.315**	**0.768**	**0.732**	**0.614**	**0.730**

Table B.12: Multiband EEGNet-Transformer: Evaluation metrics for classification model trained on traditional (clean) RSVP data.

Multiband EEGNet-Transformer : Train RSVP																								
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.665	0.638	0.544	0.658	0.627	0.598	0.586	0.646	0.165	0.148	0.132	0.171	0.58	0.549	0.639	0.632	0.257	0.233	0.219	0.27	0.662	0.626	0.586	0.689
P2	0.679	0.678	0.681	0.665	0.671	0.691	0.665	0.706	0.187	0.195	0.185	0.196	0.66	0.708	0.646	0.757	0.291	0.305	0.288	0.311	0.719	0.745	0.723	0.763
P3	0.697	0.553	0.47	0.609	0.679	0.635	0.53	0.653	0.196	0.149	0.11	0.164	0.656	0.736	0.604	0.708	0.302	0.248	0.186	0.266	0.735	0.668	0.506	0.69
P4	0.748	0.813	0.759	0.733	0.768	0.754	0.561	0.762	0.255	0.305	0.154	0.245	0.792	0.681	0.312	0.799	0.386	0.422	0.206	0.375	0.836	0.827	0.577	0.827
P5	0.824	0.74	0.731	0.782	0.779	0.753	0.603	0.78	0.327	0.245	0.172	0.284	0.722	0.771	0.444	0.778	0.45	0.372	0.248	0.416	0.85	0.811	0.632	0.844
P6	0.815	0.74	0.617	0.695	0.757	0.664	0.633	0.707	0.308	0.208	0.158	0.207	0.684	0.569	0.653	0.722	0.425	0.305	0.254	0.321	0.831	0.718	0.652	0.744
P7	0.695	0.781	0.717	0.717	0.71	0.696	0.624	0.694	0.208	0.249	0.178	0.211	0.729	0.59	0.507	0.667	0.323	0.35	0.264	0.32	0.768	0.729	0.652	0.747
P8	0.669	0.613	0.611	0.618	0.646	0.637	0.636	0.655	0.174	0.159	0.158	0.166	0.618	0.667	0.667	0.701	0.272	0.256	0.255	0.269	0.685	0.668	0.667	0.679
P9	0.789	0.792	0.706	0.757	0.793	0.767	0.682	0.704	0.294	0.289	0.201	0.236	0.799	0.736	0.653	0.639	0.43	0.415	0.307	0.345	0.867	0.823	0.728	0.778
P10	0.792	0.802	0.767	0.653	0.773	0.73	0.584	0.659	0.291	0.283	0.174	0.175	0.75	0.639	0.354	0.667	0.419	0.392	0.233	0.277	0.854	0.782	0.6	0.683
P11	0.878	0.785	0.744	0.806	0.784	0.779	0.639	0.787	0.43	0.287	0.197	0.31	0.667	0.771	0.507	0.764	0.523	0.418	0.284	0.441	0.845	0.838	0.66	0.864
P12	0.858	0.76	0.644	0.758	0.782	0.709	0.599	0.742	0.382	0.24	0.149	0.252	0.688	0.646	0.542	0.722	0.491	0.35	0.234	0.373	0.863	0.779	0.614	0.793
P13	0.707	0.71	0.59	0.681	0.677	0.632	0.569	0.718	0.199	0.18	0.13	0.205	0.639	0.535	0.542	0.764	0.304	0.269	0.209	0.324	0.737	0.677	0.568	0.782
P14	0.746	0.729	0.637	0.798	0.751	0.692	0.675	0.712	0.248	0.215	0.177	0.271	0.757	0.646	0.722	0.604	0.373	0.323	0.285	0.374	0.827	0.759	0.72	0.756
P15	0.763	0.779	0.704	0.677	0.73	0.649	0.62	0.676	0.251	0.223	0.172	0.188	0.688	0.486	0.514	0.674	0.367	0.306	0.258	0.294	0.798	0.698	0.623	0.71
P16	0.299	0.642	0.647	0.442	0.555	0.603	0.529	0.591	0.113	0.15	0.116	0.127	0.875	0.556	0.382	0.778	0.2	0.237	0.178	0.218	0.552	0.614	0.497	0.615
P17	0.699	0.571	0.34	0.588	0.703	0.626	0.568	0.641	0.207	0.148	0.117	0.156	0.708	0.694	0.854	0.708	0.32	0.244	0.206	0.256	0.735	0.653	0.583	0.674
P18	0.823	0.617	0.626	0.646	0.689	0.67	0.607	0.686	0.287	0.171	0.149	0.183	0.521	0.736	0.583	0.736	0.37	0.277	0.238	0.294	0.735	0.709	0.637	0.747
P19	0.763	0.749	0.68	0.665	0.767	0.743	0.646	0.709	0.265	0.247	0.177	0.197	0.771	0.736	0.604	0.764	0.394	0.37	0.274	0.313	0.827	0.802	0.677	0.759
P20	0.811	0.672	0.54	0.753	0.756	0.709	0.602	0.733	0.304	0.199	0.137	0.245	0.688	0.757	0.681	0.708	0.421	0.315	0.228	0.364	0.808	0.762	0.606	0.78
Avg	**0.736**	**0.708**	**0.637**	**0.685**	**0.719**	**0.686**	**0.607**	**0.698**	**0.254**	**0.214**	**0.157**	**0.209**	**0.699**	**0.660**	**0.570**	**0.714**	**0.365**	**0.320**	**0.242**	**0.321**	**0.777**	**0.734**	**0.625**	**0.746**

Table B.13: EEGNet-MoE-Transformer: Evaluation metrics for classification model trained on traditional (clean) RSVP data.

EEGNet-MoE-Transformer : Train RSVP																								
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.652	0.539	0.598	0.685	0.612	0.608	0.545	0.661	0.156	0.139	0.12	0.185	0.562	0.694	0.479	0.632	0.244	0.231	0.192	0.286	0.629	0.641	0.55	0.718
P2	0.689	0.576	0.796	0.617	0.666	0.666	0.637	0.667	0.188	0.162	0.228	0.17	0.639	0.778	0.438	0.729	0.291	0.268	0.3	0.276	0.706	0.724	0.67	0.712
P3	0.675	0.753	0.362	0.692	0.677	0.653	0.556	0.674	0.188	0.209	0.115	0.193	0.681	0.528	0.799	0.653	0.295	0.299	0.2	0.297	0.731	0.688	0.532	0.704
P4	0.701	0.778	0.456	0.726	0.721	0.741	0.571	0.724	0.214	0.266	0.122	0.227	0.747	0.694	0.715	0.722	0.333	0.385	0.208	0.345	0.783	0.806	0.571	0.783
P5	0.768	0.765	0.377	0.799	0.78	0.752	0.543	0.793	0.273	0.26	0.111	0.305	0.795	0.736	0.75	0.785	0.407	0.385	0.194	0.439	0.857	0.821	0.537	0.866
P6	0.734	0.693	0.606	0.71	0.733	0.666	0.568	0.731	0.234	0.19	0.131	0.222	0.733	0.632	0.521	0.757	0.355	0.292	0.209	0.343	0.81	0.714	0.582	0.79
P7	0.648	0.669	0.7	0.588	0.672	0.681	0.673	0.669	0.179	0.188	0.195	0.165	0.701	0.694	0.639	0.771	0.285	0.296	0.299	0.272	0.722	0.705	0.681	0.712
P8	0.633	0.572	0.617	0.674	0.645	0.59	0.652	0.665	0.165	0.136	0.165	0.183	0.66	0.611	0.694	0.653	0.264	0.222	0.266	0.286	0.684	0.588	0.685	0.701
P9	0.785	0.726	0.815	0.701	0.78	0.749	0.653	0.711	0.287	0.236	0.257	0.211	0.774	0.778	0.451	0.722	0.418	0.362	0.327	0.326	0.842	0.817	0.703	0.777
P10	0.705	0.777	0.608	0.472	0.719	0.743	0.591	0.564	0.215	0.266	0.14	0.121	0.736	0.701	0.569	0.681	0.333	0.386	0.225	0.205	0.799	0.785	0.587	0.579
P11	0.774	0.815	0.806	0.792	0.776	0.805	0.657	0.81	0.277	0.326	0.25	0.303	0.778	0.792	0.472	0.833	0.408	0.462	0.327	0.444	0.842	0.865	0.706	0.869
P12	0.799	0.764	0.622	0.669	0.759	0.721	0.583	0.726	0.292	0.247	0.139	0.204	0.708	0.667	0.535	0.799	0.414	0.361	0.22	0.325	0.82	0.782	0.589	0.771
P13	0.753	0.584	0.278	0.823	0.653	0.615	0.537	0.692	0.209	0.146	0.108	0.289	0.528	0.653	0.861	0.528	0.299	0.239	0.193	0.373	0.683	0.637	0.525	0.743
P14	0.832	0.716	0.719	0.702	0.762	0.697	0.665	0.714	0.332	0.212	0.199	0.212	0.674	0.674	0.597	0.729	0.445	0.322	0.299	0.329	0.836	0.759	0.711	0.757
P15	0.758	0.519	0.428	0.481	0.714	0.616	0.559	0.625	0.241	0.139	0.117	0.139	0.66	0.736	0.722	0.806	0.353	0.235	0.202	0.237	0.763	0.639	0.563	0.651
P16	0.704	0.566	0.652	0.63	0.546	0.598	0.56	0.6	0.131	0.138	0.132	0.147	0.347	0.639	0.444	0.562	0.19	0.227	0.203	0.233	0.531	0.606	0.561	0.599
P17	0.685	0.671	0.778	0.388	0.621	0.61	0.525	0.589	0.167	0.159	0.128	0.124	0.542	0.535	0.208	0.84	0.256	0.245	0.158	0.215	0.656	0.628	0.509	0.595
P18	0.61	0.703	0.717	0.682	0.666	0.675	0.63	0.691	0.168	0.197	0.181	0.196	0.736	0.639	0.521	0.701	0.274	0.301	0.269	0.306	0.715	0.697	0.646	0.739
P19	0.778	0.699	0.692	0.615	0.775	0.743	0.662	0.688	0.28	0.222	0.188	0.177	0.771	0.799	0.625	0.778	0.41	0.347	0.288	0.288	0.823	0.826	0.703	0.733
P20	0.726	0.676	0.739	0.66	0.774	0.697	0.577	0.679	0.244	0.196	0.159	0.185	0.833	0.722	0.375	0.701	0.378	0.309	0.223	0.292	0.837	0.752	0.594	0.725
Avg	**0.720**	**0.678**	**0.618**	**0.655**	**0.702**	**0.681**	**0.597**	**0.683**	**0.222**	**0.201**	**0.159**	**0.197**	**0.680**	**0.685**	**0.570**	**0.719**	**0.332**	**0.308**	**0.240**	**0.305**	**0.753**	**0.724**	**0.610**	**0.726**

Table B.14: Multiband EEGNet-MoE-Transformer: Evaluation metrics for classification model trained on traditional (clean) RSVP data.

Multiband EEGNet-MoE-Transformer : Train RSVP																								
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.615	0.667	0.448	0.651	0.633	0.596	0.579	0.652	0.158	0.152	0.124	0.172	0.656	0.507	0.743	0.653	0.254	0.234	0.212	0.272	0.665	0.623	0.586	0.698
P2	0.772	0.671	0.796	0.688	0.687	0.684	0.689	0.697	0.238	0.19	0.258	0.2	0.58	0.701	0.556	0.708	0.337	0.299	0.352	0.312	0.734	0.729	0.717	0.729
P3	0.671	0.701	0.124	0.618	0.708	0.661	0.513	0.652	0.199	0.19	0.102	0.165	0.753	0.611	0.998	0.694	0.314	0.29	0.186	0.267	0.755	0.705	0.486	0.693
P4	0.737	0.781	0.527	0.697	0.74	0.736	0.558	0.718	0.238	0.266	0.121	0.211	0.743	0.681	0.597	0.743	0.361	0.383	0.202	0.329	0.802	0.777	0.575	0.774
P5	0.736	0.717	0.448	0.645	0.759	0.719	0.554	0.726	0.245	0.22	0.117	0.197	0.788	0.722	0.688	0.826	0.374	0.338	0.199	0.318	0.818	0.758	0.548	0.801
P6	0.757	0.618	0.624	0.722	0.726	0.646	0.59	0.707	0.245	0.163	0.142	0.218	0.688	0.681	0.549	0.688	0.362	0.263	0.226	0.331	0.799	0.683	0.605	0.749
P7	0.744	0.781	0.621	0.742	0.711	0.684	0.629	0.721	0.231	0.243	0.157	0.234	0.67	0.562	0.639	0.694	0.344	0.34	0.252	0.35	0.771	0.724	0.645	0.776
P8	0.63	0.549	0.754	0.712	0.64	0.601	0.617	0.63	0.163	0.138	0.189	0.18	0.653	0.667	0.444	0.528	0.261	0.228	0.266	0.268	0.675	0.615	0.649	0.664
P9	0.816	0.784	0.789	0.738	0.804	0.766	0.67	0.728	0.326	0.281	0.242	0.235	0.788	0.743	0.521	0.715	0.461	0.408	0.33	0.353	0.886	0.83	0.694	0.79
P10	0.843	0.692	0.693	0.471	0.796	0.739	0.595	0.592	0.361	0.217	0.157	0.129	0.736	0.799	0.472	0.743	0.485	0.341	0.235	0.219	0.872	0.795	0.599	0.609
P11	0.835	0.765	0.565	0.799	0.766	0.78	0.65	0.805	0.338	0.271	0.155	0.309	0.681	0.799	0.757	0.812	0.452	0.404	0.258	0.447	0.83	0.836	0.687	0.867
P12	0.824	0.719	0.644	0.708	0.791	0.742	0.596	0.733	0.332	0.23	0.148	0.222	0.75	0.771	0.535	0.764	0.461	0.354	0.231	0.344	0.848	0.8	0.601	0.792
P13	0.797	0.669	0.786	0.684	0.659	0.615	0.542	0.692	0.243	0.161	0.147	0.197	0.486	0.549	0.236	0.701	0.324	0.249	0.181	0.307	0.696	0.642	0.528	0.739
P14	0.74	0.692	0.667	0.73	0.775	0.69	0.664	0.714	0.253	0.199	0.181	0.225	0.819	0.688	0.66	0.694	0.387	0.309	0.284	0.34	0.832	0.756	0.691	0.761
P15	0.822	0.592	0.584	0.778	0.728	0.647	0.611	0.676	0.306	0.159	0.145	0.237	0.611	0.715	0.646	0.549	0.407	0.26	0.237	0.331	0.779	0.687	0.629	0.722
P16	0.46	0.611	0.285	0.497	0.579	0.63	0.554	0.597	0.124	0.156	0.112	0.132	0.729	0.653	0.889	0.722	0.213	0.251	0.199	0.223	0.581	0.629	0.562	0.605
P17	0.675	0.71	0.346	0.619	0.693	0.601	0.581	0.619	0.194	0.164	0.12	0.153	0.715	0.465	0.875	0.618	0.306	0.243	0.211	0.245	0.74	0.627	0.578	0.642
P18	0.818	0.719	0.605	0.599	0.689	0.684	0.62	0.675	0.281	0.207	0.151	0.169	0.528	0.639	0.639	0.771	0.367	0.313	0.244	0.278	0.75	0.739	0.646	0.729
P19	0.799	0.881	0.656	0.697	0.771	0.74	0.661	0.696	0.297	0.429	0.177	0.203	0.736	0.562	0.667	0.694	0.423	0.486	0.279	0.314	0.84	0.813	0.7	0.755
P20	0.799	0.722	0.705	0.717	0.743	0.704	0.586	0.713	0.285	0.217	0.155	0.218	0.674	0.681	0.438	0.708	0.401	0.329	0.229	0.333	0.805	0.757	0.609	0.748
Avg	**0.744**	**0.702**	**0.583**	**0.675**	**0.719**	**0.683**	**0.602**	**0.687**	**0.252**	**0.212**	**0.155**	**0.200**	**0.689**	**0.659**	**0.627**	**0.701**	**0.364**	**0.316**	**0.240**	**0.309**	**0.774**	**0.726**	**0.617**	**0.732**

Appendix C

Appendix C: Subject Independent Classification - Noisy Data as Part of Training

Table C.1: Bayesian Ridge Regression: Evaluation metrics for classification model trained on a combination of traditional (clean) and intentionally contaminated RSVP data.

Bayesian Ridge Regression : Train Traditional RSVP + Intentionally Contaminated RSVP																								
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.573	0.705	0.519	0.735	0.59	0.605	0.557	0.606	0.136	0.165	0.12	0.175	0.611	0.479	0.604	0.444	0.222	0.245	0.201	0.251	0.622	0.623	0.556	0.65
P2	0.601	0.65	0.593	0.598	0.624	0.657	0.647	0.666	0.152	0.174	0.159	0.166	0.653	0.667	0.715	0.75	0.247	0.276	0.26	0.272	0.674	0.695	0.68	0.72
P3	0.773	0.726	0.432	0.687	0.63	0.654	0.579	0.666	0.207	0.197	0.123	0.187	0.451	0.562	0.764	0.639	0.284	0.291	0.212	0.29	0.665	0.707	0.576	0.702
P4	0.711	0.747	0.525	0.66	0.713	0.686	0.597	0.712	0.215	0.222	0.134	0.196	0.715	0.611	0.688	0.778	0.331	0.325	0.224	0.314	0.767	0.759	0.602	0.77
P5	0.751	0.747	0.522	0.789	0.771	0.739	0.596	0.759	0.258	0.244	0.133	0.283	0.795	0.729	0.688	0.722	0.39	0.366	0.223	0.406	0.841	0.801	0.613	0.828
P6	0.659	0.734	0.528	0.643	0.732	0.667	0.583	0.672	0.203	0.206	0.13	0.178	0.823	0.583	0.653	0.708	0.325	0.305	0.217	0.284	0.797	0.713	0.596	0.704
P7	0.807	0.77	0.735	0.641	0.688	0.696	0.636	0.659	0.269	0.241	0.192	0.172	0.538	0.604	0.514	0.681	0.358	0.345	0.279	0.275	0.743	0.728	0.63	0.693
P8	0.634	0.509	0.666	0.715	0.647	0.619	0.642	0.641	0.167	0.14	0.172	0.186	0.663	0.757	0.611	0.549	0.266	0.236	0.268	0.278	0.69	0.651	0.67	0.667
P9	0.724	0.826	0.624	0.581	0.689	0.727	0.671	0.687	0.212	0.31	0.173	0.17	0.646	0.604	0.729	0.819	0.319	0.409	0.28	0.281	0.757	0.764	0.728	0.74
P10	0.783	0.785	0.647	0.572	0.703	0.662	0.631	0.645	0.254	0.235	0.163	0.155	0.604	0.507	0.611	0.736	0.357	0.321	0.257	0.256	0.768	0.708	0.662	0.67
P11	0.835	0.844	0.716	0.791	0.748	0.772	0.672	0.785	0.332	0.355	0.201	0.294	0.639	0.681	0.618	0.778	0.437	0.467	0.303	0.427	0.814	0.83	0.69	0.854
P12	0.828	0.808	0.747	0.735	0.781	0.726	0.606	0.739	0.333	0.288	0.18	0.237	0.722	0.625	0.431	0.743	0.456	0.394	0.254	0.36	0.836	0.784	0.609	0.776
P13	0.746	0.71	0.407	0.707	0.674	0.62	0.556	0.695	0.215	0.174	0.116	0.207	0.583	0.507	0.743	0.681	0.315	0.259	0.2	0.317	0.705	0.634	0.553	0.742
P14	0.749	0.783	0.749	0.773	0.709	0.701	0.651	0.744	0.233	0.253	0.206	0.264	0.66	0.597	0.528	0.708	0.344	0.355	0.296	0.384	0.764	0.743	0.683	0.79
P15	0.792	0.701	0.562	0.658	0.74	0.63	0.6	0.668	0.278	0.176	0.138	0.18	0.674	0.542	0.646	0.681	0.394	0.266	0.228	0.284	0.795	0.657	0.629	0.719
P16	0.633	0.725	0.16	0.489	0.537	0.594	0.518	0.586	0.119	0.165	0.103	0.128	0.417	0.431	0.965	0.708	0.185	0.238	0.187	0.217	0.495	0.595	0.469	0.587
P17	0.601	0.603	0.719	0.699	0.652	0.634	0.551	0.601	0.162	0.156	0.137	0.162	0.715	0.674	0.34	0.479	0.264	0.253	0.195	0.242	0.676	0.658	0.548	0.605
P18	0.657	0.815	0.594	0.803	0.655	0.629	0.601	0.653	0.175	0.242	0.143	0.245	0.653	0.396	0.611	0.465	0.276	0.3	0.231	0.321	0.682	0.661	0.608	0.69
P19	0.727	0.77	0.628	0.729	0.759	0.749	0.664	0.664	0.24	0.263	0.171	0.203	0.799	0.722	0.708	0.583	0.369	0.386	0.276	0.301	0.824	0.822	0.698	0.709
P20	0.756	0.69	0.756	0.663	0.726	0.677	0.568	0.646	0.244	0.193	0.158	0.173	0.688	0.66	0.333	0.625	0.361	0.299	0.215	0.271	0.78	0.712	0.576	0.686
Avg	**0.717**	**0.732**	**0.591**	**0.683**	**0.688**	**0.672**	**0.606**	**0.674**	**0.220**	**0.219**	**0.152**	**0.198**	**0.652**	**0.596**	**0.625**	**0.663**	**0.325**	**0.316**	**0.240**	**0.301**	**0.734**	**0.712**	**0.618**	**0.715**

Table C.2: CNN-1: Evaluation metrics for classification model trained on a combination of traditional (clean) and intentionally contaminated RSVP data.

CNN-1 : Train Traditional RSVP + Intentionally Contaminated RSVP																							
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.776	0.59	0.569	0.717	0.6	0.618	0.56	0.652	0.19	0.148	0.124	0.192	0.378	0.653	0.549	0.569	0.253	0.241	0.203	0.287	0.628	0.634	0.558	0.687
P2	0.615	0.54	0.656	0.62	0.677	0.698	0.679	0.696	0.173	0.166	0.183	0.181	0.753	0.896	0.708	0.792	0.281	0.28	0.291	0.294	0.729	0.758	0.699	0.74
P3	0.9	0.9	0.9	0.9	0.5	0.5	0.5	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0.5	0.5	0.5
P4	0.792	0.821	0.576	0.749	0.764	0.765	0.623	0.753	0.287	0.318	0.148	0.251	0.729	0.694	0.681	0.757	0.412	0.437	0.243	0.377	0.832	0.823	0.635	0.802
P5	0.749	0.774	0.682	0.637	0.749	0.73	0.61	0.743	0.249	0.259	0.162	0.2	0.75	0.674	0.521	0.875	0.374	0.374	0.247	0.326	0.795	0.785	0.625	0.815
P6	0.9	0.9	0.9	0.9	0.5	0.5	0.5	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0.5	0.5	0.5
P7	0.778	0.812	0.673	0.688	0.701	0.692	0.642	0.669	0.249	0.277	0.174	0.189	0.604	0.542	0.604	0.646	0.353	0.366	0.27	0.292	0.752	0.725	0.638	0.708
P8	0.597	0.558	0.652	0.669	0.625	0.625	0.628	0.655	0.152	0.146	0.163	0.178	0.66	0.708	0.597	0.639	0.247	0.243	0.256	0.278	0.659	0.645	0.664	0.693
P9	0.657	0.701	0.619	0.785	0.689	0.695	0.693	0.701	0.188	0.205	0.179	0.254	0.729	0.688	0.785	0.597	0.298	0.315	0.292	0.357	0.74	0.746	0.745	0.744
P10	0.726	0.633	0.548	0.657	0.723	0.682	0.613	0.627	0.226	0.179	0.141	0.163	0.719	0.743	0.694	0.59	0.344	0.288	0.235	0.256	0.796	0.719	0.627	0.663
P11	0.739	0.731	0.753	0.813	0.75	0.77	0.718	0.779	0.243	0.246	0.24	0.315	0.764	0.819	0.674	0.736	0.369	0.378	0.353	0.441	0.804	0.832	0.757	0.852
P12	0.798	0.7	0.551	0.658	0.764	0.685	0.596	0.73	0.293	0.2	0.136	0.202	0.722	0.667	0.653	0.819	0.417	0.308	0.225	0.324	0.808	0.724	0.603	0.787
P13	0.682	0.54	0.426	0.684	0.638	0.581	0.527	0.655	0.174	0.13	0.108	0.182	0.583	0.632	0.653	0.618	0.268	0.216	0.185	0.281	0.664	0.592	0.512	0.687
P14	0.716	0.804	0.666	0.743	0.7	0.697	0.651	0.749	0.213	0.27	0.175	0.245	0.681	0.562	0.632	0.757	0.324	0.365	0.275	0.371	0.75	0.753	0.675	0.813
P15	0.846	0.674	0.452	0.709	0.738	0.633	0.578	0.699	0.345	0.17	0.124	0.209	0.604	0.583	0.736	0.688	0.439	0.263	0.212	0.321	0.807	0.668	0.592	0.76
P16	0.879	0.782	0.834	0.472	0.513	0.558	0.519	0.58	0.174	0.16	0.137	0.125	0.056	0.278	0.125	0.715	0.084	0.203	0.131	0.213	0.473	0.553	0.474	0.575
P17	0.462	0.536	0.328	0.658	0.581	0.603	0.531	0.584	0.125	0.137	0.108	0.145	0.729	0.688	0.785	0.493	0.213	0.229	0.189	0.224	0.602	0.621	0.519	0.585
P18	0.69	0.723	0.669	0.697	0.649	0.667	0.625	0.659	0.181	0.201	0.165	0.188	0.597	0.597	0.569	0.611	0.278	0.301	0.256	0.288	0.688	0.699	0.641	0.702
P19	0.751	0.819	0.69	0.683	0.75	0.789	0.676	0.682	0.251	0.325	0.193	0.193	0.75	0.75	0.66	0.681	0.376	0.454	0.298	0.3	0.811	0.846	0.728	0.723
P20	0.747	0.681	0.633	0.613	0.693	0.65	0.574	0.646	0.225	0.179	0.136	0.162	0.625	0.611	0.5	0.688	0.331	0.277	0.214	0.262	0.75	0.698	0.59	0.671
Avg	**0.740**	**0.710**	**0.638**	**0.702**	**0.665**	**0.656**	**0.602**	**0.662**	**0.196**	**0.185**	**0.139**	**0.178**	**0.571**	**0.589**	**0.556**	**0.613**	**0.283**	**0.276**	**0.218**	**0.274**	**0.704**	**0.691**	**0.614**	**0.700**

Table C.3: EEGNet: Evaluation metrics for classification model trained on a combination of traditional (clean) and intentionally contaminated RSVP data.

EEGNet : Train Traditional RSVP + Intentionally Contaminated RSVP																							
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.603	0.579	0.386	0.569	0.621	0.603	0.573	0.653	0.151	0.141	0.119	0.157	0.642	0.632	0.806	0.757	0.245	0.231	0.208	0.26	0.644	0.628	0.556	0.685
P2	0.61	0.704	0.66	0.701	0.663	0.694	0.632	0.667	0.167	0.205	0.166	0.193	0.729	0.681	0.597	0.625	0.272	0.315	0.26	0.295	0.708	0.729	0.676	0.683
P3	0.709	0.529	0.393	0.54	0.641	0.621	0.539	0.584	0.184	0.142	0.111	0.131	0.556	0.736	0.722	0.639	0.276	0.238	0.192	0.217	0.68	0.647	0.516	0.61
P4	0.684	0.703	0.397	0.67	0.681	0.662	0.573	0.666	0.193	0.192	0.12	0.182	0.677	0.611	0.792	0.66	0.3	0.292	0.208	0.286	0.745	0.712	0.562	0.701
P5	0.639	0.689	0.691	0.778	0.666	0.682	0.566	0.701	0.175	0.195	0.141	0.249	0.701	0.674	0.41	0.604	0.28	0.302	0.21	0.353	0.716	0.748	0.576	0.762
P6	0.77	0.714	0.396	0.665	0.693	0.671	0.584	0.669	0.24	0.2	0.123	0.182	0.597	0.618	0.819	0.674	0.342	0.302	0.213	0.287	0.753	0.708	0.588	0.702
P7	0.699	0.772	0.619	0.632	0.671	0.685	0.631	0.62	0.194	0.236	0.158	0.155	0.635	0.576	0.646	0.604	0.297	0.335	0.253	0.247	0.716	0.725	0.661	0.644
P8	0.624	0.619	0.497	0.592	0.618	0.579	0.616	0.65	0.153	0.137	0.138	0.16	0.611	0.528	0.764	0.722	0.245	0.217	0.233	0.261	0.653	0.599	0.641	0.692
P9	0.701	0.689	0.754	0.596	0.732	0.716	0.703	0.677	0.218	0.208	0.234	0.169	0.771	0.75	0.639	0.778	0.34	0.325	0.342	0.278	0.788	0.759	0.736	0.733
P10	0.72	0.731	0.607	0.592	0.738	0.677	0.603	0.662	0.229	0.21	0.145	0.164	0.76	0.611	0.597	0.75	0.352	0.312	0.233	0.269	0.8	0.714	0.624	0.677
P11	0.769	0.794	0.797	0.787	0.724	0.74	0.668	0.73	0.253	0.28	0.247	0.269	0.667	0.674	0.507	0.66	0.366	0.395	0.333	0.382	0.795	0.804	0.717	0.809
P12	0.697	0.667	0.695	0.665	0.754	0.716	0.578	0.715	0.224	0.2	0.148	0.199	0.826	0.778	0.431	0.778	0.353	0.319	0.22	0.317	0.823	0.762	0.593	0.762
P13	0.797	0.596	0.569	0.65	0.689	0.652	0.573	0.716	0.259	0.161	0.129	0.195	0.556	0.722	0.576	0.799	0.353	0.263	0.211	0.313	0.725	0.664	0.566	0.752
P14	0.847	0.679	0.726	0.803	0.736	0.704	0.7	0.78	0.345	0.2	0.217	0.304	0.597	0.736	0.667	0.75	0.438	0.315	0.328	0.433	0.812	0.77	0.741	0.836
P15	0.791	0.639	0.593	0.761	0.745	0.63	0.579	0.679	0.279	0.161	0.134	0.227	0.688	0.618	0.562	0.576	0.397	0.255	0.217	0.325	0.806	0.654	0.595	0.738
P16	0.642	0.714	0.819	0.671	0.545	0.588	0.542	0.57	0.123	0.158	0.163	0.14	0.424	0.431	0.194	0.444	0.191	0.231	0.177	0.213	0.536	0.589	0.512	0.572
P17	0.584	0.573	0.534	0.715	0.649	0.611	0.537	0.576	0.158	0.144	0.114	0.152	0.729	0.66	0.542	0.403	0.26	0.236	0.189	0.221	0.677	0.624	0.534	0.593
P18	0.671	0.6	0.706	0.764	0.703	0.66	0.596	0.671	0.197	0.165	0.16	0.225	0.743	0.736	0.458	0.556	0.311	0.269	0.237	0.32	0.746	0.71	0.595	0.729
P19	0.788	0.835	0.586	0.601	0.753	0.806	0.647	0.68	0.279	0.351	0.158	0.171	0.708	0.771	0.722	0.778	0.401	0.483	0.259	0.281	0.818	0.864	0.676	0.71
P20	0.692	0.588	0.553	0.668	0.69	0.654	0.616	0.646	0.199	0.16	0.143	0.174	0.688	0.736	0.694	0.618	0.309	0.263	0.237	0.271	0.734	0.689	0.657	0.674
Avg	**0.701**	**0.670**	**0.598**	**0.671**	**0.685**	**0.667**	**0.602**	**0.665**	**0.211**	**0.192**	**0.153**	**0.189**	**0.665**	**0.663**	**0.607**	**0.658**	**0.316**	**0.294**	**0.238**	**0.291**	**0.733**	**0.704**	**0.616**	**0.703**

Table C.4: MoE-EEGNet: Evaluation metrics for classification model trained on a combination of traditional (clean) and intentionally contaminated RSVP data.

MoE-EEGNet : Train Traditional RSVP + Intentionally Contaminated RSVP																							
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.637	0.659	0.526	0.56	0.601	0.604	0.552	0.632	0.149	0.154	0.119	0.149	0.556	0.535	0.583	0.722	0.235	0.239	0.198	0.247	0.636	0.637	0.551	0.669
P2	0.636	0.599	0.742	0.474	0.653	0.666	0.619	0.64	0.169	0.166	0.186	0.142	0.674	0.75	0.465	0.847	0.27	0.272	0.265	0.244	0.702	0.721	0.637	0.673
P3	0.604	0.568	0.635	0.51	0.64	0.674	0.6	0.641	0.158	0.163	0.148	0.146	0.684	0.806	0.556	0.806	0.257	0.272	0.234	0.247	0.696	0.708	0.6	0.678
P4	0.692	0.537	0.588	0.511	0.631	0.601	0.57	0.583	0.174	0.136	0.13	0.129	0.556	0.681	0.549	0.674	0.265	0.227	0.21	0.216	0.678	0.624	0.578	0.596
P5	0.743	0.675	0.688	0.683	0.743	0.718	0.574	0.725	0.243	0.203	0.145	0.209	0.743	0.771	0.431	0.778	0.366	0.322	0.216	0.329	0.793	0.767	0.572	0.77
P6	0.721	0.658	0.651	0.681	0.694	0.677	0.618	0.647	0.212	0.183	0.158	0.178	0.66	0.701	0.576	0.604	0.321	0.291	0.248	0.275	0.736	0.7	0.622	0.665
P7	0.752	0.778	0.639	0.671	0.685	0.704	0.645	0.681	0.224	0.251	0.167	0.189	0.601	0.611	0.653	0.694	0.326	0.356	0.266	0.297	0.74	0.74	0.672	0.694
P8	0.643	0.628	0.74	0.559	0.632	0.621	0.679	0.653	0.162	0.155	0.215	0.156	0.618	0.611	0.604	0.771	0.257	0.248	0.317	0.259	0.68	0.639	0.703	0.694
P9	0.782	0.719	0.666	0.739	0.786	0.723	0.688	0.716	0.286	0.223	0.19	0.23	0.792	0.729	0.715	0.688	0.42	0.341	0.3	0.345	0.849	0.776	0.725	0.772
P10	0.742	0.785	0.607	0.597	0.744	0.711	0.621	0.625	0.243	0.259	0.152	0.152	0.747	0.618	0.639	0.66	0.367	0.365	0.245	0.247	0.802	0.759	0.629	0.666
P11	0.706	0.747	0.741	0.706	0.738	0.748	0.649	0.76	0.223	0.248	0.201	0.23	0.778	0.75	0.535	0.826	0.346	0.372	0.292	0.36	0.791	0.812	0.676	0.811
P12	0.817	0.667	0.558	0.755	0.731	0.667	0.591	0.688	0.3	0.182	0.135	0.227	0.625	0.667	0.632	0.604	0.405	0.286	0.222	0.33	0.778	0.699	0.596	0.742
P13	0.665	0.606	0.703	0.61	0.644	0.636	0.545	0.672	0.172	0.157	0.131	0.17	0.618	0.674	0.347	0.75	0.27	0.255	0.19	0.278	0.66	0.659	0.529	0.715
P14	0.714	0.791	0.749	0.809	0.708	0.699	0.632	0.736	0.215	0.258	0.196	0.293	0.701	0.583	0.486	0.646	0.329	0.358	0.279	0.403	0.762	0.738	0.663	0.792
P15	0.79	0.747	0.526	0.758	0.735	0.656	0.625	0.674	0.274	0.207	0.143	0.223	0.667	0.542	0.75	0.569	0.388	0.3	0.24	0.32	0.785	0.708	0.646	0.729
P16	0.164	0.494	0.244	0.722	0.511	0.583	0.537	0.568	0.102	0.127	0.108	0.148	0.944	0.694	0.903	0.375	0.184	0.215	0.193	0.213	0.472	0.583	0.513	0.58
P17	0.649	0.522	0.509	0.679	0.62	0.627	0.548	0.544	0.158	0.143	0.117	0.127	0.583	0.757	0.597	0.375	0.249	0.241	0.196	0.189	0.662	0.646	0.539	0.515
P18	0.778	0.628	0.603	0.762	0.639	0.658	0.616	0.642	0.216	0.169	0.149	0.208	0.465	0.694	0.632	0.493	0.295	0.272	0.241	0.293	0.679	0.685	0.616	0.679
P19	0.742	0.807	0.442	0.819	0.727	0.76	0.622	0.64	0.236	0.301	0.135	0.254	0.708	0.701	0.847	0.417	0.354	0.421	0.233	0.316	0.774	0.827	0.671	0.676
P20	0.666	0.598	0.625	0.56	0.672	0.662	0.591	0.62	0.184	0.165	0.143	0.145	0.681	0.743	0.549	0.694	0.29	0.27	0.226	0.24	0.707	0.702	0.596	0.648
Avg	**0.682**	**0.660**	**0.609**	**0.658**	**0.676**	**0.669**	**0.606**	**0.654**	**0.205**	**0.192**	**0.153**	**0.185**	**0.670**	**0.680**	**0.602**	**0.649**	**0.309**	**0.296**	**0.240**	**0.282**	**0.719**	**0.706**	**0.616**	**0.688**

Table C.5: Multiband EEGNet: Evaluation metrics for classification model trained on a combination of traditional (clean) and intentionally contaminated RSVP data.

Multiband EEGNet : Train Traditional RSVP + Intentionally Contaminated RSVP																							
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.619	0.607	0.621	0.597	0.614	0.621	0.57	0.646	0.151	0.152	0.133	0.159	0.608	0.639	0.507	0.708	0.242	0.245	0.211	0.26	0.649	0.654	0.581	0.684
P2	0.711	0.736	0.651	0.684	0.657	0.696	0.677	0.732	0.192	0.22	0.181	0.212	0.59	0.646	0.708	0.792	0.29	0.329	0.289	0.334	0.701	0.731	0.702	0.776
P3	0.723	0.745	0.526	0.824	0.717	0.716	0.564	0.742	0.222	0.234	0.123	0.313	0.708	0.681	0.611	0.639	0.339	0.348	0.205	0.42	0.778	0.765	0.547	0.797
P4	0.728	0.651	0.605	0.806	0.755	0.713	0.605	0.75	0.239	0.194	0.145	0.295	0.788	0.792	0.604	0.681	0.367	0.312	0.234	0.412	0.826	0.778	0.632	0.821
P5	0.814	0.763	0.604	0.704	0.737	0.708	0.62	0.715	0.299	0.241	0.151	0.213	0.642	0.639	0.639	0.729	0.408	0.35	0.244	0.33	0.816	0.779	0.646	0.784
P6	0.778	0.767	0.606	0.753	0.724	0.648	0.59	0.693	0.26	0.214	0.139	0.228	0.656	0.5	0.569	0.618	0.372	0.3	0.224	0.333	0.801	0.678	0.604	0.746
P7	0.768	0.729	0.677	0.718	0.743	0.717	0.679	0.732	0.259	0.225	0.19	0.226	0.712	0.701	0.681	0.75	0.38	0.341	0.297	0.347	0.811	0.766	0.703	0.798
P8	0.543	0.676	0.638	0.445	0.641	0.604	0.645	0.615	0.15	0.157	0.166	0.133	0.764	0.514	0.653	0.826	0.251	0.241	0.265	0.23	0.689	0.624	0.658	0.652
P9	0.765	0.794	0.773	0.713	0.767	0.719	0.689	0.695	0.266	0.271	0.239	0.209	0.771	0.625	0.583	0.674	0.396	0.378	0.339	0.319	0.837	0.786	0.754	0.752
P10	0.762	0.639	0.74	0.685	0.76	0.676	0.605	0.695	0.262	0.178	0.176	0.198	0.757	0.722	0.438	0.708	0.389	0.286	0.251	0.31	0.836	0.718	0.616	0.74
P11	0.749	0.757	0.697	0.86	0.765	0.779	0.674	0.805	0.255	0.265	0.194	0.393	0.785	0.806	0.646	0.736	0.384	0.399	0.299	0.512	0.832	0.847	0.728	0.885
P12	0.774	0.735	0.787	0.758	0.776	0.723	0.641	0.733	0.277	0.231	0.224	0.248	0.778	0.708	0.458	0.701	0.408	0.349	0.301	0.367	0.841	0.777	0.638	0.783
P13	0.787	0.653	0.515	0.618	0.7	0.625	0.583	0.72	0.256	0.162	0.129	0.188	0.59	0.59	0.667	0.847	0.357	0.254	0.216	0.307	0.753	0.654	0.571	0.796
P14	0.724	0.755	0.65	0.822	0.733	0.688	0.698	0.759	0.229	0.228	0.189	0.318	0.743	0.604	0.757	0.681	0.35	0.331	0.302	0.434	0.776	0.736	0.746	0.799
P15	0.74	0.733	0.573	0.717	0.769	0.654	0.605	0.686	0.251	0.2	0.142	0.207	0.806	0.556	0.646	0.646	0.383	0.294	0.232	0.314	0.825	0.693	0.63	0.745
P16	0.82	0.71	0.483	0.678	0.554	0.638	0.568	0.611	0.179	0.183	0.122	0.161	0.222	0.549	0.674	0.528	0.198	0.274	0.207	0.247	0.546	0.651	0.569	0.643
P17	0.582	0.631	0.631	0.626	0.675	0.635	0.595	0.622	0.166	0.161	0.145	0.155	0.792	0.639	0.549	0.618	0.275	0.257	0.229	0.248	0.712	0.666	0.598	0.639
P18	0.651	0.763	0.472	0.677	0.677	0.668	0.617	0.691	0.181	0.223	0.136	0.194	0.708	0.549	0.799	0.708	0.289	0.317	0.232	0.305	0.727	0.711	0.648	0.744
P19	0.769	0.799	0.585	0.626	0.761	0.747	0.658	0.688	0.267	0.287	0.161	0.179	0.75	0.681	0.75	0.764	0.394	0.404	0.265	0.29	0.822	0.813	0.702	0.746
P20	0.787	0.613	0.648	0.632	0.74	0.68	0.635	0.7	0.273	0.174	0.165	0.185	0.681	0.764	0.618	0.785	0.39	0.283	0.26	0.299	0.804	0.741	0.643	0.75
Avg	**0.729**	**0.712**	**0.624**	**0.697**	**0.713**	**0.682**	**0.625**	**0.701**	**0.231**	**0.210**	**0.162**	**0.220**	**0.692**	**0.645**	**0.627**	**0.706**	**0.343**	**0.314**	**0.255**	**0.330**	**0.769**	**0.728**	**0.645**	**0.754**

Table C.6: Multiband MoE-EEGNet: Evaluation metrics for classification model trained on a combination of traditional (clean) and intentionally contaminated RSVP data.

Multiband MoE-EEGNet : Train Traditional RSVP + Intentionally Contaminated RSVP																							
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.753	0.662	0.369	0.576	0.613	0.624	0.538	0.678	0.187	0.163	0.11	0.166	0.438	0.576	0.75	0.806	0.262	0.255	0.192	0.275	0.659	0.636	0.531	0.716
P2	0.726	0.695	0.597	0.62	0.672	0.704	0.674	0.721	0.205	0.206	0.168	0.189	0.604	0.715	0.771	0.847	0.306	0.319	0.276	0.308	0.725	0.756	0.71	0.775
P3	0.689	0.78	0.499	0.688	0.694	0.674	0.57	0.722	0.2	0.237	0.124	0.21	0.701	0.542	0.66	0.764	0.311	0.33	0.208	0.329	0.754	0.722	0.561	0.777
P4	0.751	0.684	0.599	0.71	0.771	0.735	0.601	0.749	0.258	0.213	0.143	0.228	0.795	0.799	0.604	0.799	0.39	0.336	0.232	0.355	0.834	0.798	0.609	0.832
P5	0.737	0.739	0.638	0.699	0.672	0.667	0.654	0.645	0.21	0.209	0.17	0.182	0.59	0.576	0.674	0.576	0.31	0.306	0.271	0.277	0.72	0.715	0.7	0.688
P6	0.762	0.726	0.634	0.744	0.74	0.669	0.63	0.701	0.254	0.204	0.16	0.227	0.712	0.597	0.625	0.646	0.374	0.304	0.255	0.336	0.805	0.711	0.649	0.728
P7	0.777	0.864	0.703	0.785	0.742	0.696	0.668	0.72	0.266	0.365	0.194	0.263	0.698	0.486	0.625	0.639	0.385	0.417	0.296	0.372	0.806	0.755	0.715	0.778
P8	0.668	0.671	0.674	0.624	0.652	0.586	0.662	0.655	0.176	0.147	0.182	0.167	0.632	0.479	0.646	0.694	0.276	0.225	0.284	0.27	0.69	0.585	0.703	0.69
P9	0.67	0.792	0.574	0.65	0.755	0.727	0.655	0.682	0.214	0.273	0.159	0.183	0.861	0.646	0.757	0.722	0.343	0.384	0.262	0.292	0.835	0.775	0.7	0.743
P10	0.658	0.639	0.568	0.783	0.721	0.679	0.59	0.66	0.199	0.179	0.136	0.232	0.799	0.729	0.618	0.507	0.319	0.288	0.222	0.318	0.78	0.71	0.599	0.703
P11	0.785	0.68	0.681	0.758	0.769	0.751	0.665	0.785	0.283	0.216	0.185	0.268	0.75	0.84	0.646	0.819	0.411	0.344	0.288	0.403	0.825	0.81	0.706	0.856
P12	0.723	0.709	0.624	0.79	0.757	0.718	0.618	0.71	0.237	0.216	0.153	0.263	0.799	0.729	0.611	0.611	0.366	0.334	0.245	0.367	0.832	0.768	0.642	0.778
P13	0.639	0.553	0.344	0.756	0.673	0.61	0.552	0.716	0.177	0.141	0.113	0.241	0.715	0.681	0.812	0.667	0.284	0.233	0.199	0.354	0.72	0.64	0.563	0.776
P14	0.725	0.775	0.608	0.807	0.77	0.711	0.671	0.735	0.243	0.251	0.17	0.291	0.826	0.632	0.75	0.646	0.375	0.36	0.277	0.401	0.827	0.761	0.717	0.789
P15	0.728	0.529	0.502	0.694	0.75	0.63	0.609	0.685	0.237	0.145	0.136	0.198	0.778	0.757	0.743	0.674	0.364	0.243	0.23	0.306	0.808	0.671	0.645	0.743
P16	0.363	0.628	0.323	0.531	0.572	0.617	0.544	0.625	0.118	0.154	0.111	0.143	0.833	0.604	0.819	0.743	0.207	0.245	0.195	0.24	0.58	0.631	0.535	0.644
P17	0.699	0.65	0.497	0.618	0.672	0.654	0.572	0.63	0.195	0.173	0.124	0.157	0.639	0.66	0.667	0.646	0.298	0.274	0.209	0.253	0.709	0.671	0.582	0.659
P18	0.769	0.746	0.593	0.762	0.69	0.68	0.626	0.698	0.237	0.218	0.151	0.237	0.59	0.597	0.667	0.618	0.339	0.32	0.247	0.342	0.721	0.724	0.631	0.729
P19	0.773	0.71	0.749	0.621	0.794	0.731	0.647	0.691	0.282	0.222	0.204	0.179	0.819	0.757	0.521	0.778	0.419	0.343	0.293	0.291	0.844	0.777	0.695	0.74
P20	0.712	0.611	0.49	0.706	0.723	0.676	0.612	0.707	0.219	0.172	0.136	0.211	0.736	0.757	0.764	0.708	0.338	0.28	0.231	0.325	0.759	0.74	0.634	0.761
Avg	**0.705**	**0.710**	**0.563**	**0.696**	**0.333**	**0.761**	**0.617**	**0.695**	**0.219**	**0.205**	**0.151**	**0.211**	**0.715**	**0.657**	**0.686**	**0.695**	**0.333**	**0.307**	**0.245**	**0.320**	**0.761**	**0.717**	**0.641**	**0.745**

Table C.7: EEGTransformer: Evaluation metrics for classification model trained on a combination of traditional (clean) and intentionally contaminated RSVP data.

EEGTransformer : Train Traditional RSVP + Intentionally Contaminated RSVP																							
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.637	0.463	0.637	0.636	0.605	0.572	0.557	0.634	0.15	0.122	0.129	0.162	0.566	0.708	0.458	0.632	0.238	0.209	0.202	0.258	0.63	0.582	0.548	0.669
P2	0.548	0.718	0.546	0.684	0.63	0.64	0.618	0.698	0.147	0.187	0.143	0.199	0.733	0.542	0.708	0.715	0.245	0.278	0.238	0.312	0.663	0.667	0.661	0.77
P3	0.756	0.726	0.606	0.724	0.604	0.607	0.541	0.637	0.182	0.172	0.119	0.187	0.413	0.458	0.458	0.528	0.253	0.25	0.189	0.276	0.636	0.638	0.528	0.669
P4	0.583	0.654	0.675	0.778	0.629	0.65	0.594	0.651	0.151	0.172	0.152	0.223	0.688	0.646	0.493	0.493	0.248	0.272	0.233	0.307	0.672	0.652	0.581	0.703
P5	0.678	0.594	0.732	0.697	0.667	0.642	0.607	0.708	0.185	0.157	0.175	0.208	0.653	0.701	0.451	0.722	0.289	0.257	0.252	0.322	0.71	0.669	0.623	0.768
P6	0.715	0.488	0.688	0.772	0.709	0.669	0.592	0.636	0.215	0.152	0.154	0.211	0.701	0.896	0.472	0.465	0.33	0.259	0.232	0.29	0.776	0.727	0.598	0.668
P7	0.802	0.736	0.664	0.608	0.711	0.687	0.572	0.684	0.275	0.216	0.14	0.174	0.597	0.625	0.458	0.778	0.376	0.321	0.214	0.284	0.762	0.729	0.582	0.751
P8	0.654	0.592	0.634	0.671	0.609	0.604	0.615	0.663	0.155	0.143	0.154	0.181	0.552	0.618	0.59	0.653	0.242	0.233	0.244	0.284	0.637	0.598	0.637	0.696
P9	0.683	0.672	0.576	0.666	0.651	0.583	0.579	0.679	0.18	0.147	0.132	0.186	0.611	0.472	0.583	0.694	0.278	0.224	0.216	0.294	0.69	0.59	0.581	0.734
P10	0.741	0.552	0.751	0.794	0.737	0.625	0.581	0.706	0.24	0.146	0.165	0.265	0.733	0.715	0.368	0.597	0.361	0.242	0.228	0.367	0.803	0.649	0.601	0.749
P11	0.726	0.81	0.69	0.817	0.712	0.762	0.615	0.679	0.222	0.305	0.166	0.275	0.694	0.701	0.521	0.507	0.337	0.425	0.252	0.357	0.774	0.826	0.639	0.735
P12	0.8	0.511	0.631	0.713	0.722	0.596	0.545	0.726	0.278	0.133	0.123	0.222	0.625	0.701	0.438	0.743	0.385	0.223	0.191	0.341	0.78	0.619	0.517	0.782
P13	0.649	0.44	0.613	0.676	0.605	0.55	0.538	0.573	0.152	0.115	0.118	0.142	0.549	0.688	0.444	0.444	0.238	0.197	0.187	0.215	0.608	0.539	0.522	0.579
P14	0.641	0.74	0.652	0.774	0.696	0.643	0.649	0.776	0.185	0.197	0.171	0.277	0.764	0.521	0.646	0.778	0.299	0.286	0.271	0.408	0.754	0.697	0.677	0.829
P15	0.642	0.801	0.832	0.64	0.684	0.609	0.53	0.639	0.182	0.214	0.155	0.165	0.736	0.368	0.153	0.639	0.291	0.27	0.154	0.262	0.745	0.635	0.52	0.679
P16	0.234	0.465	0.802	0.591	0.522	0.549	0.526	0.544	0.105	0.115	0.135	0.12	0.882	0.653	0.181	0.486	0.187	0.196	0.154	0.192	0.492	0.527	0.51	0.528
P17	0.69	0.539	0.628	0.386	0.615	0.574	0.544	0.56	0.166	0.128	0.122	0.116	0.521	0.618	0.438	0.778	0.251	0.211	0.191	0.202	0.643	0.584	0.543	0.535
P18	0.717	0.552	0.698	0.623	0.639	0.588	0.61	0.618	0.186	0.133	0.166	0.153	0.542	0.632	0.5	0.611	0.277	0.22	0.249	0.245	0.673	0.606	0.635	0.646
P19	0.726	0.836	0.694	0.753	0.684	0.705	0.654	0.665	0.21	0.315	0.185	0.215	0.632	0.542	0.604	0.556	0.315	0.398	0.283	0.31	0.722	0.765	0.687	0.706
P20	0.638	0.65	0.263	0.675	0.706	0.71	0.551	0.693	0.188	0.193	0.111	0.194	0.792	0.785	0.91	0.715	0.304	0.31	0.198	0.306	0.762	0.757	0.542	0.754
Avg	**0.663**	**0.626**	**0.650**	**0.683**	**0.656**	**0.628**	**0.580**	**0.658**	**0.187**	**0.173**	**0.145**	**0.193**	**0.649**	**0.629**	**0.493**	**0.626**	**0.287**	**0.264**	**0.218**	**0.291**	**0.696**	**0.652**	**0.586**	**0.697**

Table C.8: MoE-Transformer: Evaluation metrics for classification model trained on a combination of traditional (clean) and intentionally contaminated RSVP data.

MoE-Transformer : Train Traditional RSVP + Intentionally Contaminated RSVP																							
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.614	0.778	0.733	0.597	0.582	0.605	0.558	0.625	0.137	0.195	0.145	0.152	0.542	0.389	0.34	0.66	0.219	0.26	0.203	0.246	0.599	0.625	0.553	0.639
P2	0.69	0.638	0.467	0.86	0.642	0.648	0.583	0.561	0.178	0.168	0.126	0.243	0.583	0.66	0.729	0.188	0.273	0.267	0.215	0.212	0.68	0.671	0.593	0.56
P3	0.657	0.719	0.502	0.771	0.644	0.656	0.569	0.635	0.17	0.195	0.124	0.209	0.628	0.576	0.653	0.465	0.268	0.291	0.208	0.289	0.664	0.675	0.548	0.639
P4	0.78	0.638	0.465	0.763	0.705	0.703	0.558	0.68	0.252	0.187	0.118	0.229	0.611	0.785	0.674	0.576	0.357	0.303	0.201	0.327	0.774	0.767	0.567	0.725
P5	0.74	0.736	0.74	0.582	0.744	0.684	0.581	0.666	0.242	0.215	0.161	0.163	0.75	0.618	0.382	0.771	0.366	0.319	0.227	0.269	0.809	0.744	0.55	0.715
P6	0.698	0.663	0.492	0.519	0.66	0.609	0.563	0.566	0.189	0.157	0.121	0.124	0.611	0.542	0.653	0.625	0.288	0.243	0.204	0.206	0.693	0.639	0.572	0.572
P7	0.732	0.713	0.651	0.717	0.688	0.699	0.618	0.667	0.215	0.211	0.158	0.199	0.632	0.681	0.576	0.604	0.321	0.322	0.249	0.299	0.725	0.753	0.631	0.694
P8	0.545	0.485	0.617	0.515	0.615	0.612	0.574	0.576	0.142	0.135	0.134	0.127	0.701	0.771	0.521	0.653	0.236	0.23	0.214	0.212	0.649	0.611	0.596	0.581
P9	0.393	0.701	0.735	0.646	0.628	0.704	0.696	0.683	0.133	0.208	0.22	0.182	0.92	0.708	0.646	0.729	0.233	0.322	0.328	0.292	0.661	0.738	0.709	0.731
P10	0.692	0.651	0.681	0.667	0.716	0.658	0.61	0.648	0.209	0.174	0.161	0.174	0.747	0.667	0.521	0.625	0.326	0.276	0.246	0.273	0.769	0.716	0.633	0.671
P11	0.787	0.557	0.641	0.644	0.768	0.68	0.655	0.737	0.285	0.163	0.171	0.2	0.743	0.833	0.674	0.854	0.412	0.273	0.273	0.324	0.786	0.703	0.676	0.747
P12	0.788	0.671	0.702	0.851	0.753	0.703	0.591	0.707	0.279	0.197	0.157	0.341	0.708	0.743	0.451	0.528	0.401	0.311	0.233	0.414	0.76	0.729	0.59	0.754
P13	0.726	0.613	0.532	0.693	0.681	0.597	0.561	0.688	0.209	0.143	0.123	0.198	0.625	0.576	0.597	0.681	0.314	0.23	0.203	0.307	0.705	0.616	0.545	0.725
P14	0.841	0.77	0.815	0.84	0.736	0.706	0.607	0.784	0.336	0.245	0.225	0.352	0.604	0.625	0.347	0.715	0.432	0.352	0.273	0.471	0.77	0.746	0.64	0.834
P15	0.683	0.543	0.664	0.542	0.707	0.589	0.591	0.588	0.202	0.133	0.149	0.132	0.736	0.646	0.5	0.646	0.317	0.22	0.229	0.22	0.747	0.603	0.592	0.606
P16	0.81	0.651	0.74	0.566	0.521	0.574	0.522	0.577	0.131	0.139	0.119	0.131	0.16	0.479	0.25	0.59	0.144	0.215	0.161	0.214	0.475	0.589	0.488	0.57
P17	0.654	0.692	0.447	0.512	0.641	0.613	0.557	0.593	0.169	0.166	0.117	0.132	0.625	0.514	0.694	0.694	0.265	0.25	0.201	0.221	0.676	0.619	0.548	0.576
P18	0.732	0.769	0.572	0.633	0.657	0.671	0.636	0.651	0.2	0.228	0.152	0.168	0.562	0.549	0.715	0.674	0.296	0.322	0.251	0.269	0.675	0.703	0.649	0.675
P19	0.746	0.65	0.613	0.694	0.745	0.747	0.677	0.666	0.245	0.205	0.173	0.19	0.743	0.868	0.757	0.632	0.369	0.332	0.281	0.292	0.788	0.814	0.715	0.692
P20	0.747	0.526	0.672	0.673	0.702	0.656	0.599	0.667	0.229	0.152	0.154	0.184	0.646	0.819	0.507	0.66	0.338	0.257	0.236	0.287	0.728	0.712	0.601	0.703
Avg	**0.702**	**0.658**	**0.624**	**0.664**	**0.676**	**0.655**	**0.595**	**0.648**	**0.207**	**0.180**	**0.150**	**0.191**	**0.643**	**0.652**	**0.559**	**0.628**	**0.308**	**0.279**	**0.231**	**0.282**	**0.706**	**0.688**	**0.599**	**0.670**

Table C.9: EEG Conformer: Evaluation metrics for classification model trained on a combination of traditional (clean) and intentionally contaminated RSVP data.

EEG Conformer : Train Traditional RSVP + Intentionally Contaminated RSVP																							
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.607	0.431	0.52	0.718	0.587	0.563	0.545	0.6	0.139	0.119	0.116	0.166	0.562	0.729	0.576	0.451	0.223	0.204	0.194	0.243	0.603	0.577	0.548	0.627
P2	0.656	0.632	0.676	0.616	0.636	0.669	0.659	0.722	0.167	0.174	0.181	0.188	0.611	0.715	0.639	0.854	0.262	0.28	0.283	0.308	0.662	0.69	0.684	0.767
P3	0.794	0.624	0.483	0.815	0.693	0.665	0.571	0.715	0.258	0.171	0.123	0.291	0.566	0.715	0.681	0.59	0.354	0.276	0.208	0.39	0.73	0.708	0.551	0.724
P4	0.809	0.7	0.519	0.725	0.735	0.719	0.551	0.727	0.293	0.213	0.118	0.227	0.642	0.743	0.59	0.729	0.402	0.331	0.197	0.347	0.809	0.758	0.55	0.79
P5	0.777	0.789	0.469	0.746	0.698	0.679	0.551	0.658	0.247	0.247	0.116	0.208	0.601	0.542	0.653	0.549	0.35	0.339	0.197	0.302	0.742	0.687	0.546	0.702
P6	0.707	0.75	0.509	0.752	0.743	0.685	0.582	0.705	0.225	0.223	0.128	0.233	0.788	0.604	0.674	0.646	0.35	0.326	0.215	0.343	0.817	0.719	0.592	0.742
P7	0.818	0.814	0.525	0.664	0.734	0.721	0.609	0.702	0.302	0.292	0.138	0.194	0.628	0.604	0.715	0.75	0.408	0.394	0.231	0.309	0.789	0.781	0.634	0.766
P8	0.616	0.339	0.622	0.657	0.629	0.577	0.657	0.624	0.156	0.119	0.167	0.162	0.646	0.875	0.701	0.583	0.252	0.209	0.27	0.254	0.662	0.564	0.695	0.65
P9	0.835	0.736	0.742	0.751	0.728	0.758	0.736	0.723	0.323	0.245	0.24	0.24	0.594	0.785	0.729	0.688	0.418	0.373	0.361	0.356	0.779	0.819	0.775	0.772
P10	0.816	0.738	0.599	0.797	0.762	0.675	0.623	0.702	0.311	0.212	0.151	0.265	0.694	0.597	0.653	0.583	0.43	0.313	0.246	0.364	0.836	0.722	0.648	0.739
P11	0.809	0.748	0.692	0.763	0.777	0.783	0.659	0.797	0.309	0.26	0.187	0.276	0.736	0.826	0.618	0.84	0.435	0.396	0.287	0.415	0.829	0.842	0.687	0.867
P12	0.824	0.654	0.549	0.768	0.785	0.694	0.604	0.775	0.329	0.188	0.139	0.272	0.736	0.743	0.674	0.785	0.455	0.301	0.23	0.404	0.87	0.746	0.624	0.82
P13	0.731	0.686	0.611	0.709	0.659	0.597	0.54	0.718	0.201	0.156	0.119	0.216	0.569	0.486	0.451	0.729	0.297	0.236	0.188	0.334	0.69	0.605	0.537	0.756
P14	0.812	0.786	0.75	0.847	0.695	0.693	0.58	0.773	0.278	0.252	0.165	0.359	0.549	0.576	0.368	0.681	0.369	0.35	0.227	0.47	0.732	0.729	0.592	0.821
P15	0.681	0.719	0.57	0.594	0.69	0.631	0.557	0.685	0.195	0.183	0.124	0.171	0.701	0.521	0.542	0.799	0.306	0.271	0.201	0.282	0.728	0.671	0.562	0.737
P16	0.574	0.794	0.393	0.635	0.59	0.577	0.533	0.606	0.136	0.183	0.109	0.15	0.611	0.306	0.708	0.569	0.223	0.229	0.189	0.238	0.586	0.595	0.514	0.602
P17	0.535	0.479	0.581	0.619	0.603	0.612	0.551	0.551	0.137	0.135	0.122	0.124	0.688	0.778	0.514	0.465	0.228	0.23	0.197	0.196	0.621	0.638	0.556	0.547
P18	0.738	0.769	0.532	0.659	0.7	0.696	0.62	0.706	0.223	0.24	0.142	0.194	0.653	0.604	0.729	0.764	0.333	0.344	0.238	0.309	0.76	0.744	0.643	0.754
P19	0.817	0.694	0.701	0.74	0.769	0.74	0.652	0.704	0.316	0.218	0.186	0.226	0.708	0.799	0.59	0.66	0.437	0.343	0.283	0.336	0.829	0.803	0.686	0.775
P20	0.704	0.728	0.654	0.731	0.718	0.713	0.638	0.718	0.215	0.224	0.167	0.226	0.736	0.694	0.618	0.701	0.332	0.338	0.263	0.342	0.757	0.774	0.637	0.771
Avg	**0.733**	**0.680**	**0.584**	**0.715**	**0.696**	**0.672**	**0.600**	**0.695**	**0.238**	**0.202**	**0.146**	**0.219**	**0.650**	**0.662**	**0.621**	**0.670**	**0.343**	**0.304**	**0.235**	**0.327**	**0.741**	**0.708**	**0.613**	**0.736**

Table C.10: CNN1-Transformer: Evaluation metrics for classification model trained on a combination of traditional (clean) and intentionally contaminated RSVP data.

CNN1-Transformer : Train Traditional RSVP + Intentionally Contaminated RSVP																							
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.542	0.383	0.424	0.641	0.577	0.577	0.591	0.637	0.129	0.12	0.126	0.164	0.622	0.819	0.799	0.632	0.213	0.21	0.217	0.26	0.609	0.601	0.598	0.68
P2	0.588	0.68	0.601	0.734	0.675	0.708	0.658	0.689	0.167	0.202	0.164	0.216	0.785	0.743	0.729	0.632	0.276	0.317	0.268	0.322	0.725	0.733	0.711	0.737
P3	0.705	0.751	0.267	0.764	0.688	0.664	0.546	0.724	0.203	0.214	0.11	0.249	0.667	0.556	0.896	0.674	0.311	0.309	0.196	0.363	0.747	0.728	0.525	0.77
P4	0.693	0.781	0.575	0.72	0.719	0.681	0.579	0.659	0.21	0.241	0.132	0.197	0.75	0.556	0.583	0.583	0.329	0.336	0.215	0.294	0.783	0.751	0.608	0.718
P5	0.758	0.733	0.744	0.715	0.748	0.756	0.552	0.762	0.255	0.242	0.143	0.235	0.736	0.785	0.312	0.819	0.379	0.37	0.197	0.365	0.823	0.806	0.544	0.834
P6	0.732	0.747	0.723	0.558	0.76	0.674	0.599	0.655	0.243	0.216	0.167	0.156	0.795	0.583	0.444	0.778	0.372	0.316	0.243	0.26	0.839	0.738	0.608	0.722
P7	0.687	0.726	0.665	0.639	0.702	0.709	0.616	0.667	0.202	0.22	0.161	0.175	0.722	0.688	0.556	0.701	0.316	0.334	0.249	0.28	0.769	0.759	0.652	0.716
P8	0.607	0.614	0.671	0.711	0.654	0.588	0.654	0.667	0.163	0.14	0.178	0.196	0.712	0.556	0.632	0.611	0.266	0.223	0.277	0.297	0.694	0.601	0.685	0.731
P9	0.718	0.741	0.744	0.719	0.752	0.764	0.713	0.755	0.233	0.249	0.232	0.235	0.795	0.792	0.674	0.799	0.361	0.379	0.345	0.363	0.834	0.836	0.772	0.828
P10	0.779	0.747	0.671	0.563	0.708	0.702	0.65	0.674	0.253	0.229	0.176	0.163	0.618	0.646	0.625	0.812	0.359	0.338	0.275	0.271	0.785	0.761	0.688	0.723
P11	0.793	0.798	0.576	0.694	0.786	0.764	0.653	0.769	0.296	0.293	0.158	0.228	0.778	0.722	0.75	0.861	0.429	0.417	0.261	0.36	0.824	0.822	0.705	0.829
P12	0.775	0.664	0.601	0.779	0.785	0.671	0.624	0.723	0.28	0.183	0.152	0.26	0.799	0.681	0.653	0.653	0.415	0.288	0.247	0.372	0.845	0.724	0.63	0.794
P13	0.665	0.66	0.347	0.765	0.659	0.651	0.557	0.675	0.178	0.174	0.114	0.228	0.653	0.639	0.819	0.562	0.28	0.273	0.201	0.324	0.7	0.674	0.549	0.719
P14	0.762	0.675	0.512	0.724	0.698	0.665	0.627	0.726	0.237	0.184	0.142	0.226	0.618	0.653	0.771	0.729	0.342	0.287	0.24	0.345	0.766	0.725	0.658	0.778
P15	0.74	0.621	0.515	0.742	0.729	0.644	0.598	0.736	0.236	0.163	0.134	0.24	0.715	0.674	0.701	0.729	0.355	0.262	0.224	0.361	0.798	0.687	0.614	0.771
P16	0.272	0.325	0.364	0.569	0.521	0.548	0.505	0.591	0.105	0.112	0.101	0.136	0.833	0.826	0.681	0.618	0.186	0.197	0.176	0.223	0.496	0.557	0.488	0.595
P17	0.776	0.623	0.551	0.473	0.632	0.636	0.556	0.608	0.211	0.16	0.122	0.133	0.451	0.653	0.562	0.778	0.288	0.257	0.2	0.228	0.667	0.663	0.56	0.638
P18	0.59	0.721	0.721	0.656	0.652	0.644	0.598	0.648	0.16	0.19	0.166	0.172	0.729	0.549	0.444	0.639	0.262	0.282	0.242	0.271	0.698	0.67	0.611	0.698
P19	0.647	0.793	0.628	0.783	0.665	0.743	0.649	0.716	0.176	0.28	0.166	0.26	0.688	0.681	0.674	0.632	0.28	0.397	0.266	0.368	0.722	0.788	0.694	0.768
P20	0.65	0.687	0.437	0.674	0.694	0.638	0.57	0.637	0.188	0.175	0.121	0.171	0.75	0.576	0.736	0.59	0.3	0.269	0.207	0.266	0.733	0.677	0.582	0.686
Avg	**0.673**	**0.673**	**0.566**	**0.681**	**0.690**	**0.671**	**0.604**	**0.685**	**0.206**	**0.199**	**0.148**	**0.202**	**0.710**	**0.668**	**0.652**	**0.691**	**0.315**	**0.303**	**0.237**	**0.309**	**0.742**	**0.715**	**0.624**	**0.736**

Table C.11: EEGNet-Transformer: Evaluation metrics for classification model trained on a combination of traditional (clean) and intentionally contaminated RSVP data.

EEGNet-Transformer : Train Traditional RSVP + Intentionally Contaminated RSVP																							
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.645	0.435	0.601	0.515	0.601	0.6	0.566	0.641	0.15	0.129	0.129	0.147	0.545	0.806	0.521	0.799	0.235	0.222	0.207	0.248	0.631	0.632	0.555	0.675
P2	0.644	0.708	0.717	0.677	0.67	0.662	0.652	0.694	0.177	0.193	0.192	0.195	0.701	0.604	0.569	0.715	0.283	0.293	0.287	0.307	0.709	0.72	0.687	0.734
P3	0.719	0.599	0.483	0.739	0.655	0.678	0.58	0.688	0.194	0.17	0.126	0.218	0.576	0.778	0.701	0.625	0.291	0.279	0.213	0.324	0.699	0.751	0.56	0.737
P4	0.711	0.723	0.642	0.693	0.701	0.676	0.6	0.675	0.211	0.206	0.149	0.193	0.688	0.618	0.549	0.653	0.322	0.308	0.234	0.298	0.76	0.727	0.615	0.72
P5	0.748	0.767	0.453	0.78	0.764	0.762	0.566	0.776	0.254	0.266	0.12	0.281	0.785	0.757	0.708	0.771	0.384	0.394	0.206	0.412	0.821	0.814	0.581	0.846
P6	0.755	0.697	0.74	0.738	0.715	0.687	0.593	0.715	0.239	0.2	0.17	0.229	0.667	0.674	0.41	0.688	0.352	0.308	0.24	0.344	0.779	0.739	0.614	0.752
P7	0.725	0.691	0.673	0.667	0.688	0.705	0.649	0.679	0.212	0.204	0.177	0.187	0.642	0.722	0.618	0.694	0.318	0.319	0.275	0.294	0.744	0.75	0.675	0.717
P8	0.691	0.659	0.624	0.744	0.628	0.588	0.634	0.685	0.172	0.147	0.16	0.22	0.549	0.5	0.646	0.611	0.262	0.227	0.256	0.324	0.664	0.589	0.681	0.722
P9	0.703	0.664	0.643	0.639	0.724	0.696	0.691	0.691	0.216	0.192	0.184	0.184	0.75	0.736	0.75	0.757	0.336	0.305	0.296	0.295	0.793	0.734	0.738	0.719
P10	0.763	0.69	0.706	0.631	0.76	0.732	0.623	0.687	0.263	0.214	0.174	0.18	0.757	0.785	0.521	0.757	0.39	0.336	0.261	0.291	0.826	0.791	0.642	0.741
P11	0.77	0.79	0.597	0.815	0.746	0.791	0.64	0.804	0.262	0.295	0.157	0.325	0.715	0.792	0.694	0.792	0.384	0.429	0.256	0.461	0.805	0.856	0.684	0.861
P12	0.835	0.779	0.609	0.747	0.788	0.695	0.619	0.723	0.345	0.247	0.151	0.238	0.729	0.59	0.632	0.694	0.469	0.348	0.244	0.354	0.849	0.739	0.631	0.782
P13	0.567	0.561	0.16	0.665	0.639	0.614	0.521	0.688	0.152	0.143	0.104	0.189	0.729	0.681	0.972	0.715	0.252	0.237	0.188	0.299	0.653	0.632	0.504	0.735
P14	0.856	0.767	0.715	0.763	0.747	0.701	0.638	0.742	0.368	0.241	0.184	0.256	0.611	0.618	0.542	0.715	0.46	0.347	0.275	0.377	0.806	0.747	0.683	0.808
P15	0.785	0.682	0.541	0.717	0.745	0.647	0.634	0.707	0.274	0.178	0.147	0.216	0.694	0.604	0.75	0.694	0.393	0.275	0.246	0.329	0.814	0.697	0.653	0.759
P16	0.565	0.769	0.417	0.669	0.545	0.547	0.543	0.581	0.118	0.146	0.112	0.145	0.521	0.271	0.701	0.472	0.193	0.19	0.194	0.222	0.539	0.56	0.55	0.589
P17	0.523	0.711	0.241	0.383	0.621	0.599	0.526	0.568	0.141	0.163	0.106	0.118	0.743	0.458	0.882	0.799	0.238	0.241	0.189	0.206	0.65	0.625	0.524	0.575
P18	0.642	0.703	0.497	0.658	0.659	0.671	0.625	0.653	0.173	0.195	0.14	0.174	0.681	0.632	0.785	0.646	0.275	0.298	0.238	0.274	0.699	0.703	0.652	0.696
P19	0.726	0.757	0.681	0.767	0.715	0.751	0.659	0.707	0.223	0.255	0.183	0.243	0.701	0.743	0.632	0.632	0.339	0.379	0.284	0.351	0.763	0.806	0.71	0.751
P20	0.794	0.592	0.692	0.665	0.707	0.662	0.591	0.641	0.265	0.164	0.154	0.171	0.597	0.75	0.465	0.611	0.368	0.269	0.232	0.267	0.762	0.709	0.627	0.679
Avg	**0.708**	**0.687**	**0.571**	**0.683**	**0.690**	**0.673**	**0.607**	**0.687**	**0.220**	**0.197**	**0.150**	**0.205**	**0.669**	**0.655**	**0.652**	**0.692**	**0.327**	**0.300**	**0.241**	**0.313**	**0.738**	**0.716**	**0.628**	**0.729**

Table C.12: Multiband EEGNet-Transformer: Evaluation metrics for classification model trained on a combination of traditional (clean) and intentionally contaminated RSVP data.

Multiband EEGNet-Transformer : Train Traditional RSVP + Intentionally Contaminated RSVP																							
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.602	0.587	0.377	0.655	0.598	0.598	0.546	0.657	0.142	0.14	0.112	0.175	0.594	0.611	0.757	0.66	0.23	0.228	0.196	0.277	0.628	0.62	0.534	0.694
P2	0.696	0.551	0.603	0.574	0.621	0.667	0.647	0.674	0.171	0.159	0.161	0.165	0.528	0.812	0.701	0.799	0.258	0.266	0.261	0.273	0.669	0.712	0.68	0.723
P3	0.65	0.752	0.46	0.79	0.702	0.72	0.589	0.716	0.19	0.24	0.127	0.265	0.767	0.681	0.75	0.625	0.305	0.354	0.217	0.373	0.746	0.778	0.578	0.773
P4	0.803	0.717	0.691	0.779	0.718	0.716	0.591	0.745	0.279	0.219	0.154	0.269	0.611	0.715	0.465	0.701	0.383	0.336	0.231	0.388	0.784	0.773	0.616	0.8
P5	0.711	0.768	0.677	0.751	0.736	0.745	0.614	0.775	0.224	0.26	0.162	0.26	0.767	0.715	0.535	0.806	0.347	0.381	0.249	0.393	0.809	0.814	0.65	0.844
P6	0.728	0.761	0.594	0.721	0.742	0.645	0.611	0.66	0.234	0.209	0.146	0.197	0.76	0.5	0.632	0.583	0.358	0.295	0.238	0.295	0.81	0.679	0.64	0.693
P7	0.802	0.71	0.784	0.709	0.744	0.715	0.673	0.73	0.289	0.216	0.24	0.221	0.67	0.722	0.535	0.757	0.404	0.332	0.331	0.342	0.816	0.777	0.709	0.778
P8	0.698	0.788	0.774	0.674	0.641	0.586	0.649	0.662	0.18	0.187	0.22	0.182	0.569	0.333	0.493	0.646	0.274	0.239	0.304	0.284	0.675	0.595	0.681	0.706
P9	0.745	0.799	0.71	0.79	0.752	0.755	0.679	0.708	0.248	0.29	0.201	0.262	0.76	0.701	0.639	0.604	0.374	0.411	0.306	0.366	0.817	0.812	0.733	0.772
P10	0.724	0.689	0.627	0.63	0.748	0.713	0.598	0.659	0.235	0.207	0.146	0.17	0.778	0.743	0.562	0.694	0.361	0.323	0.232	0.273	0.808	0.764	0.609	0.691
P11	0.717	0.794	0.543	0.842	0.769	0.759	0.65	0.814	0.238	0.287	0.153	0.365	0.833	0.715	0.785	0.778	0.37	0.41	0.256	0.497	0.844	0.827	0.691	0.877
P12	0.806	0.677	0.827	0.782	0.787	0.731	0.611	0.746	0.31	0.209	0.241	0.272	0.764	0.799	0.34	0.701	0.441	0.331	0.282	0.391	0.861	0.778	0.626	0.801
P13	0.766	0.699	0.663	0.735	0.682	0.62	0.551	0.726	0.231	0.17	0.129	0.232	0.576	0.521	0.41	0.715	0.33	0.257	0.196	0.35	0.714	0.637	0.554	0.772
P14	0.773	0.683	0.63	0.694	0.753	0.716	0.68	0.716	0.267	0.206	0.177	0.21	0.729	0.757	0.743	0.743	0.391	0.323	0.286	0.327	0.819	0.759	0.725	0.772
P15	0.772	0.646	0.531	0.828	0.75	0.664	0.591	0.685	0.265	0.176	0.133	0.292	0.722	0.688	0.667	0.507	0.387	0.28	0.221	0.371	0.811	0.687	0.61	0.736
P16	0.49	0.604	0.524	0.636	0.559	0.589	0.562	0.613	0.12	0.139	0.123	0.153	0.646	0.569	0.611	0.583	0.202	0.223	0.204	0.243	0.556	0.588	0.569	0.642
P17	0.672	0.731	0.683	0.677	0.66	0.653	0.543	0.608	0.181	0.198	0.127	0.159	0.646	0.556	0.368	0.521	0.283	0.292	0.189	0.244	0.702	0.682	0.537	0.637
P18	0.728	0.7	0.503	0.661	0.698	0.651	0.613	0.713	0.217	0.186	0.137	0.197	0.66	0.59	0.75	0.778	0.327	0.282	0.232	0.315	0.731	0.688	0.652	0.767
P19	0.735	0.743	0.64	0.598	0.76	0.746	0.658	0.709	0.245	0.244	0.172	0.18	0.792	0.75	0.681	0.847	0.374	0.369	0.274	0.296	0.825	0.789	0.701	0.755
P20	0.844	0.624	0.688	0.61	0.747	0.686	0.62	0.697	0.345	0.178	0.167	0.179	0.625	0.764	0.535	0.806	0.444	0.289	0.255	0.293	0.813	0.729	0.65	0.76
Avg	**0.723**	**0.701**	**0.626**	**0.706**	**0.708**	**0.683**	**0.613**	**0.700**	**0.230**	**0.206**	**0.161**	**0.220**	**0.689**	**0.662**	**0.597**	**0.692**	**0.342**	**0.311**	**0.248**	**0.329**	**0.761**	**0.724**	**0.637**	**0.749**

Table C.13: EEGNet-MoE-Transformer: Evaluation metrics for classification model trained on a combination of traditional (clean) and intentionally contaminated RSVP data.

EEGNet-MoE-Transformer : Train Traditional RSVP + Intentionally Contaminated RSVP																							
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.555	0.673	0.563	0.552	0.578	0.611	0.554	0.637	0.13	0.16	0.122	0.15	0.608	0.535	0.542	0.743	0.214	0.246	0.199	0.249	0.616	0.635	0.529	0.661
P2	0.58	0.486	0.59	0.544	0.631	0.65	0.646	0.657	0.151	0.146	0.158	0.155	0.694	0.854	0.715	0.799	0.248	0.249	0.259	0.259	0.681	0.709	0.685	0.709
P3	0.705	0.748	0.833	0.72	0.648	0.625	0.512	0.659	0.186	0.192	0.125	0.197	0.576	0.472	0.111	0.583	0.281	0.273	0.118	0.294	0.71	0.712	0.542	0.704
P4	0.755	0.731	0.737	0.687	0.646	0.604	0.554	0.647	0.206	0.173	0.143	0.18	0.51	0.444	0.326	0.597	0.294	0.249	0.199	0.276	0.724	0.665	0.609	0.694
P5	0.637	0.591	0.595	0.572	0.717	0.717	0.513	0.734	0.192	0.181	0.106	0.182	0.816	0.875	0.41	0.938	0.31	0.3	0.168	0.304	0.778	0.802	0.545	0.825
P6	0.64	0.628	0.591	0.506	0.723	0.661	0.594	0.661	0.194	0.17	0.139	0.151	0.826	0.701	0.597	0.854	0.315	0.274	0.226	0.257	0.797	0.732	0.629	0.711
P7	0.65	0.717	0.801	0.597	0.678	0.691	0.584	0.671	0.182	0.209	0.194	0.168	0.712	0.66	0.312	0.764	0.289	0.318	0.239	0.275	0.744	0.751	0.684	0.726
P8	0.646	0.818	0.757	0.681	0.612	0.578	0.624	0.628	0.155	0.202	0.195	0.17	0.569	0.278	0.458	0.562	0.244	0.234	0.274	0.261	0.653	0.6	0.697	0.685
P9	0.83	0.689	0.634	0.547	0.717	0.685	0.673	0.656	0.311	0.196	0.176	0.155	0.576	0.681	0.722	0.792	0.404	0.304	0.283	0.259	0.823	0.759	0.73	0.718
P10	0.695	0.626	0.768	0.769	0.74	0.721	0.578	0.61	0.219	0.19	0.17	0.193	0.795	0.84	0.34	0.41	0.343	0.31	0.227	0.262	0.809	0.771	0.628	0.701
P11	0.71	0.665	0.657	0.617	0.734	0.761	0.686	0.769	0.223	0.214	0.186	0.202	0.764	0.882	0.722	0.958	0.345	0.345	0.296	0.334	0.81	0.853	0.736	0.869
P12	0.761	0.798	0.871	0.765	0.75	0.669	0.518	0.709	0.257	0.249	0.172	0.243	0.736	0.507	0.076	0.639	0.381	0.334	0.106	0.352	0.839	0.766	0.644	0.793
P13	0.662	0.744	0.626	0.648	0.63	0.592	0.52	0.672	0.166	0.17	0.11	0.179	0.59	0.403	0.389	0.701	0.259	0.239	0.172	0.285	0.651	0.656	0.521	0.726
P14	0.483	0.719	0.75	0.84	0.651	0.692	0.596	0.72	0.146	0.211	0.175	0.327	0.861	0.66	0.403	0.569	0.25	0.319	0.244	0.415	0.766	0.739	0.62	0.794
P15	0.696	0.656	0.781	0.773	0.72	0.63	0.536	0.661	0.212	0.164	0.139	0.225	0.75	0.597	0.229	0.521	0.33	0.258	0.173	0.314	0.788	0.701	0.619	0.732
P16	0.654	0.65	0.557	0.558	0.53	0.574	0.507	0.578	0.117	0.139	0.103	0.13	0.375	0.479	0.444	0.604	0.178	0.215	0.167	0.215	0.554	0.588	0.504	0.613
P17	0.721	0.639	0.502	0.542	0.573	0.593	0.517	0.542	0.151	0.145	0.106	0.116	0.389	0.535	0.535	0.542	0.218	0.228	0.177	0.191	0.631	0.603	0.531	0.558
P18	0.563	0.691	0.623	0.542	0.646	0.637	0.633	0.641	0.154	0.176	0.159	0.15	0.75	0.569	0.646	0.764	0.256	0.269	0.255	0.25	0.696	0.681	0.67	0.707
P19	0.665	0.733	0.681	0.69	0.721	0.741	0.634	0.689	0.201	0.237	0.172	0.198	0.792	0.75	0.576	0.688	0.321	0.36	0.265	0.307	0.791	0.822	0.679	0.747
P20	0.688	0.781	0.618	0.675	0.709	0.681	0.584	0.64	0.205	0.241	0.139	0.173	0.736	0.556	0.542	0.597	0.32	0.336	0.221	0.269	0.773	0.736	0.608	0.666
Avg	**0.664**	**0.689**	**0.676**	**0.641**	**0.667**	**0.655**	**0.578**	**0.659**	**0.187**	**0.188**	**0.149**	**0.182**	**0.671**	**0.613**	**0.454**	**0.681**	**0.290**	**0.283**	**0.213**	**0.281**	**0.731**	**0.714**	**0.620**	**0.716**

Table C.14: Multiband EEGNet-MoE-Transformer: Evaluation metrics for classification model trained on a combination of traditional (clean) and intentionally contaminated RSVP data.

Multiband EEGNet-MoE-Transformer : Train Traditional RSVP + Intentionally Contaminated RSVP																							
Sub	**Accuracy**				**Balanced Accuracy**				**Precision**				**Recall**				**F1 Score**				**ROC-AUC**			
	RSVP	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**			**RSVP**	**IC-RSVP**		
		IC-B	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**		**IC-B**	**IC-H**	**IC-T**
P1	0.543	0.622	0.553	0.541	0.595	0.583	0.539	0.64	0.135	0.139	0.116	0.149	0.66	0.535	0.521	0.764	0.224	0.22	0.189	0.25	0.614	0.607	0.535	0.68
P2	0.634	0.526	0.653	0.599	0.667	0.678	0.687	0.688	0.174	0.159	0.186	0.173	0.708	0.868	0.729	0.799	0.279	0.268	0.296	0.285	0.712	0.728	0.713	0.719
P3	0.606	0.569	0.528	0.641	0.664	0.662	0.54	0.699	0.167	0.16	0.115	0.187	0.736	0.778	0.556	0.771	0.272	0.265	0.19	0.3	0.724	0.722	0.525	0.749
P4	0.824	0.787	0.351	0.752	0.712	0.703	0.584	0.76	0.3	0.257	0.121	0.255	0.573	0.597	0.875	0.771	0.394	0.359	0.212	0.383	0.791	0.768	0.601	0.824
P5	0.75	0.812	0.628	0.649	0.761	0.763	0.605	0.756	0.254	0.307	0.149	0.207	0.774	0.701	0.576	0.889	0.382	0.427	0.236	0.336	0.825	0.822	0.621	0.821
P6	0.665	0.677	0.751	0.773	0.73	0.642	0.596	0.738	0.204	0.174	0.175	0.261	0.812	0.597	0.403	0.694	0.326	0.27	0.244	0.38	0.794	0.666	0.611	0.766
P7	0.693	0.805	0.616	0.71	0.714	0.7	0.635	0.688	0.209	0.272	0.159	0.205	0.74	0.569	0.66	0.66	0.325	0.369	0.256	0.312	0.788	0.752	0.675	0.725
P8	0.572	0.594	0.745	0.528	0.611	0.543	0.618	0.649	0.143	0.119	0.186	0.15	0.66	0.479	0.458	0.799	0.235	0.191	0.265	0.253	0.63	0.532	0.64	0.69
P9	0.789	0.747	0.71	0.754	0.739	0.751	0.681	0.691	0.275	0.248	0.202	0.228	0.677	0.757	0.646	0.611	0.391	0.374	0.308	0.332	0.8	0.797	0.73	0.751
P10	0.74	0.743	0.702	0.646	0.763	0.734	0.597	0.664	0.249	0.24	0.16	0.176	0.792	0.722	0.465	0.688	0.379	0.36	0.238	0.28	0.834	0.783	0.612	0.704
P11	0.751	0.797	0.632	0.783	0.742	0.792	0.65	0.771	0.248	0.302	0.167	0.282	0.729	0.785	0.674	0.757	0.37	0.436	0.268	0.411	0.813	0.855	0.699	0.841
P12	0.902	0.757	0.566	0.741	0.788	0.742	0.62	0.726	0.508	0.252	0.146	0.236	0.646	0.722	0.688	0.708	0.569	0.373	0.241	0.354	0.857	0.797	0.652	0.78
P13	0.713	0.559	0.187	0.676	0.652	0.601	0.53	0.696	0.19	0.138	0.106	0.196	0.576	0.653	0.958	0.722	0.286	0.228	0.191	0.308	0.666	0.618	0.518	0.74
P14	0.792	0.755	0.669	0.722	0.752	0.716	0.705	0.722	0.282	0.24	0.197	0.224	0.701	0.667	0.75	0.722	0.402	0.353	0.312	0.342	0.828	0.774	0.735	0.781
P15	0.789	0.447	0.451	0.648	0.762	0.64	0.608	0.693	0.284	0.14	0.132	0.187	0.729	0.882	0.806	0.75	0.409	0.242	0.227	0.299	0.803	0.695	0.64	0.726
P16	0.451	0.67	0.644	0.674	0.562	0.635	0.549	0.615	0.119	0.17	0.126	0.162	0.701	0.59	0.431	0.542	0.203	0.264	0.195	0.249	0.544	0.623	0.539	0.621
P17	0.523	0.41	0.485	0.526	0.645	0.623	0.529	0.589	0.149	0.133	0.11	0.132	0.799	0.889	0.583	0.667	0.251	0.231	0.185	0.22	0.681	0.639	0.525	0.602
P18	0.658	0.775	0.656	0.724	0.705	0.674	0.642	0.68	0.194	0.234	0.169	0.208	0.764	0.549	0.625	0.625	0.309	0.328	0.267	0.312	0.723	0.712	0.662	0.724
P19	0.743	0.797	0.611	0.674	0.762	0.754	0.648	0.698	0.25	0.288	0.162	0.196	0.785	0.701	0.694	0.729	0.379	0.408	0.263	0.309	0.815	0.808	0.688	0.735
P20	0.699	0.606	0.459	0.568	0.709	0.658	0.595	0.637	0.209	0.165	0.129	0.152	0.722	0.722	0.764	0.722	0.324	0.268	0.22	0.251	0.747	0.702	0.619	0.696
Avg	**0.691**	**0.672**	**0.579**	**0.666**	**0.701**	**0.679**	**0.607**	**0.690**	**0.227**	**0.206**	**0.150**	**0.198**	**0.714**	**0.688**	**0.643**	**0.719**	**0.335**	**0.311**	**0.240**	**0.308**	**0.749**	**0.720**	**0.627**	**0.734**