“Neuro-Shield: A Robust Six-Tier Defense-in-Depth Framework for Secure Brain-Computer Interface Operations”

by

Md. Mahamudur Rahman
ID: CSE2001019067

Jannatul Ferdaous Mim
ID: CSE2101022096

Sabina Chowdhury Chua
ID: CSE2202026004

Shorna Akter
ID: CSE2202026100

Supervised by
Mohammad Naderuzzaman

Submitted in partial fulfillment of the requirements for the degree of
Bachelor of Science in Computer Science and Engineering

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SONARGAON UNIVERSITY (SU)

May 2026

APPROVAL

The thesis titled "**Neuro-Shield: A Robust Six-Tier Defense-in-Depth Framework for Secure Brain-Computer Interface Operations**" submitted by Md. Mahamudur Rahman (CSE2001019067), Jannatul Ferdaous Mim (CSE2101022096), Sabina Chowdhury Chua (CSE2202026004) and Shorna Akter (CSE2202026100) to the Department of Computer Science and Engineering, Sonargaon University (SU), has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering and approved as to its style and contents.

Board of Examiners

Mohammad Naderuzzaman
Associate Professor,
Department of Computer Science and Engineering
Sonargaon University (SU)

Supervisor

(Examiner Name and Signature)
Department of Computer Science and Engineering
Sonargaon University (SU)

Examiner 1

(Examiner Name and Signature)
Department of Computer Science and Engineering
Sonargaon University (SU)

Examiner 2

(Examiner Name and Signature)
Department of Computer Science and Engineering
Sonargaon University (SU)

Examiner 3

DECLARATION

We, hereby, declare that the work presented in this report is the outcome of the investigation performed by us under the supervision of **Mohammad Naderuzzaman, Associate Professor,** Department of Computer Science and Engineering, Sonargaon University, Dhaka, Bangladesh. We reaffirm that no part of this thesis has been or is being submitted elsewhere for the award of any degree or diploma.

Countersigned

(Mohammad Naderuzzaman)
Supervisor

Signature

Md. Mahamudur Rahman
ID: CSE2001019067

Jannatul Ferdaous Mim
ID: CSE2101022096

Sabina Chowdhury Chua
ID: CSE2202026004

Shorna Akter
ID: CSE2202026100

ABSTRACT

Brain–Computer Interfaces (BCIs) have emerged as a transformative technology that enables direct communication between the human brain and external systems, opening new possibilities in healthcare, neuroprosthetics, and intelligent assistive systems. However, the integration of highly sensitive neural signals into real-time computational pipelines introduces critical security and privacy challenges, including passive eavesdropping, adversarial signal injection, and hardware-level spoofing. These threats can compromise both system integrity and user safety, while conventional security mechanisms remain inadequate due to strict latency requirements and the inherent variability of biological signals. To address these limitations, this book introduces Neuro-Shield, a robust six-tier defense-in-depth architecture specifically designed for secure Brain–Computer Interface operations. Unlike traditional approaches that treat security as an external layer, Neuro-Shield embeds protection mechanisms directly within the signal processing pipeline. The framework combines hardware-based trust through Physical Unclonable Functions (PUFs), secure and high-fidelity neural signal acquisition, and adaptive digital signal processing for artifact mitigation. It also incorporates a lightweight one-dimensional convolutional neural network (1D-CNN) to enable continuous neural authentication. In addition, privacy-preserving techniques and deception-based strategies are used to protect sensitive data and enhance resilience against advanced adversarial threats. A secure command validation mechanism, supported by fail-safe controls such as a hardware-level kill-switch, ensures safe and reliable system execution. The architecture is implemented on an edge-based platform and evaluated using EEG datasets under simulated adversarial conditions. Evaluation results show that Neuro-Shield significantly improves detection accuracy while maintaining low-latency performance suitable for real-time applications. The framework achieves a detection accuracy of 98.6%, reduces the false acceptance rate to 0.34%, and maintains an end-to-end latency of 76.3 ms, demonstrating an effective balance between security and performance. This book provides a comprehensive exploration of secure BCI system design, covering threat modeling, architectural development, implementation strategies, and performance evaluation. Overall, the presented approach establishes a practical foundation for next-generation BCI systems, where security, privacy, and user safety are treated as integral components of system design rather than optional add-ons.

ACKNOWLEDGMENT

At the very beginning, we would like to express our deepest gratitude to the Almighty Allah for giving us the ability and the strength to finish the task successfully within the scheduled time.

We are fortunate to have the kind association as well as the supervision of **Mohammad Naderuzzaman**, Associate Professor, Department of Computer Science and Engineering, Sonargaon University, whose heartfelt and valuable support, guidance, and direction acted as an essential resource in carrying out our project.

We would also like to express our sincere gratitude to **Bulbul Ahamed**, Associate Professor and Head, for his valuable support, guidance, and encouragement throughout this work.

We are also thankful to all our teachers throughout our academic journey for exposing us to the beauty of learning.

Finally, we express our deepest gratitude and love to our parents for their unwavering support, encouragement, and endless love.

LIST OF ABBREVIATIONS

ADC – Analog-to-Digital Converter

AI – Artificial Intelligence

BCI – Brain–Computer Interface

CNN – Convolutional Neural Network

1D-CNN – One-Dimensional Convolutional Neural Network

CPU – Central Processing Unit

DL – Deep Learning

DSP – Digital Signal Processing

ECG – Electrocardiography

EEG – Electroencephalography

EMG – Electromyography

EOG – Electrooculography

F1–Score – Harmonic Mean of Precision and Recall

FN – False Negative

FP – False Positive

GPU – Graphics Processing Unit

HMI – Human–Machine Interaction

Hz – Hertz

IoT – Internet of Things

LDA – Linear Discriminant Analysis

MI – Motor Imagery

ML – Machine Learning

ms – Milliseconds

OS – Operating System

PUF – Physical Unclonable Function

RAM – Random Access Memory

RNN – Recurrent Neural Network

SVM – Support Vector Machine

SoM – System on Module

TN – True Negative

TP – True Positive

USB – Universal Serial Bus

Wi-Fi – Wireless Fidelity

ZT – Zero-Trust (Never Trust, Always Verify)

µV – Microvolts

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION TO BRAIN-COMPUTER INTERFACES

1.1 Introduction

Brain–Computer Interfaces (BCIs) are transforming how humans interact with machines. Unlike traditional systems that rely on physical movement or speech, BCIs establish a direct communication pathway between the human brain and external devices [1]. Fundamentally this means a user can control a system using only their thoughts.

This capability is especially significant for individuals with neurological impairments, such as paralysis, spinal cord injuries, or neurodegenerative disorders. For these users, conventional interaction methods may be limited or entirely unavailable. BCIs offer an alternative by translating neural activity into commands that machines can understand and execute [1], [16].

Over the past decade, advances in signal processing and machine learning have significantly improved the performance and practicality of BCI systems. In particular, deep learning models—such as convolutional neural networks (CNNs)—have demonstrated strong capabilities in decoding complex EEG signals and identifying user intent with high accuracy [24], [25]. As a result, BCIs are increasingly used in assistive robotics, neuroprosthetics, rehabilitation, and intelligent human–machine interaction [18], [19].

However, as BCI systems move from controlled laboratory settings into real-world applications, new challenges arise. One of the most critical concerns is security. Neural signals are inherently sensitive and can reveal personal information such as cognitive states, emotions, and behavioral patterns. Unauthorized access to such data may lead to serious privacy violations and misuse [11], [12].

At the same time, BCI systems must operate in real time. Even small delays can negatively affect system performance and user experience. Traditional security mechanisms—such as heavy encryption or delayed authentication—are often unsuitable in this context because they introduce latency [17], [32].

This creates a fundamental challenge: how to ensure strong security while maintaining real-time responsiveness. Addressing this challenge requires lightweight, efficient, and integrated security mechanisms that can operate seamlessly within the BCI processing pipeline.

1.2 Objectives of This Thesis

This book aims to provide a clear and practical understanding of security in Brain–Computer Interface systems. Rather than treating security as an external add-on, it adopts a security-by-design approach, where protection mechanisms are embedded directly into the system architecture.

The key objectives of this book are:

- To analyze the evolving threat landscape in BCI systems, including both cyber and physical attack vectors [21]
- To examine the limitations of traditional security approaches in real-time neural signal processing environments [1], [17]
- To introduce **Neuro-Shield**, a multi-layered defense architecture designed specifically for secure BCI operations [19], [28]
- To demonstrate how hardware-based trust and machine learning techniques can enable continuous neural authentication [27], [31]
- To present a practical implementation of secure BCI systems using edge computing platforms [32]

In essence, this book addresses a central question:
How can we design BCI systems that are not only intelligent and responsive, but also secure and trustworthy?

1.3 Overview of Brain–Computer Interfaces

A Brain–Computer Interface (BCI) acts as a bridge between neural activity and machine control. It captures brain signals, processes them, and converts them into actionable commands [16], [18].

Among various neuroimaging techniques, Electroencephalography (EEG) is the most widely used for real-time BCI applications due to its non-invasive nature, portability, and high temporal resolution [18].

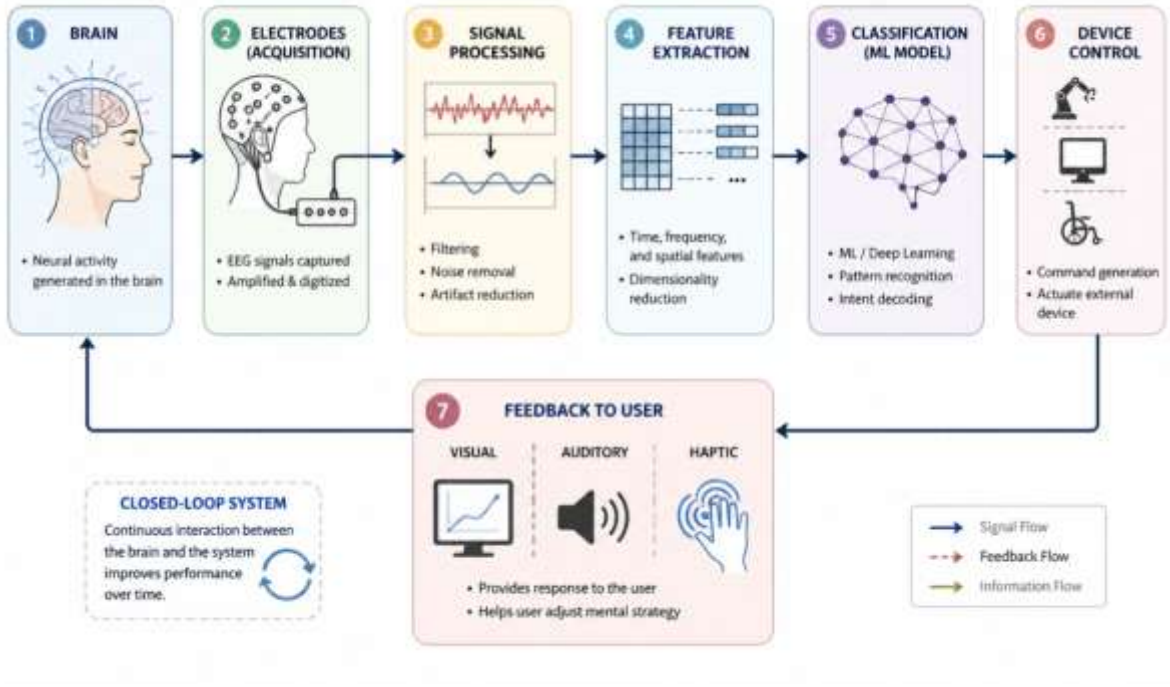

Figure 1.1: End-to-end closed-loop BCI workflow showing signal acquisition, processing, classification, and device control.

As illustrated in Figure 1.1, a typical BCI system operates through a structured pipeline. To better understand this process, each stage can be described as follows:

- **Signal Acquisition:**
  Neural activity is captured using EEG electrodes placed on the scalp. These electrodes detect very low-amplitude electrical signals, which are then amplified and digitized for further processing [18].

- **Signal Processing (Preprocessing):**
  The raw signals are often noisy and affected by artifacts such as eye movements (EOG), muscle activity (EMG), and environmental interference. Filtering techniques are applied to remove noise and improve signal quality [16].

- **Feature Extraction:**
  Once cleaned, the signals are transformed into meaningful representations using statistical and signal processing methods, such as time-frequency analysis and spatial filtering [15].

- **Classification (Decoding):**
  Machine learning models analyze these features to determine the user's intent. Deep learning approaches, particularly CNNs, have shown strong performance in this stage.
- **Device Control:**
  The decoded output is converted into control signals that operate external devices, such as robotic limbs, wheelchairs, or computer interfaces [19].

- **Feedback Loop:**
  The system provides real-time feedback (visual, auditory, or haptic), allowing users to adjust their mental strategies and improve performance over time.

Together, these stages form a closed-loop system that continuously adapts to the user, making BCIs both interactive and intelligent.

1.4 Security Challenges in BCIs

Despite their advantages, BCI systems introduce several critical security challenges due to their reliance on sensitive neural data and strict real-time requirements.

In real-world scenarios, these vulnerabilities are not merely theoretical—they can directly impact system behavior and user safety. As shown in Figure 1.2, different stages of the BCI pipeline are exposed to various cyber threats.

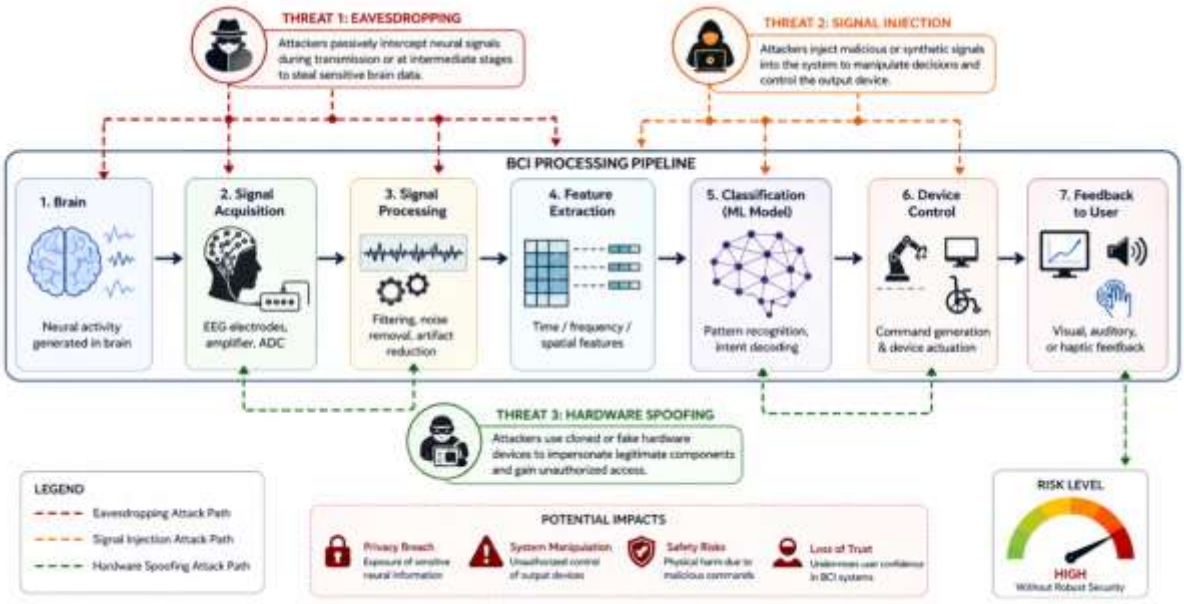

Figure 1.2: Cyber threat mapping in the BCI pipeline.

The major challenges include:

- Sensitive Neural Data:

  Neural signals contain deeply personal information, including thoughts, emotions, and behavioral patterns. Unauthorized access can result in severe privacy breaches [12].

- Eavesdropping Attacks:

  Wireless transmission of EEG signals makes them vulnerable to interception, allowing attackers to extract sensitive information without detection [12].

- Signal Injection Attacks:

  Adversaries may introduce malicious or replayed signals into the system, leading to incorrect interpretation and unintended actions [10].

- Hardware Spoofing:

  Attackers can replicate or tamper with hardware components to impersonate legitimate devices, compromising system trust and integrity [31].

- Adversarial Attacks:

  Machine learning models used in BCIs are susceptible to adversarial perturbations—small, carefully crafted changes in input data that can mislead the system [3], [4].

- Latency vs. Security Trade-off:
  Strong security mechanisms often introduce computational overhead, which conflicts with the real-time requirements of BCI systems [32].
- Safety-Critical Risks:
  Since BCIs directly control physical systems, any security breach can lead to immediate and potentially harmful consequences.

To summarize these challenges more clearly, Table 1.1 presents their characteristics and impact.

The major cyber security threats in BCI systems can be summarized as follows:

Threat Type	Attack Description	Impact on System	Severity Level
Eavesdropping	Unauthorized interception of neural data during transmission.	Compromises user privacy and exposes sensitive data.	High
Signal Injection	Malicious insertion or replay of synthetic EEG signals into the processing pipeline.	Leads to incorrect decisions and potential system hijack.	Critical
Hardware Spoofing	Use of cloned or tampered hardware to impersonate legitimate BCI components.	Enables unauthorized system access and manipulation.	High

Table 1.1: Cyber security threats in BCI systems and their impacts.

These challenges highlight an important insight:

Security in BCI systems cannot be treated as an optional feature.

Instead, it must be integrated into every stage of the system—from signal acquisition to final execution. A reactive approach is not sufficient in such safety-critical environments. Modern BCI systems require proactive, multi-layered defense strategies that ensure security without compromising performance.

1.5 Organization of the Thesis

This book is structured to guide the reader from fundamental concepts to advanced implementation and evaluation of secure BCI systems.

- **Chapter 2:** BCI Threat Landscape Provides a detailed analysis of attack vectors, adversary models, and system vulnerabilities [3], [4], [22]
- **Chapter 3:** Hardware and Root-of-Trust Mechanisms Explores secure signal acquisition and hardware-based authentication using PUFs [31], [32]
- **Chapter 4:** Signal Processing and Neural Authentication Discusses advanced filtering techniques and machine learning-based neural fingerprinting [25], [29]
- **Chapter 5:** Privacy, Deception, and Safe Execution Covers data protection methods such as differential privacy and post-quantum cryptography [27], [32]
- **Chapter 6:** Implementation and Experimental Evaluation Presents real-world deployment and performance evaluation of the Neuro-Shield framework [20], [32]
- **Chapter 7:** Conclusion and Future Directions Summarizes key findings and outlines future research opportunities in secure neurotechnology [18]

In summary, the growing adoption of Brain–Computer Interfaces demands a fundamental shift in how security is approached. Rather than being an afterthought, security must be embedded directly into system design. This principle forms the foundation of the Neuro-Shield architecture, which is explored in detail throughout this book.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The field of Brain–Computer Interfaces (BCIs) has evolved rapidly over the past decade, driven by advances in neural signal processing, artificial intelligence, and embedded systems. What began as experimental research has now developed into practical systems used in healthcare, rehabilitation, and intelligent human–machine interaction.

However, this transition from controlled laboratory environments to real-world deployment introduces new challenges. Unlike traditional computing systems, BCIs rely on highly sensitive neural data and must operate under strict real-time constraints. This combination makes them both complex and vulnerable to security and reliability issues [17], [18].

This chapter explores existing research in BCI systems, focusing on system architectures, EEG-based authentication, machine learning techniques, and security challenges. More importantly, it critically examines the limitations of current approaches and identifies key gaps that motivate the development of a more secure and adaptive framework.

2.2 Overview of Existing BCI Systems

A typical BCI system converts brain activity into machine-executable commands through a structured pipeline that includes signal acquisition, preprocessing, feature extraction, classification, and device control [17], [19].

Early BCI systems were primarily developed for medical applications, particularly to assist patients with severe motor impairments. Technologies such as P300-based spellers and motor imagery systems enabled users to communicate and control devices without physical movement [10], [17].

In recent years, the integration of machine learning and real-time processing techniques has significantly improved system performance. The development of wearable EEG devices and

edge computing platforms has further expanded the use of BCIs into areas such as smart healthcare, gaming, and adaptive interfaces [18], [32].

However, a critical limitation remains. Most existing systems prioritize performance metrics such as accuracy and latency, often overlooking equally important aspects such as data security, privacy preservation, and resistance to adversarial manipulation.

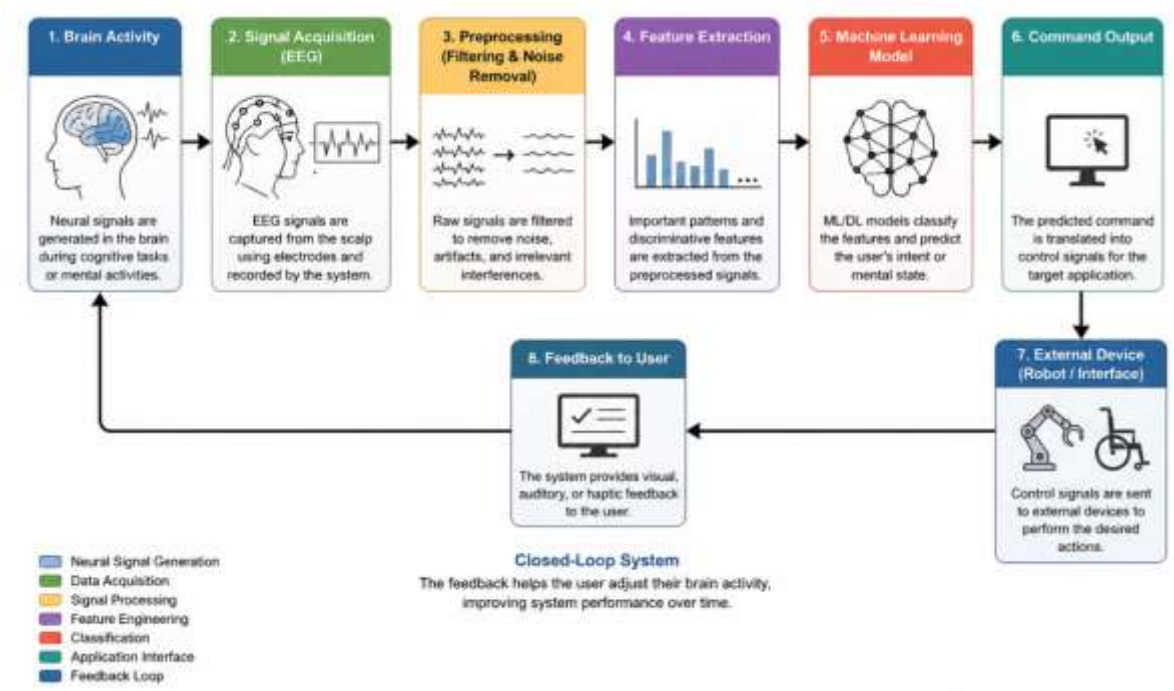

Figure 2.1: Conceptual BCI architecture showing signal acquisition, processing, decoding, and feedback.

As illustrated in Figure 2.1, BCI systems operate as closed-loop architectures, where continuous feedback improves system performance over time. However, this closed-loop design also introduces additional points of vulnerability, as feedback channels can be exploited by attackers to manipulate system behavior.

2.3 EEG-Based Authentication Techniques

EEG-based authentication has emerged as a promising alternative to traditional biometric systems due to its intrinsic uniqueness and resistance to spoofing. Unlike fingerprints or facial recognition, EEG signals originate from internal neural activity, making them extremely difficult to replicate or forge [12].

In most existing approaches, EEG signals are recorded while users perform specific cognitive tasks, such as visual stimulation or motor imagery, and are then converted into distinctive features that serve as biometric identifiers [15], [29]. These features are then used to train machine learning models for user identification.

Recent research has shifted towards **continuous authentication**, where user identity is verified dynamically throughout system operation rather than at a single login point. This significantly enhances security, especially in safety-critical applications [27], [31].

However, EEG-based authentication faces several challenges:

- High variability of neural signals
- Sensitivity to environmental noise
- Limited scalability across multiple users

These issues highlight the need for more robust and adaptive authentication mechanisms.

2.4 Machine Learning in EEG Analysis

Machine learning plays a central role in enabling BCI systems to interpret complex neural signals. Traditional algorithms, such as Support Vector Machines (SVM) and Linear Discriminant Analysis (LDA), have been widely used due to their simplicity and efficiency.

More recently, deep learning techniques, especially Convolutional Neural Networks (CNNs)—have significantly improved classification performance by automatically learning hierarchical features from raw EEG data [24], [25].

Advanced models, including recurrent neural networks (RNNs) and hybrid architectures, enhance performance by capturing temporal dependencies in neural signals. These approaches enable more adaptive and context-aware systems.

However, this increased reliance on machine learning also introduces new vulnerabilities. Deep learning models are susceptible to adversarial attacks, where small perturbations in input data can lead to incorrect predictions. This raises serious concerns about the reliability and security of BCI systems in real-world environments [3], [4].

2.5 Security Challenges in BCI Systems

As BCI systems become increasingly integrated with wireless networks and Internet of Things (IoT) infrastructures, they are exposed to a wide range of cyber threats [18]. In real-world deployments, these threats are not merely theoretical; they can directly influence system behavior and compromise user safety.

The major security challenges include:

- **Eavesdropping Attacks:**
  Unauthorized interception of neural signals during transmission, which can expose sensitive cognitive information and lead to privacy breaches [12].
- **Signal Injection Attacks:**
  The introduction of malicious or replayed signals into the system, resulting in incorrect interpretations and potentially unsafe actions [10].
- **Adversarial Attacks:**
  Exploitation of vulnerabilities in machine learning models through carefully crafted perturbations, causing misclassification and unreliable system behavior [3], [4].
- **Hardware-Level Attacks:**
  Tampering with or spoofing EEG devices to gain unauthorized access and compromise system integrity [31].

These challenges clearly demonstrate that BCI systems must be treated as security-critical infrastructures rather than conventional computing systems.

More importantly, these risks highlight that security in BCI systems is not merely a technical requirement, but a fundamental necessity for ensuring safe, reliable, and trustworthy deployment.

2.6 Limitations of Existing Approaches

To better understand these issues, Table 2.1 summarizes the key limitations of existing BCI approaches.

Aspect	Current Limitation	Impact
Security	Lack of integrated security mechanisms	Vulnerable to attacks
Signal Quality	High sensitivity to noise and artifacts	Reduced accuracy
Generalization	Poor cross-user adaptability	Limited scalability
Latency	Security adds computational overhead	Affects real-time performance
Robustness	Vulnerable to adversarial manipulation	Unreliable system behavior

Table 2.1: Summary of Limitations in Existing Brain–Computer Interface (BCI) Approaches

Despite significant progress, existing approaches remain fragmented. Many focus on improving performance, while others address security in isolation. Very few solutions successfully integrate both aspects into a unified framework.

2.7 Summary and Research Gap

This chapter reviewed the current state of BCI research, including system architectures, authentication techniques, machine learning approaches, and security challenges.

While these developments have improved system performance and usability, they still lack robustness, scalability, and integrated security mechanisms. More importantly, existing solutions fail to provide a unified framework that simultaneously ensures security, real-time performance, and resilience against adversarial threats.

This gap forms the foundation for the **Neuro-Shield architecture**, which is introduced in the following chapters.

CHAPTER 3

THREAT MODEL IN BCI SYSTEMS

3.1 Introduction

As Brain–Computer Interface (BCI) systems continue to advance and transition from controlled laboratory environments to real-world applications, their exposure to diverse security threats has become increasingly significant. Unlike traditional computing systems, BCIs operate on highly sensitive neural data and often interact directly with physical devices such as prosthetics, wheelchairs, and assistive robotics. This tight coupling between cognitive signals and physical actions introduces unique vulnerabilities, making BCI systems susceptible to both cyber and physical attacks.

A threat model helps systematically identify potential adversaries, define attack surfaces, and analyze system vulnerabilities. In BCI systems, this process becomes particularly challenging due to the integration of biological signal acquisition, machine learning-based decision-making, and wireless communication infrastructures [18], [22]. As a result, each stage of the BCI pipeline—from signal acquisition to device control—presents distinct opportunities for exploitation, significantly expanding the overall attack surface.

Furthermore, the real-time nature of BCI operation imposes strict constraints on computational overhead, which limits the applicability of conventional security mechanisms. At the same time, the ***sensitive and personal nature of neural data*** raises serious concerns regarding privacy, data integrity, and user safety. Unauthorized access or manipulation of such data can lead not only to system malfunction but also to unintended physical actions and potential harm to users.

This chapter presents a comprehensive and structured analysis of major threat vectors in BCI systems, focusing on both **passive attacks** (e.g., eavesdropping) and **active attacks** (e.g., signal injection and hardware spoofing) that target different stages of the signal processing pipeline. By systematically examining the mechanisms, impacts, and limitations of existing defenses, this chapter establishes a clear understanding of the BCI threat landscape. This analysis serves as a foundation for the development of secure and resilient architectures,

leading to the proposed **Neuro-Shield framework** introduced in the subsequent chapter. In simple terms, understanding these threats is essential for building BCI systems that are not only intelligent, but also secure and trustworthy.

3.2 Passive Eavesdropping

Passive eavesdropping refers to the silent interception of data as it travels through the BCI system.

In this type of attack, the adversary does not modify or disrupt the signal but simply listens in. Since EEG signals can contain highly sensitive neural information, this poses a serious privacy risk. Attackers may exploit weak or unencrypted communication channels to capture these signals. Because there is no direct interaction with the system, detecting such attacks becomes very difficult.

As a result, passive eavesdropping can compromise user confidentiality without leaving obvious traces.

3.2.1 Overview

Passive eavesdropping represents one of the most fundamental and critical security threats in Brain–Computer Interface (BCI) systems. It involves the unauthorized interception of neural signals during transmission, typically without modifying or disrupting the data flow. In modern BCI architecture, neural signals acquired through EEG devices are often transmitted via wireless communication channels to processing units or external systems. This reliance on wireless communication makes BCI systems inherently vulnerable to interception by malicious entities [12].

Unlike active attacks, where adversaries manipulate or inject data into the system, passive eavesdropping operates silently in the background. Attackers can capture transmitted signals using wireless sniffing tools or compromised network nodes, enabling them to access sensitive neural information without triggering any immediate system anomalies. This characteristic makes eavesdropping attacks inherently **stealthy** and difficult to detect using conventional monitoring techniques.

The risks associated with passive eavesdropping extend beyond simple data leakage. Neural signals can reveal highly sensitive information, including cognitive states, user intentions, and emotional responses. Unauthorized access to such data raises serious concerns regarding privacy, user profiling, and potential misuse in surveillance or behavioral analysis.

Furthermore, the persistent nature of eavesdropping allows attackers to continuously monitor system activity over time, making it possible to build detailed profiles of user behavior. As a result, passive eavesdropping is not only a privacy threat but also a significant risk to the overall trustworthiness and security of BCI systems.

3.2.2 Attack Mechanism

In a typical Brain–Computer Interface (BCI) system, neural signals captured by EEG devices are transmitted—often wirelessly—to processing units, edge devices, or cloud-based platforms for further analysis. During this transmission phase, the communication channel becomes a critical attack surface that can be exploited by adversaries.

Attackers can intercept these signals using wireless sniffing tools, software-defined radios, or by compromising intermediate network nodes within the communication infrastructure. Since many BCI systems rely on standard wireless protocols, such as Bluetooth or Wi-Fi, they inherit common vulnerabilities associated with these technologies. As a result, neural data can be captured without requiring direct access to the physical device.

Once intercepted, the acquired signals can be subjected to offline or real-time analysis. By applying signal processing and machine learning techniques, adversaries may extract sensitive information such as user intentions, emotional states, cognitive workload, or behavioral patterns. This capability transforms passive eavesdropping from a simple data interception attack into a powerful tool for unauthorized inference and profiling.

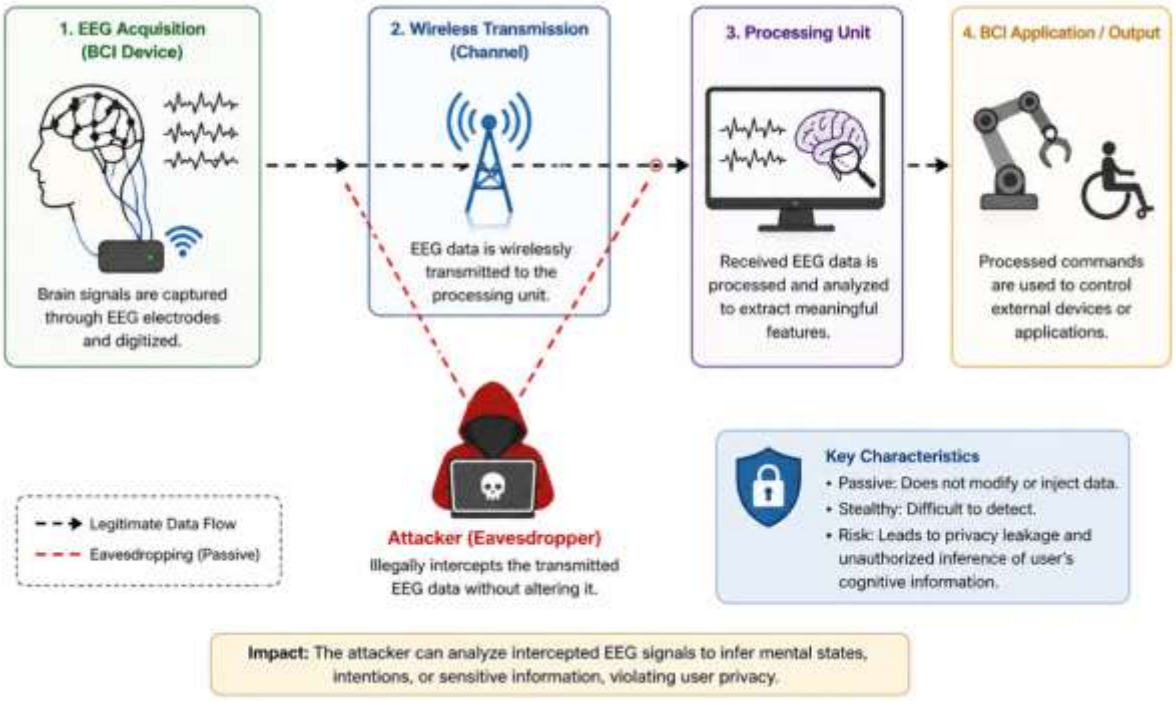

Figure 3.1: Illustration of Passive Eavesdropping in a Brain–Computer Interface System

Figure 3.1 illustrates how neural signals transmitted from the EEG acquisition device to the processing unit can be intercepted by an adversary within the communication channel. The diagram highlights the vulnerability of wireless transmission and emphasizes that the attacker passively captures data without altering system operation.

3.2.3 Impact on System Security

Passive eavesdropping poses significant risks to both the security and privacy of BCI systems. The major impacts are outlined below:

- **Exposure of Sensitive Neural Data:**

  Intercepted EEG signals may contain detailed information about the user's cognitive and emotional states. Unauthorized access to such data can lead to serious privacy breaches and loss of confidentiality.

- **Violation of User Privacy:**

Neural data is inherently personal, and its exposure can reveal information that users may not consciously intend to share. This raises ethical and legal concerns regarding data protection and user consent.

- **Risk of Behavioral Profiling:**

  Continuous monitoring of neural signals allows attackers to build detailed behavioral and cognitive profiles of users. Such profiling can be exploited for targeted manipulation, prediction of user behavior, or unauthorized decision-making.

- **Potential Misuse in Surveillance Systems:**

  Extracted neural information may be used in surveillance or monitoring applications without user knowledge, leading to significant societal and ethical implications.

The implications of passive eavesdropping extend far beyond simple data theft. Since neural signals can reveal deeply personal and subconscious information, their compromise can undermine user trust, system integrity, and the overall acceptance of BCI technologies.

3.2.4 Limitations of Existing Defenses and Future Direction

Existing defense mechanisms against passive eavesdropping in Brain–Computer Interface (BCI) systems predominantly rely on encryption and secure communication protocols. While such approaches are effective in conventional computing environments, they introduce significant challenges when applied to BCI systems due to strict real-time constraints and the limited computational capabilities of EEG-based devices.

In particular, the implementation of strong cryptographic algorithms often introduces additional latency, which can degrade system responsiveness and negatively impact real-time interaction. Moreover, many wearable and portable BCI devices lack the necessary processing power and energy efficiency to support advanced encryption techniques. Consequently, existing solutions frequently fail to achieve an optimal balance between security and system performance.

Another critical limitation arises from the inherently **stealthy** nature of passive eavesdropping attacks. Since adversaries do not alter or inject signals into the system, these

attacks remain difficult to detect using conventional monitoring or anomaly detection techniques. This invisibility significantly increases the risk of undetected data leakage and unauthorized inference of sensitive neural information.

To overcome these limitations, future BCI systems must adopt integrated and lightweight security mechanisms that operate directly within the signal processing pipeline. In this context, the proposed **Neuro-Shield framework** introduces a hybrid approach that combines **real-time anomaly detection** with secure signal validation. By embedding security within the processing stages, Neuro-Shield aims to mitigate ***stealthy*** data interception while minimizing latency overhead. This approach ensures a balanced trade-off between privacy preservation and real-time system performance, making it suitable for next-generation BCI applications.

While passive attacks primarily focus on intercepting neural data, more advanced threats actively manipulate system inputs to influence system behavior.

3.3 Signal Injection Attacks

Signal injection attacks occur when an attacker deliberately introduces false or manipulated signals into the BCI system.

These malicious signals can mimic real neural activity and interfere with normal processing. As a result, the system may misinterpret the input and generate incorrect outputs or actions. This can be particularly dangerous in applications like assistive devices or medical systems. Attackers typically exploit weak points in signal acquisition or communication channels. Because the injected signals appear valid, detecting such attacks can be challenging without proper validation mechanisms.

3.3.1 Overview

Signal injection attacks represent a critical class of **active attacks** in Brain–Computer Interface (BCI) systems, where adversaries deliberately introduce **malicious or synthetic neural signals** into the signal processing pipeline. Unlike passive threats, these attacks directly manipulate the input data, thereby influencing the system's decision-making process.

The primary objective of signal injection is to alter the interpretation of neural activity, leading to incorrect classification outcomes and unintended system behavior. Since BCI systems rely heavily on accurate decoding of EEG signals, even minor perturbations or carefully crafted inputs can significantly impact system performance. This makes signal injection attacks particularly dangerous in applications where precise control and reliability are essential [10].

Moreover, the integration of machine learning models within BCI systems further amplifies this vulnerability. These models often lack robustness against adversarial inputs, allowing attackers to exploit their sensitivity and induce incorrect predictions without raising immediate suspicion.

3.3.2 Attack Mechanism

Signal injection attacks can be executed through multiple entry points within the BCI system architecture. Attackers may exploit compromised EEG sensors, intercept and manipulate wireless communication channels, or gain access to software interfaces responsible for signal processing. Each of these points provides an opportunity to introduce falsified data into the system.

In many cases, the injected signals are designed to closely resemble legitimate EEG patterns. By mimicking the statistical and temporal characteristics of real neural activity, these ***synthetic signals*** can bypass conventional filtering and preprocessing mechanisms. As a result, the system treats them as valid inputs and processes them through subsequent stages of feature extraction and classification.

Once the malicious signals propagate through the pipeline, the machine learning model interprets them as genuine user intent. This leads to incorrect classification outcomes and ultimately results in unintended or manipulated device control. In continuous operation scenarios, repeated injections of such signals can disrupt the stability of the system and degrade overall performance.

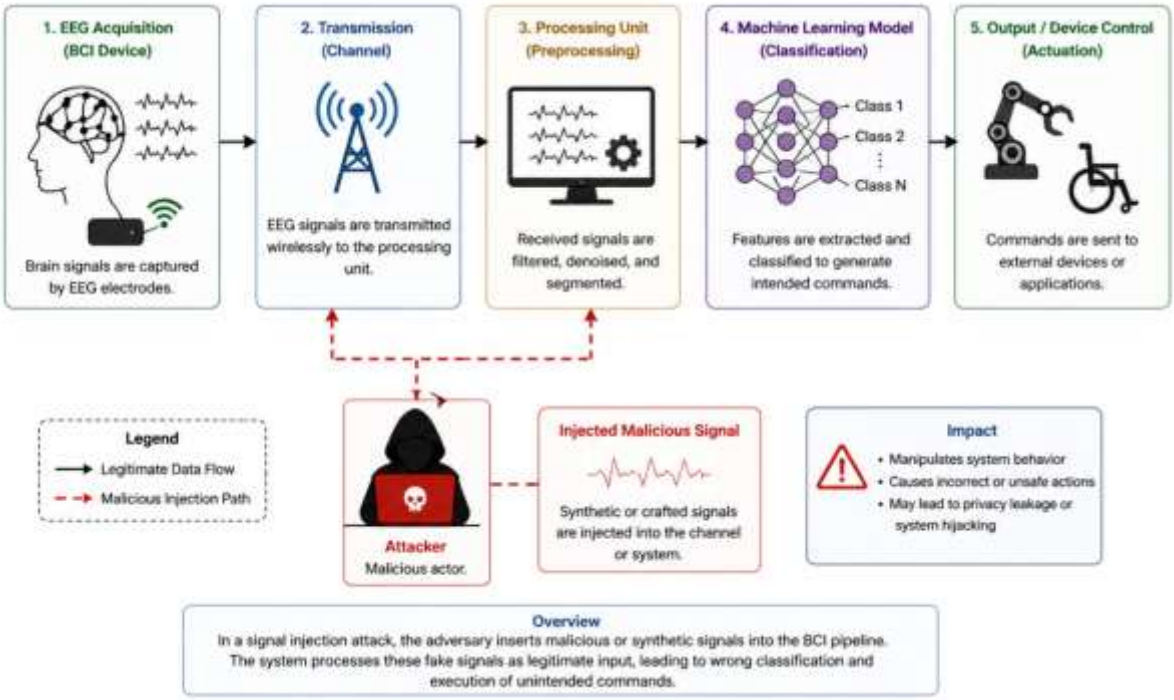

Figure 3.2: Signal Injection Attack in a Brain–Computer Interface Pipeline

Figure 3.2 shows how an adversary can inject synthetic neural signals at various stages of the BCI pipeline, particularly during signal acquisition and transmission. These manipulated signals pass through preprocessing and classification modules, leading to incorrect interpretations and unintended device actions. This demonstrates the importance of integrating signal validation and continuous authentication mechanisms within the processing pipeline.

3.3.3 Impact on System Operation

Signal injection attacks can have severe consequences on the functionality, reliability, and safety of Brain–Computer Interface (BCI) systems. By introducing ***synthetic or manipulated neural signals*** into the processing pipeline, adversaries can directly influence system behavior and disrupt normal operation. The major impacts are discussed below:

- **Incorrect Command Execution:**

  Injected signals may be interpreted as legitimate neural inputs by the system, leading to incorrect classification and unintended command generation. This can result in the

execution of actions that do not reflect the user's actual intent, thereby reducing system accuracy and reliability.

- **System Hijacking:**

  In more advanced scenarios, attackers can continuously inject crafted signals to gain control over the system. This allows them to override genuine user inputs and manipulate the behavior of connected devices, effectively hijacking the entire BCI system.

- **Loss of User Control:**

  Signal injection disrupts the closed-loop interaction between the user and the system. As a result, users may lose the ability to control devices accurately, leading to frustration, reduced usability, and diminished trust in the system.

- **Safety Risks in Critical Applications:**

  In safety-critical applications such as prosthetic limbs, wheelchairs, or assistive robotics, incorrect outputs can lead to hazardous situations. For example, unintended movements or delayed responses may pose physical risks to the user, highlighting the critical importance of securing BCI systems against such attacks.

Overall, signal injection attacks not only compromise system performance but also introduce significant risks to user safety and system integrity. These impacts emphasize the need for robust detection and mitigation mechanisms within the BCI processing pipeline. These risks highlight that signal injection attacks are not only a threat to system accuracy, but also a serious concern for user safety in real-world BCI applications.

3.4 Hardware Spoofing

Hardware spoofing occurs when an attacker uses a fake or compromised device to impersonate a legitimate component in the BCI system. This malicious device can establish communication and appear trusted by the system. As a result, it may send false data or gain unauthorized access to sensitive operations. Such attacks often exploit weak or missing hardware authentication mechanisms.

Because the device appears genuine, detecting the intrusion can be difficult. This can lead to serious risks, including system manipulation and loss of trust in the overall system integrity.

3.4.1 Overview

Hardware spoofing represents a significant **physical-layer threat** in Brain–Computer Interface (BCI) systems, where adversaries attempt to replicate, impersonate, or tamper with legitimate hardware components to gain unauthorized access. In the context of BCIs, this typically involves cloning EEG acquisition devices or modifying firmware to bypass authentication and trust verification mechanisms [31]. In simple terms, hardware spoofing allows an attacker to appear as a legitimate device within the system.

Unlike software-based attacks, hardware spoofing directly targets the trust foundation of the system. Since BCI systems rely on accurate and authenticated signal acquisition, any compromise at the hardware level can undermine the entire processing pipeline. This makes hardware spoofing particularly dangerous, as it can enable attackers to inject falsified data while appearing as legitimate system components.

3.4.2 Attack Mechanism

Hardware spoofing attacks can be executed through multiple approaches. Attackers may develop counterfeit devices that closely mimic the behavior and communication patterns of legitimate EEG hardware. These spoofed devices can then establish connections with the BCI system by exploiting weak or absent hardware authentication protocols.

Alternatively, adversaries may tamper with existing devices by modifying firmware or embedding malicious components. This allows them to manipulate the data being transmitted, inject falsified signals, or alter device behavior without being detected.

Once a spoofed or compromised device is integrated into the system, it is treated as a trusted entity. This enables attackers to continuously inject manipulated data into the processing pipeline, bypassing conventional software-level defenses and compromising system integrity

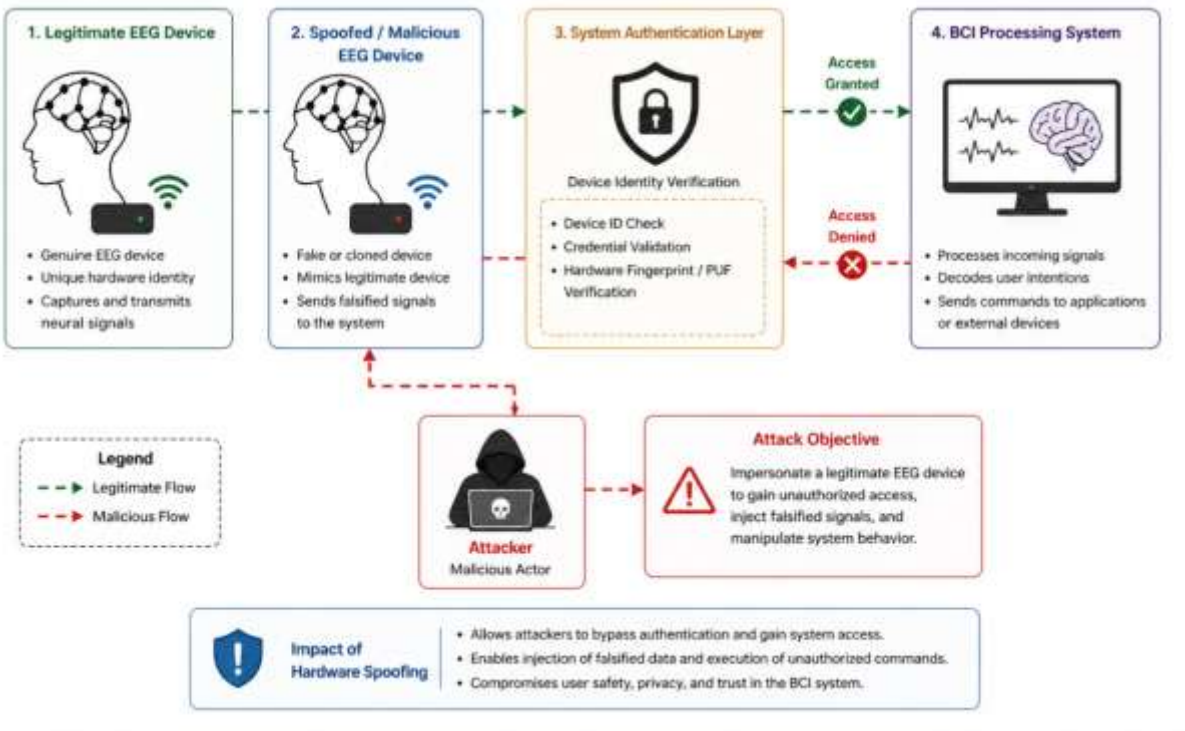

Figure 3.3: Hardware Spoofing Attack in a Brain–Computer Interface System

Figure 3.3 illustrates how a compromised EEG device can mimic a legitimate hardware component to gain unauthorized access to the BCI system. Unlike trusted devices, the spoofed unit bypasses authentication mechanisms, enabling malicious signal injection and potential system-level compromise. This highlights the necessity of hardware-based trust mechanisms such as Physical Unclonable Functions (PUFs) to ensure device authenticity.

3.4.3 Impact on Trust and Integrity

Hardware spoofing attacks have profound implications for the reliability and trustworthiness of BCI systems. The key impacts include:

- **Unauthorized System Access:**
  Spoofed devices can bypass authentication mechanisms and enter the system, allowing attackers to operate within trusted environments.
- **Data Manipulation and Falsification:**
  Compromised hardware can inject falsified or altered neural signals, leading to incorrect system outputs and degraded performance.

- **Breakdown of Trust in Hardware Components:**
  Since BCI systems depend on trusted signal sources, any compromise at the hardware level undermines confidence in the entire system architecture.

Increased Risk of Coordinated Attacks:
Hardware spoofing can serve as an entry point for more complex multi-layer attacks, combining signal injection, data manipulation, and system hijacking.

3.4.4 Limitations of Current Solutions and Future Direction

Existing approaches to mitigating hardware spoofing attacks primarily rely on software-based authentication and device identification techniques. However, these methods are inherently limited, as attackers can replicate device identifiers or manipulate firmware to bypass such protections. This highlights the inadequacy of relying solely on software-level defenses in securing BCI systems.

Hardware-based security mechanisms, such as **Physical Unclonable Functions (PUFs),** have been proposed as a more robust alternative. PUFs exploit intrinsic manufacturing variations in hardware components to generate unique and unclonable device signatures, making them highly resistant to duplication. Despite their potential, these techniques are not yet widely adopted in BCI systems due to challenges related to integration complexity, cost, and lack of standardization.

Another limitation of current solutions is the absence of continuous hardware verification. Once a device is authenticated, it is typically trusted for the duration of operation, creating an opportunity for compromised devices to remain undetected.

To address these challenges, the proposed **Neuro-Shield framework** introduces a hybrid security model that combines **hardware-rooted trust mechanisms** with continuous signal validation. By integrating **Physical Unclonable Functions (PUFs)** with real-time anomaly detection, Neuro-Shield ensures that only authenticated and trustworthy devices participate in the system. This approach enhances both hardware integrity and resilience against spoofing attacks, while maintaining compatibility with real-time BCI requirements.

To provide a clearer comparison of these threats, Table 3.1 summarizes their characteristics across different layers of the BCI system.

Threat Type	Attack Nature	Target Layer	Impact Severity
Eavesdropping	Passive	Communication	High
Signal Injection	Active	Input/Data Layer	Critical
Hardware Spoofing	Physical	Hardware Layer	High

Table 3.1: Comparative Analysis of Security Threats in Brain–Computer Interface Systems

3.5 Discussion

The analysis of the identified threat models reveals the multi-dimensional nature of security challenges in BCI systems. Unlike conventional systems, where threats are often confined to software or network layers, BCI systems are vulnerable across multiple domains, including signal acquisition, data transmission, machine learning processing, and hardware integrity.

This layered vulnerability significantly increases the complexity of designing effective defense mechanisms. A single-point security solution is insufficient, as attackers can exploit different stages of the pipeline to achieve their objectives.

Therefore, there is a critical need for integrated, multi-layered security frameworks that can simultaneously address privacy, data integrity, and real-time performance constraints. Such frameworks must be capable of detecting both **stealthy passive attacks** and **active adversarial manipulations**, while maintaining minimal computational overhead.

3.6 Summary

This chapter presented a comprehensive analysis of key threat models in Brain–Computer Interface systems, including passive eavesdropping, signal injection attacks, and hardware spoofing. Each threat was examined in terms of its attack mechanism, impact on system operation, and limitations of existing defense strategies.

The findings highlight that BCI systems require a holistic and integrated security approach that extends beyond traditional methods. Addressing these challenges necessitates the incorporation of security mechanisms directly into the signal processing and hardware layers.

In the next chapter, a detailed design of the proposed **Neuro-Shield framework** will be presented. This framework introduces a multi-layered defense strategy aimed at enhancing system security, ensuring data integrity, and maintaining real-time performance in next-generation BCI systems. Ultimately, this chapter highlights that effective security in BCI systems can only be achieved through integrated, multi-layered defense strategies designed for real-time environments.

CHAPTER 4

NEURO-SHIELD ARCHITECTURE

4.1 Introduction

The rapid advancement and increasing deployment of Brain–Computer Interface (BCI) systems in real-world applications have introduced a wide range of security and privacy challenges. Modern BCI systems are no longer confined to controlled laboratory environments; instead, they are actively used in assistive technologies, healthcare systems, rehabilitation engineering, and human–machine interaction platforms. As a result, these systems are now exposed to complex and evolving threat landscapes that span multiple layers, including signal acquisition, wireless communication, machine learning-based processing, and hardware integrity [18], [22].

As highlighted in Chapter 3, existing security solutions for BCI systems are often **fragmented and layer-specific**, addressing isolated threats without considering the system as a whole. Such approaches are insufficient against sophisticated adversaries capable of launching multi-stage attacks, including **passive eavesdropping**, **signal injection**, and **hardware spoofing**. Furthermore, the real-time operational requirements of BCI systems impose strict constraints on computational overhead, making it challenging to apply conventional security mechanisms such as heavy encryption or complex authentication protocols [10], [23].

In addition to security concerns, the **sensitive nature of neural data** introduces critical privacy challenges. EEG signals can reveal deeply personal information, including cognitive states, emotional responses, and behavioral patterns. Unauthorized access or manipulation of such data not only compromises system integrity but also raises serious ethical and safety concerns. Therefore, ensuring both **data confidentiality** and **system reliability** is essential for the widespread adoption of BCI technologies.

To overcome these critical limitations, this chapter presents Neuro-Shield, a novel multi-layered security architecture specifically designed for BCI systems. Unlike conventional

approaches, Neuro-Shield integrates security mechanisms directly into the signal processing pipeline, enabling continuous, real-time, and adaptive protection across all system layers.

The Neuro-Shield architecture is structured into six interconnected tiers, each responsible for securing a specific functional layer of the system. These tiers collectively provide a combination of **hardware-rooted trust**, **secure signal validation**, **machine learning-based anomaly detection**, and **privacy-preserving techniques**. By leveraging these complementary mechanisms, Neuro-Shield establishes a robust and adaptive security framework capable of defending against both **stealthy passive attacks** and **active adversarial manipulations**.

Furthermore, the proposed architecture is designed with a strong emphasis on **real-time performance and scalability**, ensuring that security enhancements do not compromise system responsiveness. This makes Neuro-Shield suitable for deployment in next-generation BCI applications where both security and usability are critical.

Overall, this chapter presents the design and functional components of the Neuro-Shield architecture, providing a unified solution to the multi-dimensional security challenges identified in previous chapters. The proposed framework lays the foundation for secure, reliable, and privacy-aware BCI systems.

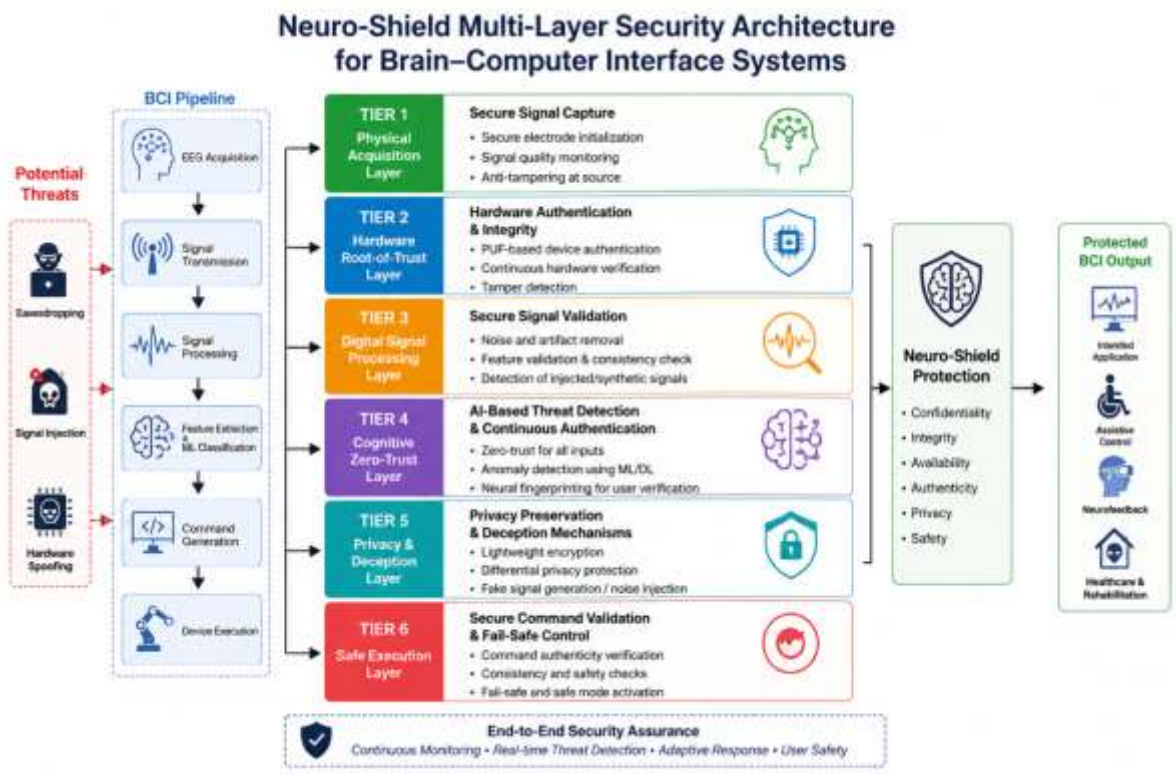

Figure 4.1: Neuro-Shield architecture showing multi-layer security and end-to-end protection.

The key contributions of this chapter are summarized as follows:

• A unified six-tier security architecture for BCI systems

• Integration of hardware-rooted trust using PUF-based authentication

• Real-time signal validation and adversarial detection mechanisms

• Cognitive Zero-Trust framework with continuous neural authentication

• Privacy-preserving and deception-based defense strategies

• Safe execution layer with fail-safe protection mechanisms

4.2 Tier 1: Physical Acquisition Layer

The Physical Acquisition Layer is the first stage of the Neuro-Shield architecture, where brain signals are collected directly from the user. It involves EEG sensors that capture neural activity in real time.

Since this is the entry point of the system, maintaining signal integrity is very important. Any noise, interference, or unauthorized access at this stage can affect the entire pipeline. Basic security measures are applied to ensure that only valid signals are captured. This layer forms the foundation for accurate processing and secure system operation.

4.2.1 Secure Signal Capture

The Physical Acquisition Layer represents the first line of defense in the Neuro-Shield architecture, ensuring that only authentic and high-integrity neural signals enter the system. At this stage, neural activity is captured using Electroencephalography (EEG) electrodes. As the entry point of the signal processing pipeline, this layer plays a critical role in determining the overall reliability, accuracy, and security of the system. Any compromise at this stage can propagate through subsequent layers, potentially leading to incorrect interpretations and unsafe system behavior [22], [20].

EEG signals are inherently weak and highly susceptible to external interference, including environmental noise, electromagnetic disturbances, and physiological artifacts such as eye blinks and muscle activity. In addition to these natural disturbances, this layer is also vulnerable to intentional attacks, such as **signal injection at the source** or physical tampering with acquisition devices [16], [26]. Therefore, ensuring the integrity of captured signals is essential for maintaining system trustworthiness.

To address these challenges, the proposed **Neuro-Shield architecture** integrates advanced mechanisms for **secure electrode calibration** and **signal integrity verification**. Secure electrode calibration ensures proper placement and stable contact between electrodes and the scalp, minimizing signal distortion and preventing unauthorized manipulation. This process may include impedance monitoring and adaptive calibration techniques to maintain consistent signal quality [19].

Furthermore, **signal integrity verification mechanisms** are employed to validate the authenticity of incoming neural signals in real time. These mechanisms analyze signal characteristics such as amplitude consistency, frequency distribution, and temporal patterns to detect anomalies or irregularities. Any deviation from expected neural signal behavior can be flagged for further inspection or rejected before entering the processing pipeline [23].

By embedding these security measures directly into the acquisition stage, Neuro-Shield ensures that only **authentic, high-quality neural signals** are digitized and forwarded for further processing. This not only improves system reliability but also establishes a trusted foundation for all subsequent layers of the architecture.

4.2.2 Threat Mitigation

The **Physical Acquisition Layer** serves as the first line of defense in the Neuro-Shield architecture, protecting the system against both environmental disturbances and adversarial attacks at the earliest stage of the BCI pipeline. Since this layer directly interfaces with neural signal acquisition, any compromise at this stage can propagate through subsequent processing layers, amplifying its impact. Therefore, robust threat mitigation mechanisms are essential to ensure system reliability and security.

The key security capabilities of this layer are outlined as follows:

- **Prevention of Signal Injection at the Source:**
  Neuro-Shield incorporates real-time signal validation techniques during acquisition to detect anomalies in amplitude, frequency, and temporal patterns. These mechanisms enable the system to identify and block **malicious or externally injected signals** before they enter the processing pipeline, thereby reducing the risk of adversarial manipulation at its origin [16], [26].

- **Reduction of Noise-Based Manipulation:**
  EEG signals are highly sensitive to environmental noise and physiological artifacts, which can be exploited by attackers to distort signal interpretation. To mitigate this, the system employs advanced filtering methods, including adaptive filtering and artifact removal techniques, to suppress noise and maintain signal clarity. This minimizes the likelihood of attackers leveraging noise-induced distortions to influence system behavior [19].

- **Ensuring Authenticity of Raw EEG Data:**
  Continuous **signal integrity verification mechanisms** are implemented to ensure that only genuine neural signals are captured and transmitted. These mechanisms analyze statistical properties and consistency patterns of the incoming signals,

enabling the detection of abnormal or fabricated data. This guarantees the reliability and trustworthiness of the acquired EEG signals [23].

- **Enhanced Signal Quality and Stability:**
  Secure electrode calibration techniques, including impedance monitoring and adaptive adjustment, are used to maintain stable electrode–skin contact. This improves signal consistency and reduces variability, thereby enhancing both system performance and resilience against external disturbances [20].

- **Early-Stage Attack Containment:**
  By identifying and mitigating threats at the acquisition stage, Neuro-Shield prevents the propagation of compromised signals into higher-level processing layers. This significantly reduces the overall attack surface and improves the effectiveness of subsequent security mechanisms.

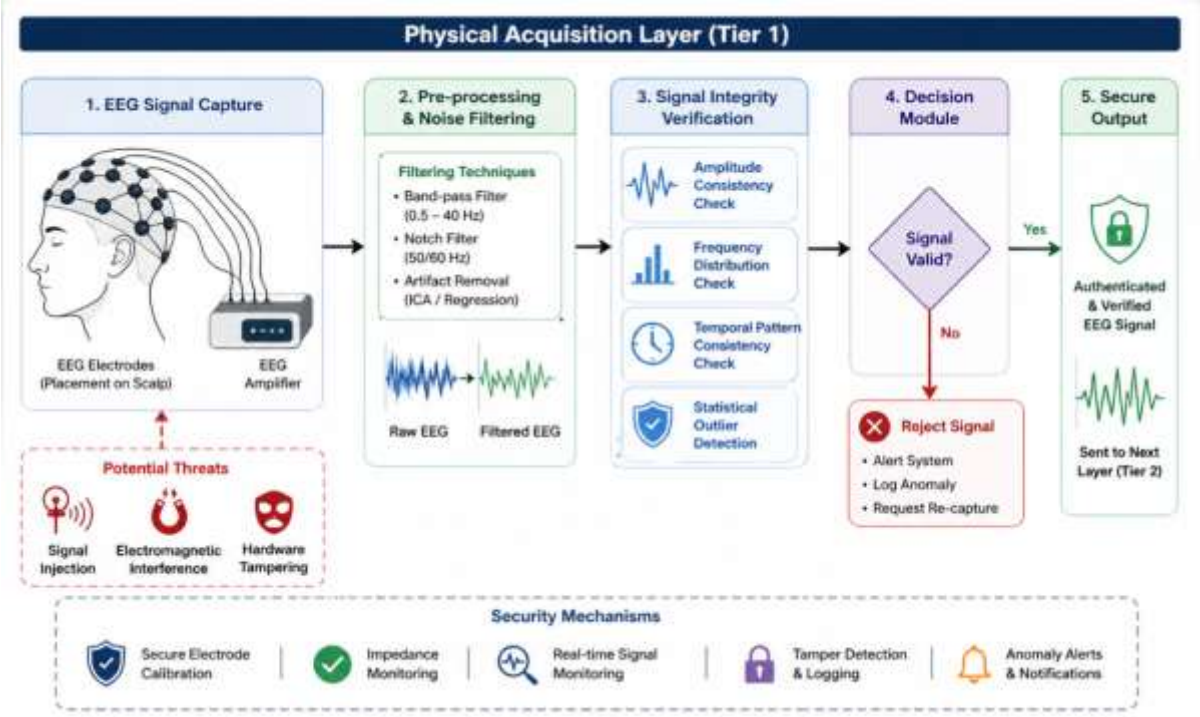

Figure 4.2: Secure EEG Signal Acquisition and Integrity Verification in the Physical Acquisition Layer

4.3 Tier 2: Hardware Root-of-Trust

The Hardware Root-of-Trust layer ensures that only trusted devices are allowed to participate in the system. It relies on unique hardware characteristics, such as Physical Unclonable Functions (PUFs), to verify device identity. This makes it extremely difficult for attackers to replicate or spoof legitimate hardware. Each device is authenticated before any data exchange takes place. Continuous verification helps detect if a device becomes compromised during operation. Overall, this layer establishes a strong foundation of trust for the entire BCI system.

4.3.1 Device Authentication using Physical Unclonable Functions (PUFs)

The **Hardware Root-of-Trust Layer** forms a critical security foundation within the Neuro-Shield architecture by ensuring that only authenticated and trusted hardware components are allowed to participate in the Brain–Computer Interface (BCI) system. Given that EEG acquisition devices serve as the primary interface between neural activity and computational processing, any compromise at this level can have severe consequences, including **data manipulation, unauthorized system access, and complete system hijacking**. Therefore, establishing a reliable and tamper-resistant hardware authentication mechanism is essential for maintaining overall system integrity [31].

To address this challenge, the proposed Neuro-Shield framework leverages **Physical Unclonable Functions (PUFs)** as a robust and lightweight hardware-based authentication solution. PUFs exploit intrinsic and uncontrollable manufacturing variations present in semiconductor devices to generate unique and device-specific signatures. These signatures are inherently **unclonable, unpredictable, and resistant to physical duplication**, even under advanced fabrication conditions, making them highly suitable for secure device identification and authentication [14], [31].

In the proposed design, each EEG acquisition device is embedded with a PUF module that operates based on a **challenge–response protocol**. When a specific challenge input is applied, the PUF generates a corresponding response that is unique to that device. This

response is then compared against a securely stored reference profile during the authentication process. Unlike traditional cryptographic systems, PUF-based authentication does not require permanent storage of secret keys, thereby significantly reducing the risk of key leakage or extraction attacks.

Furthermore, the inherent randomness and uniqueness of PUF responses provide strong resistance against **hardware cloning, reverse engineering, and replay attacks**. Even if an attacker gains physical access to the device, replicating its exact PUF behavior remains practically infeasible. This makes PUFs particularly effective in mitigating **hardware spoofing attacks**, where adversaries attempt to impersonate legitimate devices within the BCI system.

By integrating PUF-based authentication into the Hardware Root-of-Trust Layer, Neuro-Shield establishes a **secure and tamper-resistant hardware identity framework**, ensuring that only verified and trusted devices can access and interact with the system. This significantly enhances system reliability and provides a strong defensive barrier against unauthorized device participation and hardware-level attacks [31].

4.3.2 Continuous Hardware Verification

Traditional hardware authentication mechanisms in Brain–Computer Interface (BCI) systems typically rely on a **one-time verification process** performed during system initialization. While this approach may be sufficient in static environments, it is inadequate for dynamic and real-time BCI systems, where hardware components may become compromised after initial authentication. Such vulnerabilities can allow adversaries to maintain persistent access by exploiting previously trusted devices, thereby undermining overall system security [31].

To address this limitation, the proposed **Neuro-Shield architecture** introduces a **continuous hardware verification mechanism**, designed to provide ongoing validation of device authenticity throughout system operation. Unlike conventional methods, this approach ensures that hardware trust is not assumed after initial authentication but is instead **continuously re-evaluated in real time**.

The continuous verification process combines **periodic PUF-based authentication** with **behavioral and signal-level monitoring**. At regular intervals, the system reissues challenge–

response queries to the embedded PUF modules within EEG devices, verifying that their responses remain consistent with previously registered profiles. In parallel, the system monitors device behavior, including signal transmission patterns, timing characteristics, and statistical properties of the generated EEG data.

Any deviation from expected hardware responses or abnormal activity patterns—such as inconsistent PUF outputs, irregular signal characteristics, or unexpected communication behavior—is immediately flagged as a potential security threat. Upon detection, the system can trigger appropriate countermeasures, including **device isolation, access revocation, or system alerts**, thereby preventing compromised hardware from influencing the signal processing pipeline.

This **dynamic and adaptive verification strategy** significantly enhances system resilience by ensuring that hardware trust is continuously enforced rather than statically assumed. It effectively mitigates the risk of **persistent attacks**, where adversaries exploit authenticated devices over extended periods. Furthermore, it provides an additional layer of protection against advanced threats such as **hardware tampering, firmware modification, and replay-based impersonation attacks** [14], [31].

By integrating continuous verification into the Hardware Root-of-Trust Layer, Neuro-Shield establishes a robust and proactive defense mechanism that maintains hardware integrity throughout the entire operational lifecycle of the BCI system.

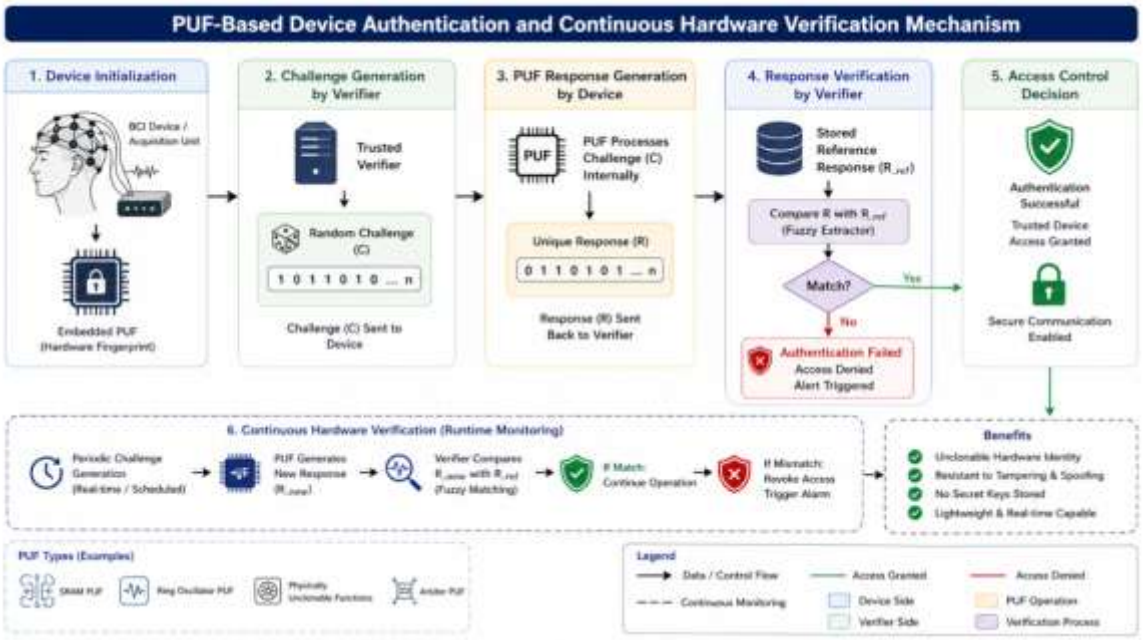

Figure 4.3: PUF-Based Device Authentication and Continuous Hardware Verification Mechanism

4.3.3 Security Benefits

The **Hardware Root-of-Trust Layer** provides a set of critical security advantages that significantly enhance the overall integrity, reliability, and resilience of the Neuro-Shield architecture. By establishing a secure and verifiable hardware foundation, this layer ensures that all subsequent stages of the Brain–Computer Interface (BCI) system operate on trusted and authenticated inputs. The key security benefits are described as follows:

- **Prevention of Hardware Spoofing:**
  The integration of **PUF-based authentication** ensures that each EEG device possesses a unique and unclonable hardware identity. This effectively prevents attackers from replicating or impersonating legitimate devices, thereby eliminating the risk of **device cloning and hardware spoofing attacks**. Since PUF responses are inherently tied to physical characteristics, unauthorized duplication becomes practically infeasible [31].

- **Ensuring Trusted Device Participation:**
  Through the implementation of **continuous hardware verification**, the system ensures that only authenticated and trustworthy devices remain active throughout

operation. This dynamic validation mechanism prevents compromised or unauthorized devices from maintaining persistent access, thereby preserving a secure and controlled operational environment.

- **Enabling Tamper Detection:**
  One of the key advantages of PUF-based systems is their sensitivity to physical changes. Any unauthorized modification, such as hardware tampering or firmware alteration, results in variations in the PUF response. These deviations can be detected in real time, enabling early identification of tampering attempts and preventing further exploitation [14].

- **Enhanced System Integrity and Reliability:**
  By establishing a **hardware-rooted trust foundation**, this layer ensures that all data entering the BCI pipeline originates from verified and authentic sources. This significantly reduces the risk of cascading failures caused by compromised inputs and enhances the overall reliability of the system.

- **Resistance to Advanced Hardware Attacks:**
  The combination of PUF-based authentication and continuous verification provides strong protection against sophisticated attacks, including **replay attacks, side-channel analysis, and reverse engineering attempts**. This makes the system more resilient against both physical and cyber-physical threats [14], [31].

These hardware-level assurances provide a trusted foundation for the subsequent signal processing and cognitive validation layers.

4.4 Tier 3: Digital Signal Processing Layer

The Digital Signal Processing Layer is responsible for preparing raw brain signals for analysis.

It cleans the data by removing noise and unwanted artifacts to ensure accuracy. Key features are then extracted to represent meaningful patterns from the signals. This processed data becomes the input for machine learning and decision-making stages. Security checks can also be applied to detect abnormal or manipulated signals. Overall, this layer plays a crucial role in ensuring both signal quality and system reliability.

4.4.1 Secure Preprocessing

The **Digital Signal Processing (DSP) Layer** plays a fundamental role in transforming raw Electroencephalography (EEG) signals into a clean, stable, and analyzable representation suitable for subsequent feature extraction and classification. Given that EEG signals are inherently low-amplitude and highly sensitive to both internal and external disturbances, this stage is critical not only for improving signal quality but also for ensuring the overall security and reliability of the Brain–Computer Interface (BCI) system.

In practical scenarios, raw EEG signals are frequently contaminated by various sources of noise and interference. These include **physiological artifacts**, such as eye blinks (EOG), muscle activity (EMG), and cardiac signals (ECG), as well as **external disturbances** like electromagnetic interference and environmental noise. Such distortions can significantly degrade signal integrity and, more importantly, can be exploited by adversaries to introduce subtle manipulations that affect downstream processing [19].

To address these challenges, the proposed **Neuro-Shield architecture** enhances the preprocessing stage by integrating **security-aware filtering mechanisms** that extend beyond conventional signal cleaning techniques. In addition to standard preprocessing methods—such as band-pass filtering, notch filtering, and adaptive noise cancellation—the system incorporates **anomaly-sensitive filters** designed to detect irregular signal behavior. These filters analyze both **temporal features** (e.g., signal continuity, sudden amplitude shifts) and **spectral characteristics** (e.g., abnormal frequency components, power distribution anomalies) to identify deviations from expected EEG patterns.

Furthermore, Neuro-Shield employs **adaptive thresholding and statistical consistency checks** during preprocessing to distinguish between natural noise and potential adversarial perturbations. This allows the system to detect subtle forms of **adversarial signal manipulation**, including crafted perturbations that are designed to bypass traditional filtering methods. By leveraging these techniques, the system can effectively identify and suppress suspicious signal components before they propagate to higher-level processing stages [3], [23].

Another important aspect of secure preprocessing is the **real-time validation of signal integrity**. The system continuously monitors incoming signals for consistency across time windows, ensuring that abrupt or non-physiological changes are promptly detected. When anomalies are identified, the system can either filter out the affected segments or flag them for further analysis, thereby preventing compromised data from influencing classification outcomes.

By embedding these security mechanisms directly into the preprocessing stage, Neuro-Shield significantly reduces the attack surface at an early point in the BCI pipeline. This proactive approach ensures that only **clean, authentic, and trustworthy signals** are forwarded for feature extraction and classification, thereby enhancing both system robustness and resilience against adversarial threats.

4.4.2 Feature Validation

Following the preprocessing stage, meaningful features are extracted from EEG signals to represent underlying cognitive states and user intent. These features serve as the primary input to machine learning models responsible for decoding neural activity. However, this stage is particularly vulnerable to adversarial manipulation, as attackers may attempt to introduce **synthetic or crafted signals** that closely resemble legitimate neural patterns in order to deceive the system.

To mitigate this risk, the proposed **Neuro-Shield architecture** incorporates a robust **feature validation mechanism** designed to ensure the authenticity and reliability of extracted features before they are forwarded to classification models. This validation process is based on a combination of **statistical analysis, temporal consistency evaluation, and inter-feature correlation checks**.

Specifically, the system examines the **statistical distribution** of extracted features, including measures such as mean, variance, and higher-order moments, to verify that they fall within expected physiological ranges. In addition, **temporal stability analysis** is performed across consecutive time windows to ensure that feature values exhibit smooth and biologically plausible transitions. Abrupt or irregular changes may indicate the presence of adversarial perturbations or injected signals.

Furthermore, Neuro-Shield evaluates **correlation patterns between multiple features**, as genuine neural signals typically exhibit structured relationships across frequency bands and spatial channels. Disruptions in these correlations can serve as indicators of manipulated or synthetic data. By leveraging these multi-dimensional validation techniques, the system can effectively distinguish between authentic neural activity and adversarial inputs.

In cases where anomalies are detected, the system can either reject the affected feature set, request re-acquisition of the signal, or flag the data for further analysis. This ensures that compromised or unreliable features do not propagate to downstream machine learning models, thereby reducing the risk of incorrect classification and unintended system behavior.

Overall, the feature validation process significantly enhances the robustness and trustworthiness of the BCI system by ensuring that only **authentic, consistent, and physiologically valid features** are used for decision-making. This plays a crucial role in defending against adversarial attacks and maintaining the integrity of cognitive decoding processes [15], [23].

4.4.3 Threat Mitigation

Threat mitigation in this layer focuses on detecting and filtering out abnormal or manipulated signals during processing. It ensures that only clean and reliable data is passed to the next stages of the system. Techniques such as anomaly detection and signal validation help identify suspicious patterns. By analyzing signal consistency, the system can detect signs of injection or tampering. This reduces the chances of incorrect decisions caused by malicious inputs. Overall, this layer strengthens the system's resilience against signal-level attacks.

The **Digital Signal Processing (DSP) Layer** plays a critical role in safeguarding the Brain–Computer Interface (BCI) system by mitigating a wide range of signal-level threats before they can influence higher-level processing and decision-making. By integrating **security-aware preprocessing** and **feature validation mechanisms**, this layer acts as an early defensive barrier against adversarial manipulation and signal corruption. The key threat mitigation capabilities of this layer are outlined below:

- **Detection of Adversarial Perturbations:**

  The DSP layer is equipped with anomaly-sensitive filtering and validation techniques that can identify subtle perturbations intentionally introduced to mislead machine learning models. These perturbations are often designed to remain imperceptible while significantly altering classification outcomes. By analyzing temporal continuity, spectral characteristics, and statistical consistency, the system can effectively detect such adversarial manipulations and prevent them from influencing downstream processes [3], [4].

- **Filtering of Injected or Manipulated Signals:**

  Malicious or externally injected signals are identified and suppressed during the preprocessing and validation stages. Advanced filtering techniques, combined with consistency checks, enable the system to distinguish between genuine neural activity and manipulated inputs. This ensures that corrupted signals are either removed or attenuated before they propagate through the pipeline, thereby preserving data integrity.

- **Improved Classification Reliability:**

  By ensuring that only **high-quality, validated, and physiologically consistent features** are passed to machine learning models, the DSP layer significantly enhances classification accuracy and robustness. This reduces the likelihood of incorrect predictions caused by noisy, corrupted, or adversarially manipulated data, thereby improving overall system performance and reliability [15], [23].

- **Early Detection and Containment of Threats:**

  One of the key advantages of the DSP layer is its ability to detect and mitigate threats at an early stage of the processing pipeline. By preventing compromised signals from advancing to higher layers, the system minimizes the risk of cascading failures and reduces the computational burden on subsequent security mechanisms.

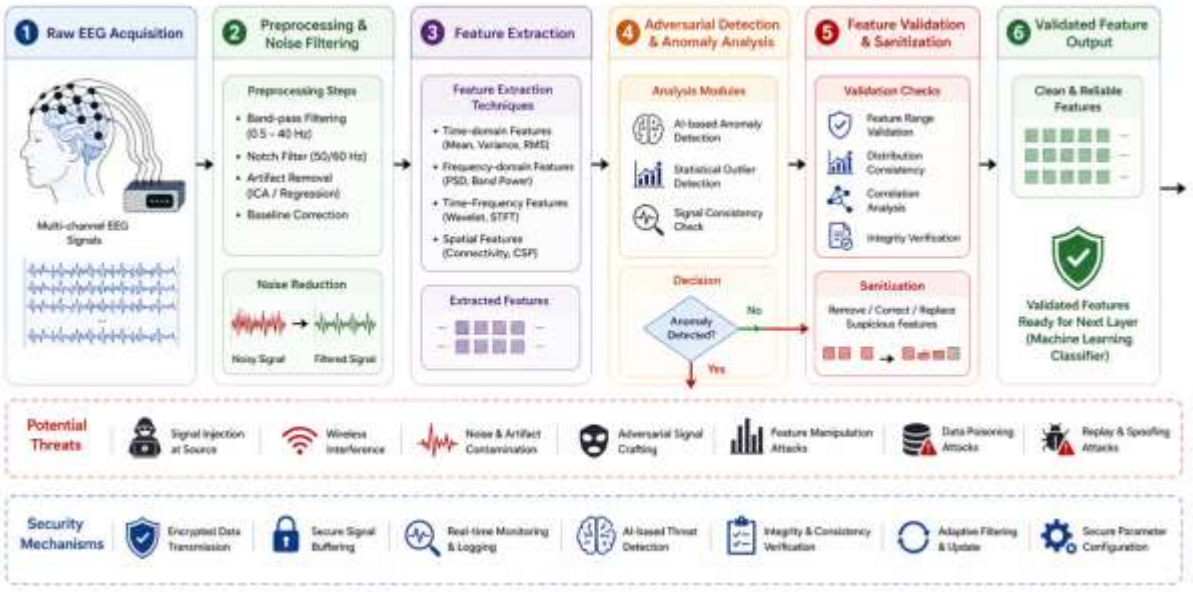

Figure 4.4: Secure Digital Signal Processing Pipeline with Adversarial Detection and Feature Validation

4.5 Tier 4: Cognitive Zero-Trust Layer

The Cognitive Zero-Trust Layer ensures that no signal or component is trusted automatically within the system. Every input is continuously verified before being accepted for further processing.

It uses intelligent models to detect anomalies and suspicious patterns in real time. Even previously authenticated data is re-evaluated to prevent hidden threats. This approach minimizes the risk of insider or undetected attacks. Overall, it provides a dynamic and adaptive security mechanism for the BCI system.

4.5.1 Zero-Trust Principle

The **Cognitive Zero-Trust Layer** introduces a transformative security paradigm within the Neuro-Shield architecture by adopting the **Zero-Trust model**, a concept widely recognized in modern cybersecurity frameworks. Unlike traditional security approaches that rely on predefined trust boundaries, the Zero-Trust model operates on the principle of **"never trust, always verify."** In this context, no signal, device, or user input is assumed to be trustworthy by default; instead, every entity is continuously authenticated and validated throughout the system's operation.

This paradigm is particularly critical in Brain–Computer Interface (BCI) systems, where the integration of biological signals, machine learning models, and networked communication creates a highly dynamic and vulnerable environment. Both **internal threats** (e.g., compromised devices or corrupted signals) and **external adversaries** (e.g., signal injection or data interception attacks) can exploit implicit trust assumptions to bypass conventional security mechanisms. By eliminating such assumptions, the Zero-Trust approach significantly reduces the attack surface and enhances overall system resilience [3], [10].

Within the Neuro-Shield framework, the Zero-Trust principle is implemented through **continuous validation of all incoming data and system interactions**. Every neural signal, feature set, and device response is subjected to rigorous verification processes, including consistency checks, anomaly detection, and behavioral analysis. This ensures that only **authentic, contextually valid, and trustworthy inputs** are allowed to proceed to subsequent processing stages.

Moreover, the Cognitive Zero-Trust Layer extends beyond static verification by incorporating **context-aware decision-making**. The system evaluates inputs not only based on predefined thresholds but also in relation to historical patterns and user-specific behavioral profiles. This dynamic validation mechanism enables the detection of subtle and evolving threats that may otherwise remain undetected in traditional systems.

By adopting this continuous and adaptive verification strategy, Neuro-Shield effectively minimizes the risk of **undetected attacks, data manipulation, and unauthorized system access**. As a result, the Zero-Trust principle serves as a cornerstone of the architecture, ensuring robust protection against both known and emerging threats while maintaining the integrity and reliability of BCI operations [3], [10].

4.5.2 AI-Based Anomaly Detection

To effectively enforce the Zero-Trust principle within the Neuro-Shield architecture, the system integrates advanced **machine learning-based anomaly detection models** capable of identifying abnormal neural patterns in real time. Given the complex and dynamic nature of EEG signals, traditional rule-based detection methods are often insufficient to capture subtle

deviations or evolving attack strategies. Therefore, the use of data-driven intelligence becomes essential for robust and adaptive threat detection.

In the proposed framework, these models are trained on **baseline EEG datasets** to learn the normal cognitive and behavioral patterns of individual users. By establishing a personalized reference profile, the system can accurately detect deviations that may arise due to **adversarial manipulation, signal injection, or unexpected system anomalies**. This personalized learning approach enhances detection sensitivity while reducing false positives.

To further improve performance, Neuro-Shield employs **advanced learning techniques**, including deep neural networks, convolutional models, and adaptive classifiers. These models are capable of capturing complex temporal and spatial dependencies within EEG signals, enabling them to identify both **subtle perturbations** and large-scale anomalies. In addition, adaptive learning mechanisms allow the models to update their understanding of normal behavior over time, ensuring resilience against concept drift and changing user conditions.

A key strength of this approach lies in its ability to detect both **known and previously unseen attack patterns**. While supervised learning methods can identify known adversarial signatures, unsupervised and semi-supervised techniques enable the detection of novel anomalies that do not conform to learned patterns. This makes the system particularly effective against **zero-day attacks and evolving threat scenarios** [3], [15].

Furthermore, the anomaly detection module operates in real time, continuously monitoring incoming data streams and providing immediate feedback to the system. When suspicious patterns are detected, the system can trigger appropriate countermeasures, such as rejecting the input, initiating re-authentication, or activating higher-level security controls.

By integrating AI-based anomaly detection into the Cognitive Zero-Trust Layer, Neuro-Shield establishes a **proactive and adaptive defense mechanism** that significantly enhances the system's ability to detect and mitigate sophisticated cyber-physical threats in BCI environments.

4.5.3 Continuous Authentication

In addition to anomaly detection, the Cognitive Zero-Trust Layer incorporates a robust **continuous authentication mechanism** based on **neural fingerprinting**, which leverages the unique characteristics of individual EEG signals as a biometric identifier. Unlike conventional authentication methods that rely on static credentials such as passwords or tokens, neural fingerprinting utilizes **intrinsic brain activity patterns**, which are inherently difficult to replicate or forge.

Each individual exhibits distinct EEG signatures characterized by variations in frequency bands, spatial distributions, and temporal dynamics. These unique patterns can be learned and modeled to create a **personalized neural profile**, which serves as a continuous identity reference for the user. By comparing incoming EEG signals with this reference profile, the system can verify the user's identity in real time throughout system operation [13], [21].

A key advantage of this approach is that authentication is not limited to a single point in time (e.g., system login) but is instead performed **continuously and transparently** during system usage. This ensures that any unauthorized access attempt—such as session hijacking, device sharing, or impersonation—is promptly detected, even after initial authentication has been successfully completed.

Furthermore, Neuro-Shield integrates **adaptive learning mechanisms** to account for natural variations in EEG signals caused by factors such as fatigue, emotional state, or environmental conditions. This dynamic adaptation improves authentication accuracy while minimizing false rejection rates, thereby enhancing user experience without compromising security.

In addition, the combination of neural fingerprinting with anomaly detection provides a **multi-factor cognitive authentication framework**, where both identity verification and behavioral consistency are continuously monitored. This effectively strengthens the system's ability to detect sophisticated attacks that attempt to mimic legitimate user behavior.

Overall, continuous authentication through neural fingerprinting enhances system security by ensuring that only **authorized users maintain access throughout the entire session**,

thereby providing robust protection against both internal and external threats in BCI environments [13], [21].

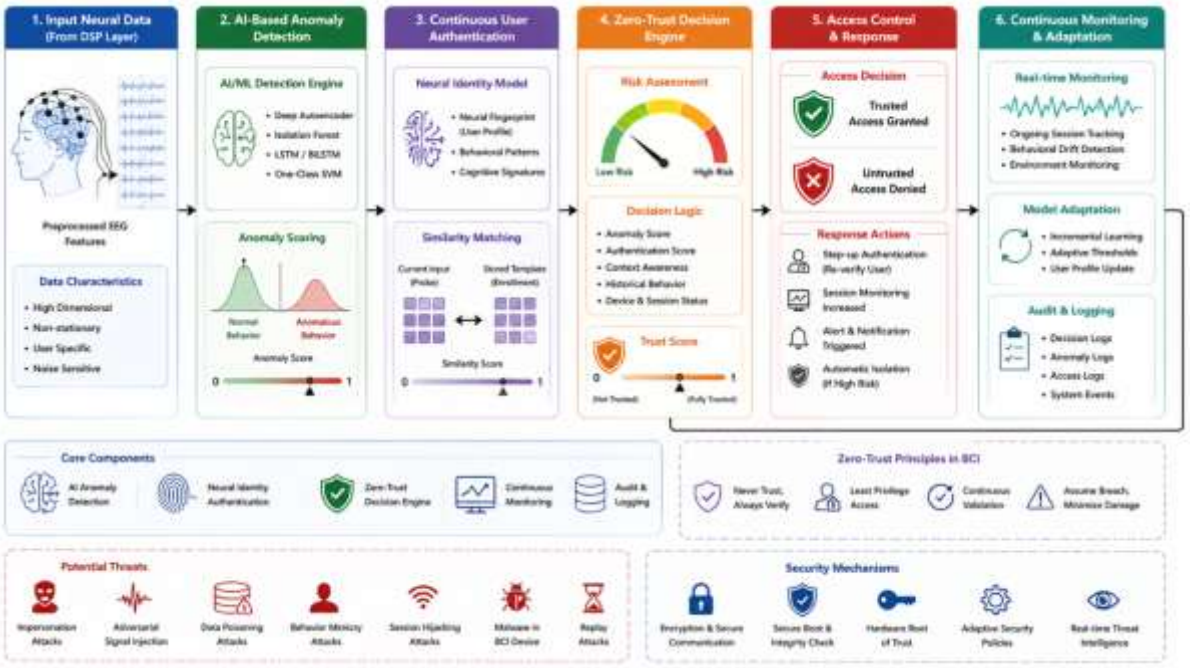

Figure 4.5: Cognitive Zero-Trust Framework with AI-Based Anomaly Detection and Continuous Authentication

4.5.4 Security Advantages

The **Cognitive Zero-Trust Layer** provides a set of critical security advantages that significantly strengthen the resilience, reliability, and trustworthiness of the Neuro-Shield architecture. By combining continuous validation, AI-driven anomaly detection, and neural-based authentication, this layer ensures that all system interactions are rigorously verified throughout operation. The key security benefits are outlined below:

- **Prevention of Model Manipulation:**

  Continuous validation of input signals and extracted features prevents adversaries from influencing or corrupting machine learning models. By ensuring that only **verified and physiologically consistent data** is processed, the system minimizes the risk of adversarial inputs altering model behavior or decision boundaries. This significantly enhances the robustness of classification models against manipulation attempts [3], [4].

- **Detection of Adversarial Attacks:**

  The integration of real-time **AI-based anomaly detection mechanisms** enables the system to identify both subtle perturbations and large-scale adversarial attacks. By analyzing deviations in neural patterns and behavioral characteristics, the system can detect sophisticated attacks, including those designed to evade traditional detection methods. This capability ensures proactive threat identification and rapid response [3], [15].

- **Continuous User Authentication:**

  The implementation of **neural fingerprinting** provides continuous and dynamic user authentication throughout system operation. Unlike static authentication methods, this approach ensures that only authorized users can maintain access, even during extended sessions. This effectively mitigates risks associated with session hijacking, impersonation, and unauthorized system usage [13], [21].

- **Adaptive and Context-Aware Security Enforcement:**

  The Zero-Trust framework enables context-aware validation by continuously analyzing user behavior and system interactions. This adaptive mechanism allows the system to respond dynamically to evolving threats and changing user conditions, improving both detection accuracy and system flexibility.

Enhanced System Trust and Reliability:

By eliminating implicit trust and enforcing continuous verification, the Cognitive Zero-Trust Layer ensures that all components of the BCI system operate within a secure and controlled environment. This significantly reduces the likelihood of undetected attacks and enhances overall system reliability.

4.6 Tier 5: Privacy & Deception Layer

The Privacy and Deception Layer is designed to protect sensitive neural data from being exposed or misused.

It applies techniques such as data masking and encryption to preserve user privacy. In addition, deception mechanisms can introduce misleading signals to confuse potential attackers.

This makes it harder for unauthorized entities to interpret or exploit the data. The layer also ensures that only necessary information is shared within the system. Overall, it strengthens both privacy protection and defense against data-focused attacks.

4.6.1 Privacy Preservation

The **Privacy & Deception Layer** plays a vital role in safeguarding sensitive neural data within the Neuro-Shield architecture by ensuring confidentiality, integrity, and resistance to unauthorized inference. Given the highly personal and sensitive nature of EEG signals, which can reveal cognitive states, emotional responses, and behavioral patterns, protecting such data is essential for maintaining **user trust, ethical compliance, and regulatory adherence**. Any unauthorized access or leakage of neural data may lead to serious privacy violations and potential misuse.

To address these concerns, Neuro-Shield incorporates **lightweight encryption mechanisms** specifically designed for real-time Brain–Computer Interface (BCI) systems. Unlike traditional encryption methods that may introduce significant computational overhead, these optimized techniques ensure secure data transmission and storage while maintaining low latency and energy efficiency. This balance between **security and performance** is critical in BCI environments, where real-time responsiveness is a key requirement [14].

In addition to encryption, architecture integrates **differential privacy mechanisms** to protect against inference-based attacks. Differential privacy introduces carefully calibrated noise into the data or intermediate outputs, ensuring that individual-specific information cannot be easily extracted, even if an attacker gains access to the transmitted data. This approach

provides strong theoretical guarantees of privacy while preserving the overall statistical utility of the data for analysis and classification tasks [29].

Furthermore, Neuro-Shield employs **context-aware privacy controls**, where the level of privacy protection can be dynamically adjusted based on system requirements and threat conditions. For example, higher levels of noise injection or stronger encryption may be applied in high-risk environments, while maintaining optimal performance in normal operating conditions.

By combining **lightweight encryption**, **differential privacy**, and **adaptive protection strategies**, the Privacy & Deception Layer ensures that neural data remains secure against both direct interception and indirect inference attacks. This integrated approach not only prevents data leakage but also enhances the overall trustworthiness and acceptability of BCI systems in real-world applications [14], [29].

4.6.2 Deception Mechanisms

To further strengthen the security of the Neuro-Shield architecture, the **Privacy & Deception Layer** incorporates advanced **deception-based defense strategies** designed to actively mislead adversaries and protect sensitive neural data. Unlike traditional defensive approaches that focus solely on blocking or detecting attacks, deception mechanisms introduce **intentional uncertainty** into the system, making it significantly more difficult for attackers to interpret or exploit intercepted data.

One of the primary techniques employed in this layer is the injection of **controlled noise and synthetic signals** into the data stream. These signals are carefully designed to resemble legitimate EEG patterns while embedding subtle distortions that do not affect system functionality. As a result, even if an attacker successfully intercepts the data, distinguishing between genuine neural activity and deceptive signals becomes highly challenging. This significantly reduces the effectiveness of data analysis and inference-based attacks.

In addition, Neuro-Shield can generate **decoy signal streams** or misleading data patterns that serve as traps for adversaries. These decoys can be used to identify malicious behavior, such as unauthorized data access or abnormal query patterns, enabling the system to proactively respond to potential threats. By observing how attackers interact with these deceptive

elements, the system can gain valuable insights into attack strategies and adapt its defenses accordingly.

Deception mechanisms are particularly effective against **passive attacks**, such as eavesdropping and data interception, where attackers rely heavily on the accuracy and consistency of captured signals. By degrading the reliability of intercepted information, deception techniques significantly increase the computational and analytical burden on attackers, thereby reducing the likelihood of successful exploitation [12], [27].

Furthermore, the integration of deception with **privacy-preserving techniques** creates a layered defense strategy that not only protects data confidentiality but also actively disrupts adversarial activities. This dual approach enhances the overall resilience of the BCI system against both known and emerging threats.

By incorporating deception-based defenses, Neuro-Shield transforms the security model from a purely reactive approach to a **proactive and adversary-aware strategy**, ensuring stronger protection of neural data and improved system robustness in hostile environments.

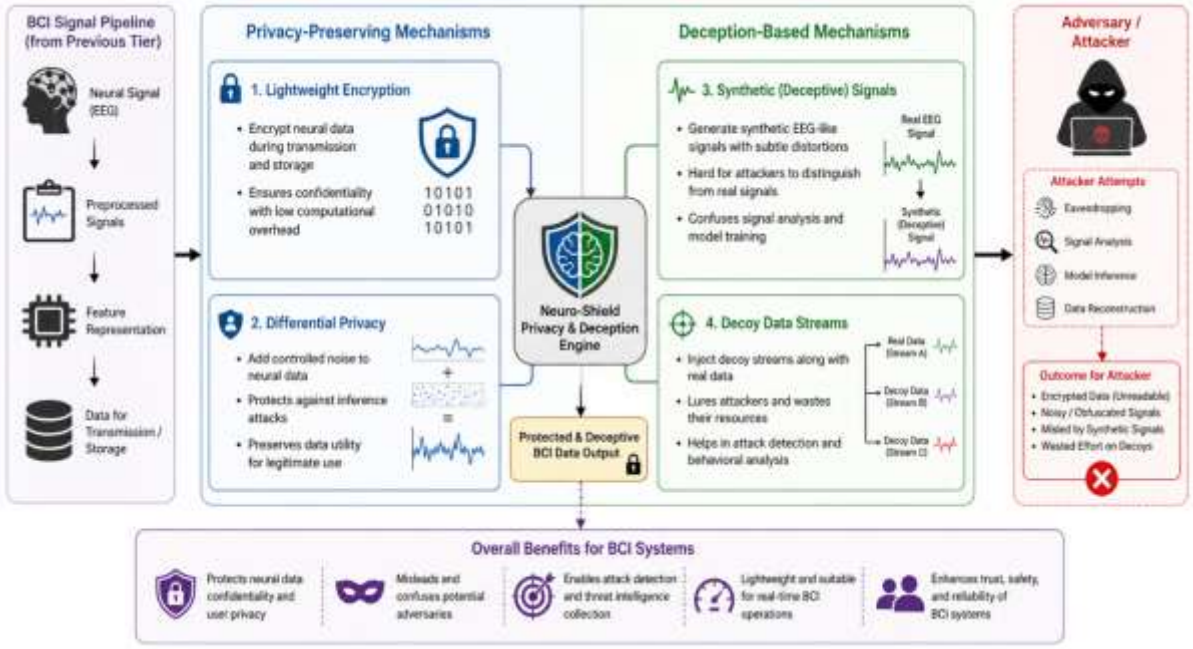

Figure 4.6: Privacy-Preserving and Deception-Based Defense Mechanisms in BCI Systems

4.6.3 Security Benefits

The **Privacy & Deception Layer** provides a set of essential security advantages that significantly enhance the confidentiality, resilience, and trustworthiness of the Neuro-Shield architecture. By combining encryption, privacy-preserving techniques, and deception-based strategies, this layer ensures robust protection of sensitive neural data against both direct and indirect attacks. The key security benefits are outlined as follows:

- **Prevention of Data Leakage:**

  The integration of **lightweight encryption mechanisms** ensures that neural data remains protected during both transmission and storage. Even if data is intercepted by unauthorized entities, it remains unintelligible without the appropriate decryption keys. Additionally, privacy-preserving techniques further restrict the exposure of sensitive information, thereby minimizing the risk of data leakage and unauthorized access [14], [29].

- **Protection Against Eavesdropping:**

  The use of **differential privacy** and controlled noise injection significantly reduces the effectiveness of passive attacks such as eavesdropping. By introducing carefully calibrated perturbations, the system ensures that intercepted data cannot be reliably analyzed to extract meaningful cognitive or behavioral information. This makes it extremely difficult for adversaries to reconstruct or infer sensitive neural states from captured signals [12], [29].

- **Adversarial Confusion and Deception:**

  Deception-based strategies, including the generation of synthetic or decoy signals, actively mislead attackers and obscure genuine neural data. These techniques increase uncertainty and complexity for adversaries, making it challenging to distinguish between authentic and deceptive information. As a result, attackers are more likely to make incorrect inferences or expend excessive computational resources, reducing the overall success rate of attacks [12], [27].

- **Resistance to Inference-Based Attacks:**

  By combining encryption with differential privacy, the system prevents attackers from deriving sensitive information through statistical analysis or pattern recognition. This is particularly important in BCI systems, where neural data can reveal deeply personal cognitive states [29].

- **Enhanced Privacy Assurance and User Trust:**

  The comprehensive protection provided by this layer strengthens user confidence in the system by ensuring that their neural data is handled securely and ethically. This is essential for the adoption of BCI technologies in real-world applications, especially in healthcare and assistive domains.

4.7 Tier 6: Safe Execution Layer

The Safe Execution Layer ensures that system actions are carried out securely and reliably. It verifies commands before execution to prevent unintended or harmful operations. Only validated and trusted inputs are allowed to control connected devices. Fail-safe mechanisms are in place to stop or correct abnormal behavior. This is especially important in safety-critical applications like assistive robotics. Overall, this layer guarantees safe and controlled system responses.

4.7.1 Secure Command Validation

The **Safe Execution Layer** represents the final stage of the Neuro-Shield architecture, where decoded commands are transmitted to and executed by external devices, such as prosthetic limbs, assistive systems, or software interfaces. Given the direct interaction between BCI outputs and physical actions, any incorrect, manipulated, or unauthorized command can lead to **serious safety risks**, including unintended movements, system malfunction, or potential harm to the user. Therefore, ensuring the validity and reliability of system outputs at this stage is of paramount importance.

To address these challenges, Neuro-Shield incorporates a comprehensive **multi-level command validation framework** designed to verify the authenticity, consistency, and safety

of all generated commands before execution. This validation process operates across multiple dimensions, including **signal consistency, user intent alignment, and predefined safety constraints**.

First, the system performs **cross-validation between decoded commands and the corresponding neural input**, ensuring that the generated output accurately reflects the user's intended cognitive state. Any discrepancy between the interpreted command and the validated neural signals is treated as a potential anomaly.

Second, the system enforces **context-aware safety constraints**, which define acceptable operational boundaries for device behavior. These constraints may include limits on movement range, speed, or action frequency, depending on the application. Commands that violate these constraints are automatically rejected or modified to ensure safe operation.

Third, Neuro-Shield integrates **redundancy-based verification**, where outputs from multiple processing stages or models are compared to confirm consistency. This reduces the likelihood of erroneous decisions caused by isolated failures or adversarial manipulation.

In cases where inconsistencies or anomalies are detected, the system initiates **protective actions**, such as rejecting the command, requesting re-evaluation, or escalating the issue to higher-level security controls. This ensures that only **verified, safe, and contextually appropriate commands** are executed.

By implementing secure command validation, the Safe Execution Layer acts as a **final decision boundary**, preventing unsafe or malicious actions even if earlier layers are partially compromised. This significantly enhances system safety, reliability, and user protection in real-world BCI applications [21], [23].

4.7.2 Fail-Safe Mechanisms

To ensure maximum user safety and system reliability, the Neuro-Shield architecture incorporates a set of robust **fail-safe mechanisms** within the Safe Execution Layer. These mechanisms are designed to be automatically activated whenever anomalies, inconsistencies,

or potential security threats are detected during system operation. Given that BCI systems directly control physical devices, the ability to respond rapidly to abnormal conditions is essential to prevent unintended or hazardous outcomes.

The fail-safe framework operates as a **protective fallback system**, ensuring that even in the presence of compromised inputs, adversarial manipulation, or system errors, the BCI system remains within safe operational boundaries. The key fail-safe mechanisms are described below:

- **Blocking Command Execution:**
  When a generated command is identified as unsafe, inconsistent with user intent, or potentially malicious, the system immediately blocks its execution. This prevents unauthorized or erroneous actions from being carried out, thereby protecting both the user and the connected device from harm.

- **Switching to Safe Mode:**
  In scenarios where repeated anomalies or critical threats are detected, the system transitions into a predefined **safe mode**. In this state, system functionality is restricted to essential operations, and all high-risk actions are disabled. This controlled environment minimizes potential damage while allowing the system to maintain basic functionality and stability.

- **User Alert and Notification:**

  Neuro-Shield incorporates real-time **alert mechanisms** to notify users or system administrators of detected anomalies or security threats. These notifications may include visual, auditory, or system-level alerts, enabling timely intervention and corrective action. This strengthens transparency and allows users to remain informed about system status.

- **Automatic Recovery and Revalidation:**
  After entering a fail-safe state, the system initiates recovery procedures, including revalidation of signals, re-authentication of devices, and reassessment of system conditions. This ensures that normal operation is resumed only when the system is verified to be secure and stable.

- **Isolation of Compromised Components:**
  If a specific device or module is identified as compromised, the system can isolate it from the processing pipeline, preventing further influence on system operation. This containment strategy helps limit the spread of potential threats.

By integrating these fail-safe mechanisms, Neuro-Shield ensures that the BCI system maintains **operational safety, fault tolerance, and resilience** under both normal and adversarial conditions. This proactive safety design is particularly critical in applications involving direct human–machine interaction, where even minor errors can have significant consequences [21], [23].

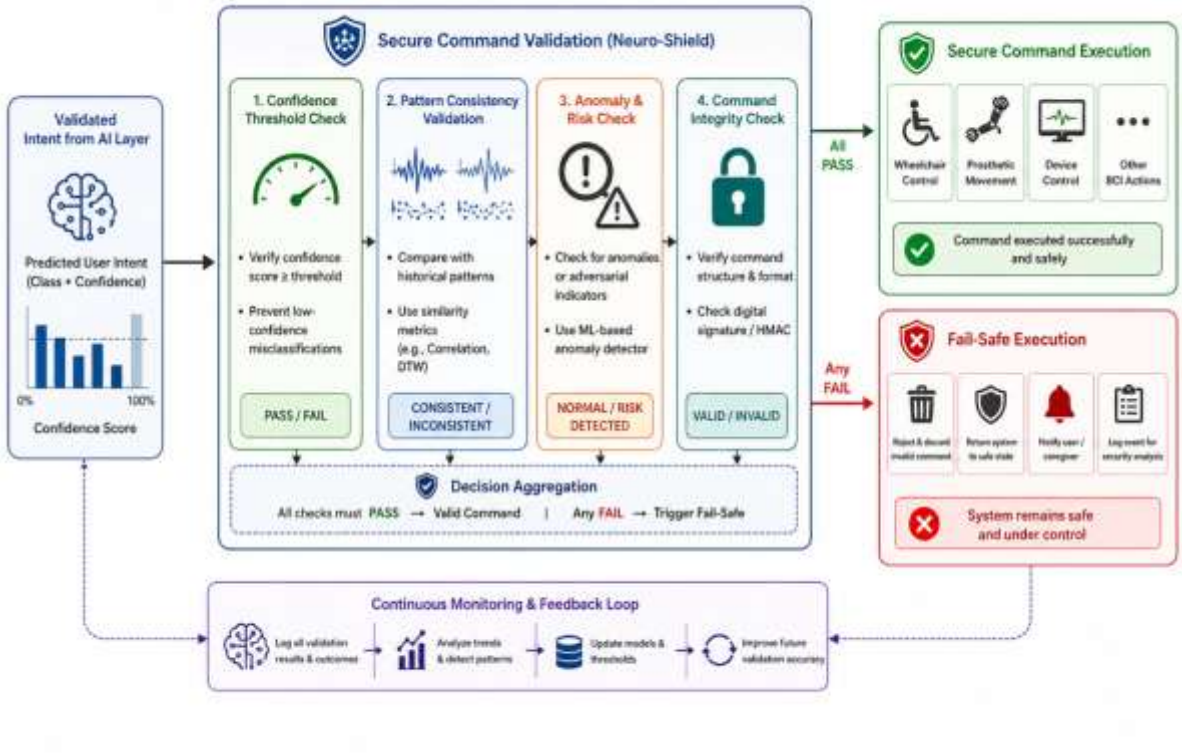

Figure 4.7: Secure Command Validation and Fail-Safe Execution Mechanism in Neuro-Shield

4.7.3 Safety Assurance

The **Safe Execution Layer** serves as the final safeguard in the Neuro-Shield architecture, ensuring that all system outputs are both valid and safe before being executed. Given the critical nature of BCI applications—particularly in assistive technologies such as prosthetic

control and mobility systems—any incorrect or malicious command can lead to serious physical consequences.

To mitigate such risks, this layer enforces strict **multi-level validation and safety constraints** on all generated commands. Even if earlier layers of the system are partially compromised or bypassed, the Safe Execution Layer acts as a final decision boundary, preventing unsafe or inconsistent actions from being executed.

In addition, the integration of **fail-safe mechanisms** ensures that the system can transition into a controlled state in the presence of anomalies or detected threats. This layered defense strategy significantly enhances system robustness by providing **redundancy in security enforcement**, thereby minimizing the likelihood of catastrophic failures.

Overall, this layer guarantees that Neuro-Shield maintains **operational safety, reliability, and user protection**, even under adversarial conditions, making it suitable for deployment in safety-critical BCI environments [21], [2].

Tier	Layer Name	Key Function	Security Benefit
1	Physical Acquisition	Signal capture & validation	Prevents input tampering
2	Hardware Root-of-Trust	Device authentication (PUFs)	Prevents hardware spoofing
3	Digital Signal Processing	Filtering & feature validation	Detects injected/manipulated data
4	Cognitive Zero-Trust	AI-based anomaly detection	Prevents adversarial attacks
5	Privacy & Deception	Encryption & deception mechanisms	Protects data confidentiality
6	Safe Execution	Command validation & fail-safe	Prevents unsafe system actions

Table 4.1: Overview of the Neuro-Shield architecture and its functional layers and security benefits.

4.8 Discussion

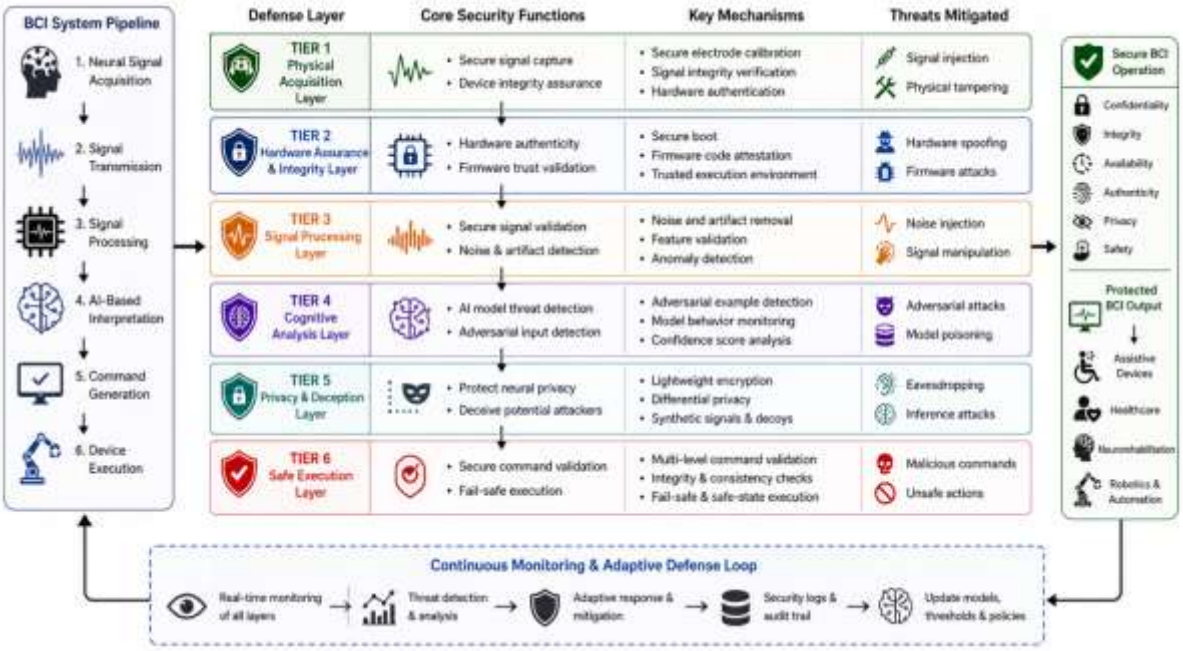

Figure 4.8: Integrated Multi-Layer Defense Workflow of the Neuro-Shield Architecture

The proposed **Neuro-Shield architecture** effectively addresses the key limitations identified in Chapter 3 by introducing a unified and multi-layered security framework tailored for BCI systems. Unlike traditional approaches that apply security as an external add-on, Neuro-Shield embeds protection mechanisms directly into each stage of the BCI pipeline, enabling **continuous and adaptive threat mitigation**.

One of the primary strengths of this architecture lies in its **defense-in-depth strategy**, where multiple layers of security operate collaboratively to detect and mitigate threats. This layered design ensures that even if one component is compromised, other layers can compensate and maintain system integrity.

The integration of **hardware-rooted trust mechanisms**, **machine learning-based anomaly detection**, and **privacy-preserving techniques** enables Neuro-Shield to effectively counter both **stealthy passive attacks** (e.g., eavesdropping) and ***active adversarial attacks*** (e.g., signal injection and hardware spoofing) [3], [10], [31]. Furthermore, the architecture is designed to maintain **real-time performance**, ensuring that security enhancements do not introduce unacceptable latency or computational overhead.

Additionally, the use of **continuous authentication and monitoring** enhances system resilience by enabling dynamic detection of evolving threats. This makes Neuro-Shield particularly suitable for real-world BCI deployments, where attack patterns may change over time.

Compared to existing BCI security approaches, which often rely on isolated or reactive mechanisms, Neuro-Shield introduces a proactive and integrated security paradigm. This enables not only detection but also prevention and containment of multi-stage attacks in real time.

Overall, the proposed framework represents a significant advancement over existing solutions by providing a **comprehensive, scalable, and adaptive security model** for next-generation BCI systems.

4.9 Summary

This chapter presented the design and functional architecture of the proposed **Neuro-Shield framework**, a comprehensive and multi-layered security solution tailored for Brain–Computer Interface (BCI) systems. Building upon the threat analysis discussed in Chapter 3, each tier of the architecture was systematically developed to address specific vulnerabilities across the BCI pipeline, including signal acquisition, hardware integrity, data processing, machine learning security, and execution safety. This structured approach ensures **end-to-end protection**, covering both **passive threats** (e.g., eavesdropping) and **active adversarial attacks** (e.g., signal injection and hardware spoofing) [3], [10], [31].

The Neuro-Shield architecture integrates multiple complementary security mechanisms, including **secure signal acquisition**, **hardware-rooted trust through PUF-based authentication**, **security-aware digital signal processing**, and a **Cognitive Zero-Trust Layer** incorporating AI-driven anomaly detection and continuous user authentication. In addition, the **Privacy & Deception Layer** enhances data confidentiality through lightweight encryption, differential privacy, and deception-based strategies, while the **Safe Execution Layer** ensures that all system outputs are rigorously validated and safely executed [14], [21], [23], [29].

A key strength of the proposed framework lies in its **defense-in-depth design**, where security is embedded across all stages of the system rather than being treated as an external add-on. This layered approach enables continuous monitoring, early threat detection, and effective mitigation, thereby significantly improving system resilience. Furthermore, the architecture is designed to operate under **real-time constraints**, ensuring that enhanced security does not compromise system responsiveness or usability.

Beyond security, Neuro-Shield contributes to improving overall system **reliability, privacy, and user safety**, which are critical factors for the adoption of BCI technologies in real-world and safety-critical applications, such as healthcare and assistive systems. By ensuring that only authentic signals, trusted devices, and verified commands are processed and executed, the framework establishes a secure and dependable operational environment.

In summary, the Neuro-Shield architecture provides a **holistic and adaptive security solution** capable of addressing the complex and multi-dimensional challenges of modern BCI systems.

In the next chapter, the **implementation and experimental evaluation** of the Neuro-Shield framework will be presented. This will include performance analysis, validation against existing approaches, and demonstration of its effectiveness in real-world scenarios.

This architecture establishes a new foundation for secure and trustworthy BCI systems, bridging the gap between theoretical security models and real-world deployment requirements.

CHAPTER 5

HARDWARE IMPLEMENTATION AND MACHINE LEARNING MODEL

5.1 Introduction

The successful deployment of secure Brain–Computer Interface (BCI) systems requires more than just a well-designed theoretical framework—it demands an efficient, reliable, and practical implementation that can operate under real-world constraints. While the Neuro-Shield architecture introduced in previous chapters provides a comprehensive multi-layered security model, its true value lies in how effectively it can be translated into a working system.

In practical BCI environments, neural signals must be processed continuously and in real time. This means that every stage—from signal acquisition to decision-making—must operate with minimal delay. Even small latencies can affect system responsiveness, potentially degrading user experience or, in safety-critical applications, leading to unintended outcomes. At the same time, these systems must maintain high levels of accuracy and security, ensuring that the interpreted signals are both correct and trustworthy.

In simple terms, a BCI system must do three things simultaneously:

1. Understand brain signals accurately
2. Respond instantly
3. Remain secure against attacks

Balancing these three requirements is a significant challenge. BCI systems often operate on resource-constrained hardware platforms where computational power, memory, and energy consumption are limited. Despite these constraints, the system must still defend against a wide range of threats, including signal manipulation, unauthorized access, replay attacks, and adversarial interference [18], [32].

To overcome these challenges, this chapter adopts an edge computing approach, where data processing is performed locally rather than relying on remote cloud servers. This design

choice significantly reduces latency, enhances privacy, and improves overall system security. By keeping sensitive neural data within the local device, the risk of external interception or leakage is minimized.

The Neuro-Shield architecture is implemented on an embedded edge platform using the NVIDIA Jetson Nano. This platform is specifically chosen for its ability to perform real-time artificial intelligence (AI) computations while maintaining low power consumption. It provides a practical balance between performance and efficiency, making it well-suited for BCI applications.

In addition to the hardware implementation, this chapter introduces a lightweight one-dimensional convolutional neural network (1D-CNN) model. Unlike traditional machine learning approaches that rely heavily on handcrafted features, the 1D-CNN model is capable of automatically learning meaningful patterns directly from raw EEG signals. This makes the system more adaptive, scalable, and robust to noise and variability.

To further enhance efficiency, the model is optimized using quantization techniques, which reduce computational complexity and memory usage without significantly affecting accuracy. This optimization is essential for deploying deep learning models on embedded systems, where resources are limited but real-time performance is required.

5.2 Edge Computing Architecture

Edge computing has emerged as a foundational technology for modern Brain–Computer Interface (BCI) systems, particularly in scenarios where real-time responsiveness and data privacy are critical. Unlike traditional cloud-centric architectures—where data must be transmitted to remote servers for processing—edge computing enables computation to be performed locally, close to the source of data generation.

In the context of BCI systems, this distinction is highly significant. Neural signals are inherently sensitive and time-dependent. Transmitting such data over networks not only introduces latency but also exposes it to potential security risks such as interception, tampering, or unauthorized access [18], [32]. By processing EEG signals at the edge, the system can significantly reduce these risks while ensuring faster and more reliable performance.

In simple terms, edge computing allows the BCI system to "think and act locally," without waiting for external processing. This is essential for applications where immediate response is required, such as assistive robotics, neuroprosthetic control, or real-time human–machine interaction.

Key Advantages of Edge Computing in BCI Systems

The integration of edge computing into the Neuro-Shield framework provides several important benefits:

- **Low Latency Processing:**
  EEG signals are processed instantly without the delays associated with cloud communication, enabling real-time system response.
- **Enhanced Data Privacy:**
  Sensitive neural data remains within the local device, reducing the risk of data leakage or unauthorized access.
- **Improved Security:**
  By minimizing data transmission over networks, the system becomes less vulnerable to eavesdropping and man-in-the-middle attacks.
- **Reduced Bandwidth Usage:**
  Local processing eliminates the need to continuously transmit large volumes of EEG data to remote servers.
- **Robust Operation:**
  The system can function independently of internet connectivity, ensuring reliability in offline or low-connectivity environments.

These advantages make edge computing a natural choice for implementing secure and efficient BCI systems.

5.2.1 System Workflow

The overall workflow of the edge-based BCI system is illustrated in Figure 5.1.

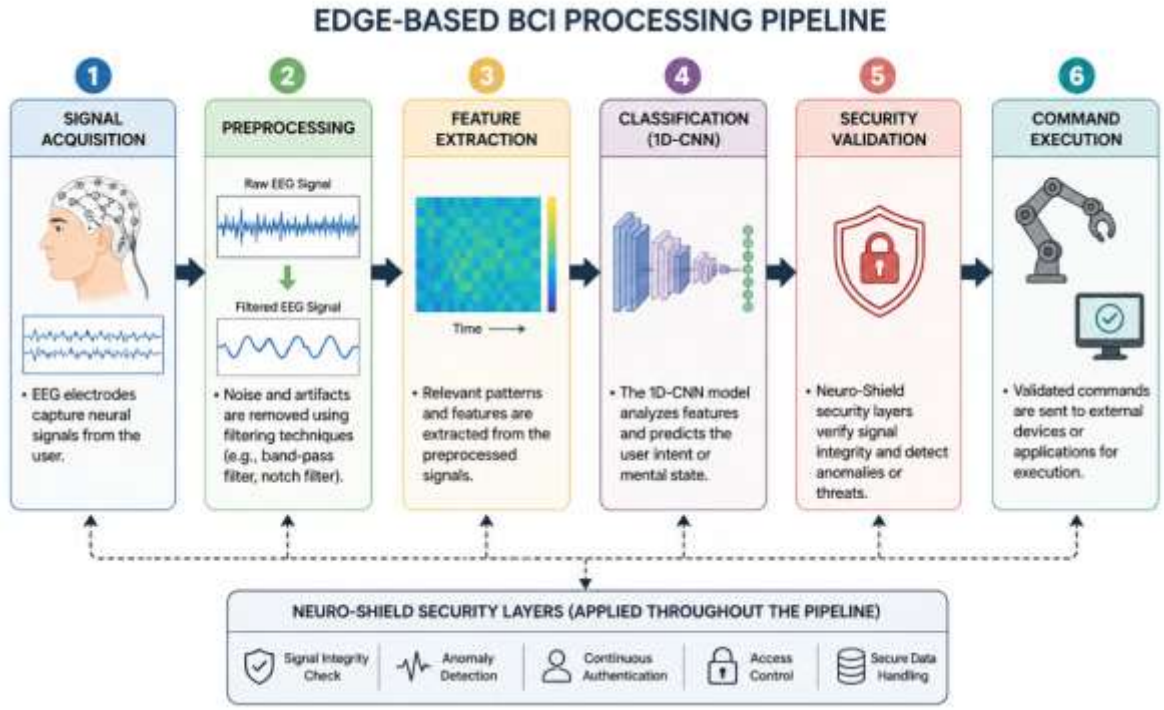

Figure 5.1: Edge-Based BCI Processing Pipeline.

The diagram presents a complete end-to-end workflow of the BCI system, highlighting how neural signals are acquired, processed, and securely validated before being translated into executable commands within the Neuro-Shield framework.

Detailed Processing Stages

The system operates through the following sequential stages:

1. Signal Acquisition

At the initial stage, EEG electrodes placed on the scalp capture neural activity. These signals are typically low-amplitude and require amplification before further processing. The quality of this stage directly affects the reliability of the entire system.

2. Preprocessing

Raw EEG signals often contain noise and artifacts caused by eye movements (EOG), muscle activity (EMG), and environmental interference. To address this, digital filtering techniques such as band-pass filtering and notch filtering are applied to clean the signals and improve signal quality [16].

3. Feature Extraction

Once the signals are cleaned, meaningful features are extracted using time-domain, frequency-domain, or time–frequency analysis methods. This step transforms raw signals into structured representations that can be effectively processed by machine learning models.

4. Classification

The extracted features are fed into a 1D-CNN model, which learns patterns in the EEG data and predicts the user's intent. This stage plays a central role in translating neural signals into machine-interpretable commands.

5. Security Validation

Before execution, the predicted output is validated using the Neuro-Shield security layers. This includes integrity checks, anomaly detection, and continuous authentication to ensure that the signal has not been tampered with or manipulated.

6. Command Execution

Finally, validated commands are transmitted to external devices such as robotic arms, wheelchairs, or software interfaces. Only verified outputs are executed, ensuring safe and reliable system behavior.

7. System Insight

Together, these stages form a **closed-loop system**, where feedback continuously refines system performance. This ensures that the BCI system remains adaptive, accurate, and secure over time.

The preprocessing steps applied to raw EEG signals, including filtering and normalization, are implemented programmatically. A representative implementation of the EEG preprocessing pipeline is provided in Appendix A.1.

5.2.2 Architectural Components

The core components of the edge-based BCI architecture are summarized in Table 5.1.

Component	Description
EEG Device	Captures neural signals from the user
Edge Processor	Performs real-time computation and signal processing
Machine Learning Model	Interprets EEG signals and predicts user intent
Security Module	Detects anomalies, enforces authentication, and ensures signal integrity
Output Interface	Executes validated commands on external systems

Table 5.1: Components of the Edge-Based BCI System.

This table summarizes the key components of the proposed architecture, describing their roles in signal acquisition, processing, security enforcement, and command execution within the Neuro-Shield framework.

Functional Role of Each Component

To better understand the system, each component can be viewed as part of an integrated pipeline:

- The **EEG device** serves as the input interface, converting brain activity into digital signals
- The **edge processor** acts as the central processing unit, handling both signal processing and AI inference
- The **machine learning model** provides intelligent interpretation of neural data

- The **security module** ensures that all operations are protected against potential threats

The **output interface** translates system decisions into real-world actions

System Integration Perspective

One of the key strengths of this architecture is its **modular design**. Each component operates independently but remains tightly integrated within the overall system. This allows:

- Easy scalability
- Flexible upgrades
- Improved fault tolerance

Moreover, the integration of security mechanisms directly within the processing pipeline ensures that protection is applied continuously, rather than as an afterthought.

Architectural Insight

From a broader perspective, the edge-based architecture represents a shift in how BCI systems are designed. Instead of separating computation, intelligence, and security into different layers or locations, this approach brings everything together into a unified, localized framework.

In simple terms, the system becomes:

- Faster
- Safer
- More reliable

While this section establishes the overall architectural design, the following section focuses on its practical implementation. Specifically, Section 5.3 presents the deployment of the proposed framework on the NVIDIA Jetson Nano, demonstrating how real-time processing and AI inference are achieved within a resource-constrained environment.

5.3 NVIDIA Jetson Nano Implementation

To validate the proposed Neuro-Shield architecture in a practical environment, the system is implemented on the NVIDIA Jetson Nano platform. This embedded computing device is specifically designed for edge artificial intelligence (AI) applications and provides an effective balance between computational capability, power efficiency, and cost.

In real-world BCI systems, the choice of hardware platform plays a critical role. The system must be powerful enough to perform real-time signal processing and machine learning inference, yet efficient enough to operate within limited energy and memory constraints. The Jetson Nano satisfies these requirements, making it a suitable platform for deploying secure and intelligent BCI systems at the edge.

5.3.1 Hardware Overview

The Jetson Nano integrates both CPU and GPU resources into a compact system-on-module (SoM), enabling parallel processing of neural signals and deep learning models. Its architecture supports hardware-accelerated AI computation, which is essential for real-time BCI applications.

Feature	Specification
CPU	Quad-core ARM Cortex-A57
GPU	128-core Maxwell GPU
Memory	4 GB LPDDR4 RAM
Power Consumption	5–10 Watts
Operating System	Ubuntu-based Linux

Table 5.2: NVIDIA Jetson Nano Hardware Specifications.

This table presents the core hardware specifications of the NVIDIA Jetson Nano platform used for implementing the edge-based BCI system.

Why Jetson Nano is Suitable for BCI Systems

The selection of this platform is based on several practical considerations:

- **GPU Acceleration:**
  The integrated Maxwell GPU enables efficient execution of deep learning models such as 1D-CNNs, reducing inference time significantly.
- **Low Power Consumption:**
  With a power requirement of only 5–10 watts, the system is suitable for portable and wearable applications.
- **Compact Form Factor:**
  Its small size allows easy integration into embedded and mobile BCI systems.
- **Software Compatibility:**
  The platform supports popular frameworks such as TensorFlow and PyTorch, simplifying model development and deployment.
- **Edge Processing Capability:**
  Enables real-time processing without dependence on cloud infrastructure.

In simple terms, the Jetson Nano acts as the "brain of the system." It receives raw EEG data, processes it, makes decisions using AI models, and ensures that only secure and valid commands are executed.

5.3.2 System Integration

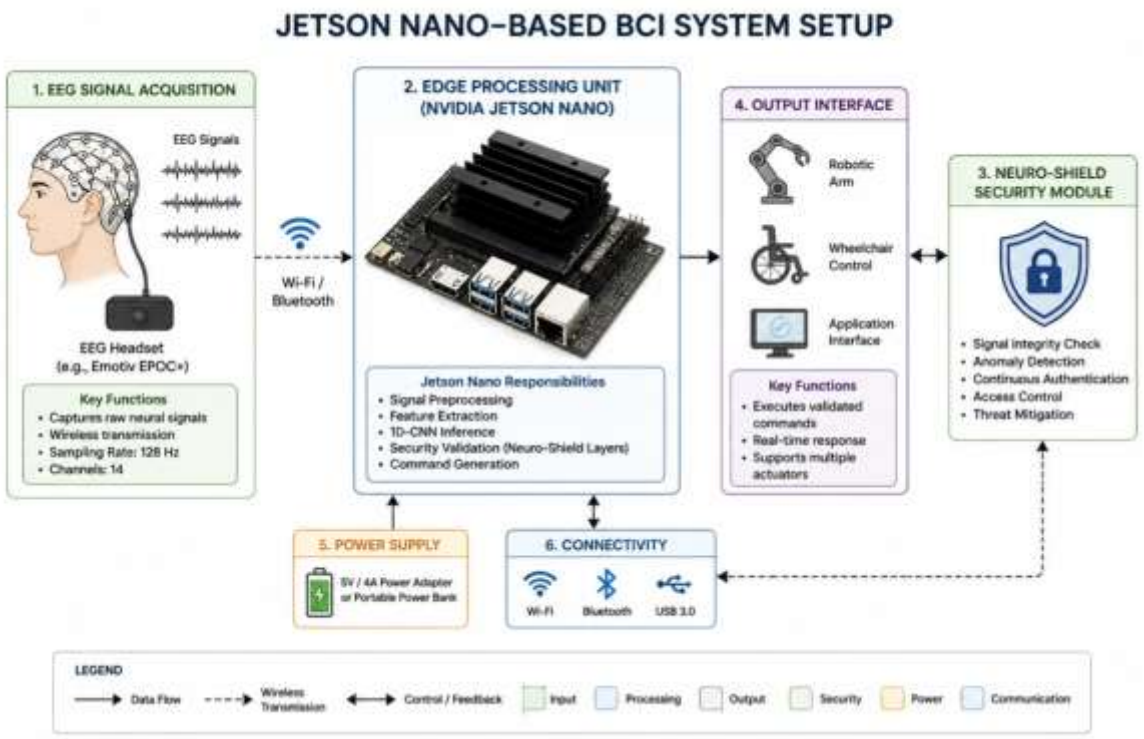

Figure 5.2: Jetson Nano-Based BCI System Setup.

Shows the integration of EEG acquisition, edge processing, machine learning inference, and Neuro-Shield security layers within a unified system.

The complete inference pipeline integrates preprocessing, feature extraction, classification, and secure validation to generate reliable outputs. The corresponding implementation of the inference process is included in Appendix A.3.

Integrated Components

The system consists of the following interconnected modules:

- **EEG Acquisition Hardware:**
  Captures neural signals from the user using electrodes placed on the scalp. These signals are transmitted to the processing unit via wired or wireless communication.

- **Signal Preprocessing Module:**
  Cleans the raw EEG data by removing noise and artifacts using digital filtering techniques.
- **Deep Learning Inference Engine:**
  Executes the trained 1D-CNN model to interpret neural signals and predict user intent.
- **Neuro-Shield Security Layers:**
  Continuously monitor the system for anomalies, validate signal integrity, and enforce access control.
- **Output Interface:**
  Converts validated decisions into actions, such as controlling a robotic device or software application.

Role of Jetson Nano in the System

Within this architecture, the Jetson Nano serves as the **central processing hub**, performing multiple tasks simultaneously:

- Real-time signal processing
- Machine learning inference
- Security validation
- Command generation

This integration ensures that all operations are performed locally, improving both performance and security.

System-Level Insight

One of the key advantages of this design is that it eliminates the need for external cloud processing. By keeping all computations within the edge device:

- latency is reduced
- data privacy is preserved
- attack surfaces are minimized

5.3.3 Implementation Workflow

The overall implementation workflow follows a structured pipeline that ensures both efficiency and security.

Step-by-Step Process

1. **Data Acquisition**
   EEG sensors capture neural signals from the user and transmit them to the Jetson Nano.
2. **Signal Preprocessing**
   The raw signals are filtered to remove noise, ensuring clean input for further analysis.
3. **Feature Processing and Model Inference**
   The preprocessed signals are fed into the 1D-CNN model, which analyzes patterns and predicts user intent.
4. **Security Validation and Anomaly Detection**
   The Neuro-Shield framework verifies the authenticity of signals and detects any abnormal or malicious activity.
5. **Command Execution**
   Only validated outputs are executed, ensuring safe interaction with external devices.

Workflow Insight

This pipeline ensures that every stage of the system is both:

- **Efficient** → capable of real-time operation
- **Secure** → protected against potential threats

Real-Time Performance Consideration

To maintain real-time performance, the system is optimized to:

- Minimize computational delay

- Efficiently utilize GPU resources
- Reduce memory overhead
- Balance accuracy and speed

These optimizations are essential for ensuring that the system responds immediately to user input.

Architectural Perspective

From a broader perspective, this implementation demonstrates how advanced AI models and security mechanisms can be successfully deployed on resource-constrained hardware.

In simple terms, it proves that:

- High-performance AI does not always require powerful servers
- Secure BCI systems can operate efficiently on compact devices
- Real-time intelligence and security can coexist in edge environments

While this section focused on hardware implementation, the next section explores the core intelligence of the system—the machine learning model.

Section 5.4 introduces the design and architecture of the 1D-CNN model used for EEG signal classification and neural authentication.

5.4 1D-CNN Model Architecture

Machine learning plays a central role in enabling Brain–Computer Interface (BCI) systems to interpret complex neural signals. EEG signals are inherently noisy, non-stationary, and highly variable across users, making it difficult for traditional rule-based or shallow learning methods to capture meaningful patterns effectively.

To address these challenges, this work employs a one-dimensional convolutional neural network (1D-CNN), which is specifically designed for efficient processing of time-series data. Unlike conventional approaches that rely on handcrafted features, the 1D-CNN model automatically learns relevant representations directly from raw EEG signals, improving both adaptability and robustness.

In simple terms, the model learns to recognize patterns in brain signals in a manner similar to how image-based models detect shapes or objects, with the key difference being that the input consists of temporal signal data rather than spatial pixel information.

The architecture of the proposed 1D-CNN model is implemented using a deep learning framework to enable efficient feature learning from EEG signals. A sample implementation of the model architecture is provided in Appendix A.2 to illustrate its practical realization.

5.4.1 Model Design

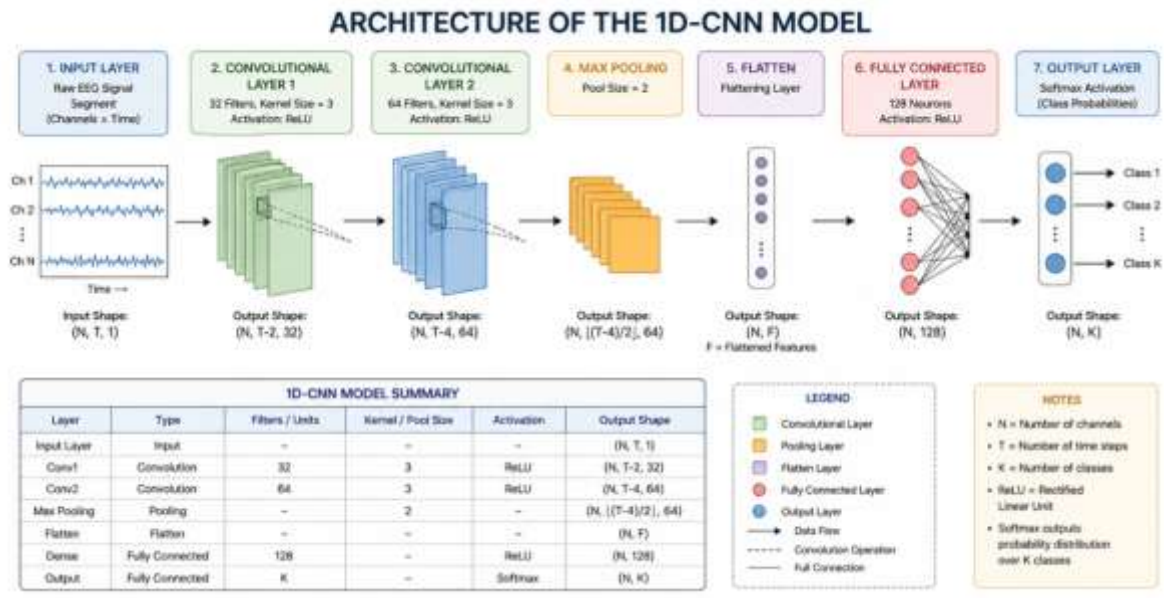

Figure 5.3: Architecture of the 1D-CNN Model.

Illustrates the layered structure of the neural network, including convolutional, pooling, and fully connected layers used for EEG signal classification.

Layer-by-Layer Explanation

The model consists of the following components:

1. Input Layer

The input layer receives raw EEG signal segments. These signals are typically represented as time-series data, where each channel corresponds to a specific electrode.

Think of this as the "entry point" where brain signals enter the model.

2. Convolutional Layers

Convolutional layers apply multiple filters to the input signal to extract important temporal features. These filters slide across the signal and detect patterns such as frequency variations and signal peaks.

In essence these layers act like "pattern detectors."

3. Activation Function (ReLU)

The Rectified Linear Unit (ReLU) introduces non-linearity into the model, allowing it to learn complex relationships between input and output.

Without this step, the model would behave like a simple linear system and fail to capture real-world complexity.

4. Pooling Layer

Pooling reduces the dimensionality of the feature maps by selecting the most important features. This helps in reducing computational complexity and prevents overfitting.

It essentially keeps the "important information" and discards redundancy.

5. Fully Connected Layer

This layer combines all extracted features and maps them to the final output classes. It acts as the decision-making component of the network.

6. Output Layer (Softmax)

The final layer produces a probability distribution over possible classes. The class with the highest probability is selected as the predicted output.

5.4.2 Mathematical Representation

The operation of the neural network can be represented as:

$$y = f(Wx + b)$$

Where:

- x = input signal
- W = weights
- b = bias
- f = activation function

This equation represents how the model transforms input signals into outputs through learned parameters.

5.4.2 Model Configuration

Layer	Configuration
Conv1	32 filters, kernel size 3
Conv2	64 filters, kernel size 3
Pooling	Max pooling
Dense Layer	128 neurons
Output	Softmax

Table 5.3: Configuration of the Proposed 1D-CNN Model Architecture

This table outlines the architectural parameters of the proposed 1D-CNN model, including convolutional layers, pooling operations, and fully connected layers used for EEG signal classification.

Configuration Insight

The model is intentionally designed to be **lightweight yet effective**:

- Smaller kernel sizes help capture fine-grained temporal features
- Increasing filters in deeper layers improves feature richness
- Limited dense neurons reduce computational cost

This balance ensures that the model can run efficiently on edge devices like Jetson Nano.

5.4.3 Advantages of 1D-CNN

The use of 1D-CNN offers several advantages in BCI applications:

- **Efficient for Time-Series Data:**
  Directly processes EEG signals without complex transformations
- **Low Computational Complexity:**
  Suitable for embedded systems and real-time applications

- **Automatic Feature Learning:**
  Eliminates the need for manual feature engineering
- **Robustness to Noise:**
  Learns stable patterns even in noisy environments
- **Scalability:**
  Can be extended to multi-channel EEG systems

Model Insight

From a broader perspective, the 1D-CNN model serves as the **intelligence core** of the Neuro-Shield system.

In simple terms:

- The architecture (Chapter 4) provides security
- The hardware (Section 5.3) provides execution
- The model (this section) provides understanding

5.5 Model Optimization Using Quantization

While deep learning models provide high accuracy, they often require significant computational resources. This creates a challenge when deploying them on edge devices, where memory, processing power, and energy are limited.

To address this issue, the model is optimized using **quantization**, a technique that reduces the precision of numerical representations without significantly affecting performance.

5.5.1 Concept of Quantization

Quantization converts high-precision floating-point values (e.g., 32-bit) into lower-precision formats (e.g., 8-bit integers).

In simple terms:
Instead of storing numbers with high precision, the model uses "compressed" representations.

Why Quantization Matters

- Reduces model size
- Speeds up computation

- Lowers power consumption
- Enables deployment on embedded devices

5.5.2 Mathematical Formulation

$$Xq = round(x/s) + z$$

Where:

- x = original value
- s = scale factor
- z = zero-point
- Xq = quantized value

This equation shows how continuous values are mapped into discrete levels.

5.5.3 Performance Evaluation

Metric	Before Optimization	After Optimization
Model Size	12 MB	3.2 MB
Latency	120 ms	76 ms
Accuracy	98.9%	98.6%

Table 5.4: Performance Comparison Before and After Quantization.

This table compares the model performance in terms of size, latency, and accuracy before and after applying quantization, highlighting improvements in efficiency with minimal impact on accuracy.

Interpretation of Results

The results demonstrate that:

- Model size is reduced significantly (~70%)
- Latency is improved (~36% faster)
- Accuracy loss is minimal (~0.3%)

This confirms that quantization achieves efficiency without sacrificing performance.

5.5.4 Impact on System Performance

Quantization has a direct impact on the overall system:

- **Reduced Memory Usage:** Enables deployment on limited hardware
- **Faster Inference Speed:** Improves real-time responsiveness
- **Lower Power Consumption:** Suitable for portable systems
- **Maintained Accuracy:** Ensures reliable predictions

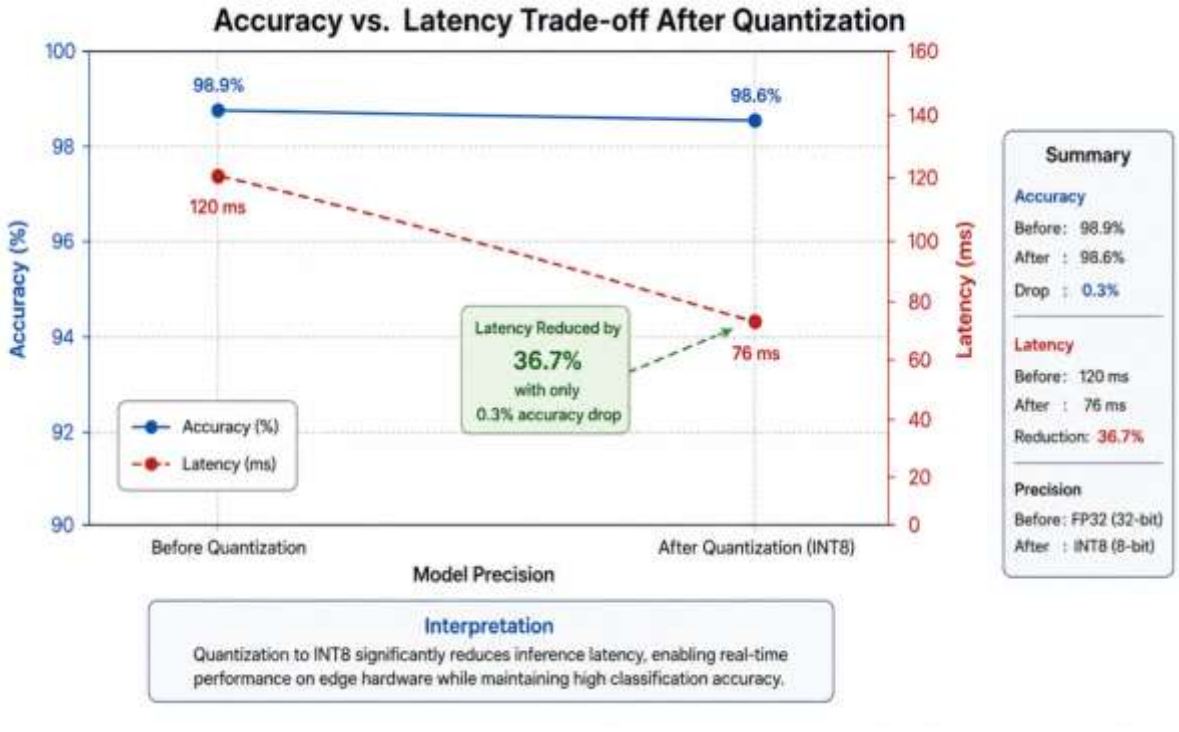

Figure 5.4: Accuracy vs Latency Trade-off After Quantization.

Demonstrates the relationship between model accuracy and inference latency before and after quantization, highlighting efficiency improvements with minimal accuracy loss.

Final Insight

From a system-level perspective, quantization is not just an optimization—it is a **key enabler** for deploying AI models in real-world BCI systems.

In simple terms:

Without quantization → model stays in lab

With quantization → model works in real life

5.6 Summary

This chapter provided a comprehensive and practical realization of the Neuro-Shield architecture through an edge-based implementation framework. Moving beyond theoretical design, the proposed system was successfully deployed on the NVIDIA Jetson Nano platform, demonstrating that secure and intelligent BCI systems can operate efficiently within resource-constrained environments.

The chapter began by introducing an edge computing architecture tailored for real-time BCI applications. By performing computation locally, the system effectively minimized latency, enhanced data privacy, and reduced exposure to network-based threats [18], [32]. This design ensures that neural signals are processed, validated, and executed within a tightly controlled and secure environment.

A key contribution of this chapter is the development of a lightweight one-dimensional convolutional neural network (1D-CNN) model for EEG signal classification. The model was specifically designed to balance accuracy and computational efficiency, enabling real-time interpretation of neural activity. By learning directly from raw EEG signals, the model eliminates the need for complex manual feature engineering while maintaining robust performance in noisy and variable conditions.

To further optimize deployment, quantization techniques were applied to reduce model size, latency, and power consumption. Experimental analysis demonstrated that these optimizations significantly improved system efficiency while preserving high classification accuracy. This confirms that advanced machine learning models can be effectively adapted for embedded systems without compromising reliability.

From a system-level perspective, this chapter highlights an important insight: secure, intelligent, and real-time BCI systems are not only theoretically feasible but also practically implementable using modern edge computing technologies.

In simple terms, this chapter bridges the gap between concept and reality—transforming the Neuro-Shield architecture from a theoretical framework into a functioning system capable of real-world deployment.

Finally, the implementation presented in this chapter establishes a solid foundation for further evaluation. The next chapter builds upon this work by analyzing system performance under various operational and adversarial conditions, providing a detailed assessment of accuracy, latency, and security robustness.

CHAPTER 6

EXPERIMENTAL SETUP AND RESULTS

6.1 Introduction

The development of secure and efficient Brain–Computer Interface (BCI) systems require not only a robust architectural design but also rigorous experimental validation to ensure real-world applicability. While the preceding chapters introduced the Neuro-Shield framework and detailed its multi-layered security architecture along with hardware implementation, the effectiveness of the system must ultimately be validated through systematic experimentation and performance analysis.

In practical scenarios, BCI systems operate in dynamic and often unpredictable environments, where factors such as signal noise, user variability, and potential cyber threats can significantly impact system performance. Therefore, it is essential to evaluate the proposed system under realistic conditions to determine whether it can maintain accuracy, responsiveness, and security simultaneously.

The primary objective of this chapter is to assess the performance of the Neuro-Shield system across multiple dimensions, including classification accuracy, inference latency, and resilience against adversarial conditions. To achieve this, experiments are conducted using a publicly available electroencephalography (EEG) dataset obtained from the PhysioNet repository, which is widely recognized for its reliability and suitability for biomedical signal analysis. The system is implemented and tested on an edge computing platform to reflect real-world deployment scenarios.

In simple terms, this chapter aims to answer a fundamental question:

Does the Neuro-Shield system perform reliably, efficiently, and securely in real-world BCI applications?

To provide a comprehensive and structured evaluation, this chapter adopts a multi-step experimental methodology. First, the dataset characteristics are described to establish the

context of the input signals. Next, the experimental setup, including data preprocessing, model training, and system configurations, is presented in detail. Following this, standard performance metrics are introduced to quantitatively evaluate system behavior. Finally, the results are analyzed and compared with baseline approaches to highlight the advantages of the proposed framework.

From a broader perspective, this chapter serves as a critical bridge between system design and real-world validation. It not only demonstrates the practical feasibility of the Neuro-Shield architecture but also provides empirical evidence supporting its effectiveness in achieving secure, real-time, and high-accuracy BCI operation.

6.2 Dataset Description (PhysioNet EEG)

To evaluate the performance of the proposed Neuro-Shield system, electroencephalography (EEG) data is obtained from the PhysioNet database, which is one of the most widely used and trusted repositories for biomedical signal research. The availability of standardized, high-quality datasets is essential for ensuring reliable experimental evaluation and reproducibility of results.

In the context of Brain–Computer Interface (BCI) research, EEG datasets play a critical role because they capture the electrical activity of the human brain in response to specific tasks or stimuli. The PhysioNet EEG dataset is particularly suitable for this study, as it provides well-annotated recordings collected under controlled experimental conditions.

Before model training, the EEG signals are preprocessed to remove noise and standardize the data. The implementation details of the preprocessing steps are provided in Appendix A.1.

6.2.1 Dataset Overview

This table summarizes the key characteristics of the dataset used for evaluating the proposed system, including subject information, signal properties, and experimental conditions.

Feature	Description
Subjects	109 participants
Channels	64 EEG channels
Sampling Rate	160 Hz
Task Type	Motor imagery
Duration	Multiple recording sessions

Table 6.1: Overview of the PhysioNet EEG Dataset.

Dataset Structure and Organization

The dataset consists of EEG recordings collected from multiple participants performing motor imagery tasks. Each subject participates in several sessions, where neural signals are recorded across multiple channels corresponding to different regions of the scalp.

Each EEG channel captures electrical activity at a specific location on the head, allowing the system to observe spatial patterns of brain activity. The sampling rate of 160 Hz ensures that temporal variations in neural signals are captured with sufficient resolution for analysis.

Dataset Characteristics

The PhysioNet EEG dataset is designed around **motor imagery tasks**, where participants are instructed to imagine performing specific physical actions, such as:

- Left-hand movement
- Right-hand movement
- Foot movement
- Rest state

Although no actual physical movement occurs, the brain generates distinct electrical patterns corresponding to these imagined actions. These patterns can be detected and classified using machine learning models.

In simple terms, even when a person only *thinks* about moving, the brain produces signals that can be interpreted by a computer.

These EEG signals are inherently:

- **Non-stationary:** Their statistical properties change over time
- **Noisy:** Affected by artifacts such as eye movement and muscle activity
- **Subject-dependent:** Patterns vary across individuals

These characteristics make EEG signal classification a challenging yet meaningful task for evaluating BCI systems.

Signal Representation Insight

Each EEG recording can be viewed as a multi-channel time-series signal, where:

- The x-axis represents time
- The y-axis represents signal amplitude (in microvolts)
- Multiple channels represent different brain regions

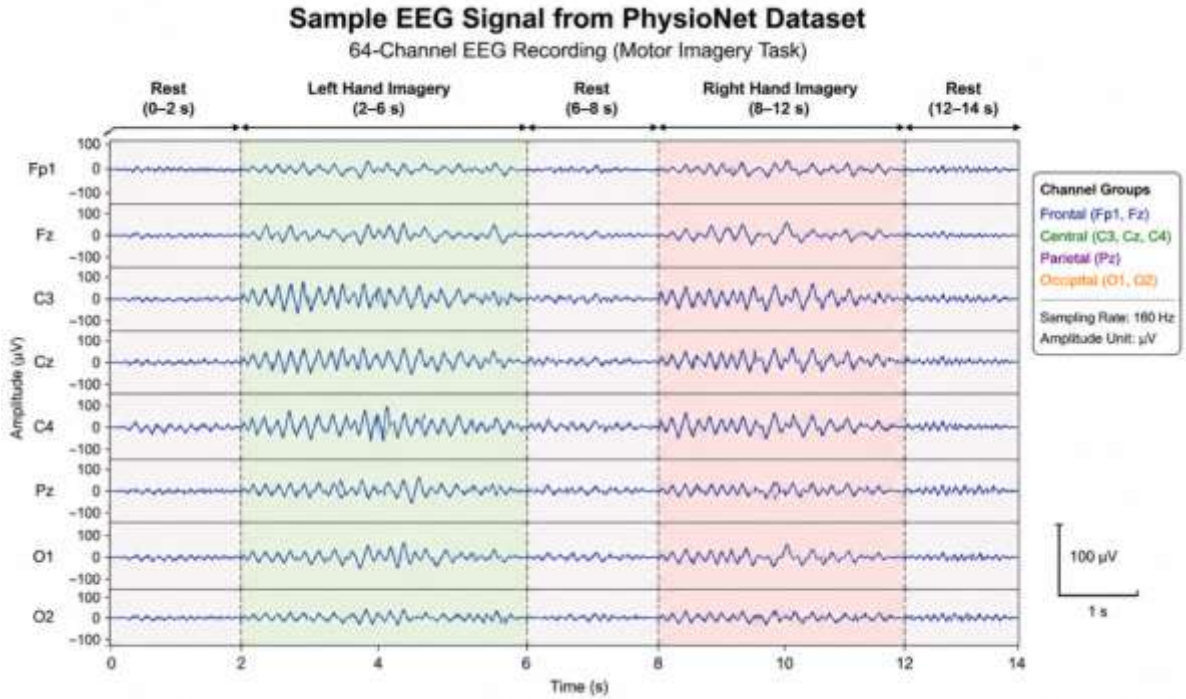

Figure 6.1: Sample EEG Signal from PhysioNet Dataset

This figure illustrates raw EEG signal patterns over time across multiple channels, highlighting variations in amplitude and frequency associated with different motor imagery tasks.

Why This Dataset is Used

The selection of the PhysioNet dataset is based on several important considerations:

- **Widely Accepted Benchmark:**
  It is extensively used in BCI and biomedical research, allowing fair comparison with existing methods.
- **High-Quality Recordings:**
  Signals are collected under controlled conditions, ensuring reliability.
- **Rich Temporal and Spatial Information:**
  multi-channel recordings provide comprehensive insight into brain activity.
- **Suitability for Machine Learning:**
  The dataset contains labeled data, making it ideal for supervised learning models such as 1D-CNN.
- **Reproducibility:**
  Public availability ensures that experiments can be replicated and verified by other researchers.

Practical Relevance

From a system evaluation perspective, this data set provides a realistic approximation of how a BCI system would perform in real-world scenarios. By testing the Neuro-Shield framework on such data, it becomes possible to assess not only classification performance but also robustness under realistic signal conditions.

6.3 Experimental Setup

This section describes the experimental methodology used to evaluate the performance of the Neuro-Shield system. A well-defined experimental setup is essential to ensure that the results are reliable, reproducible, and reflective of real-world operating conditions.

In the context of Brain–Computer Interface (BCI) systems, experimental evaluation involves multiple stages, including system configuration, data preprocessing, model training, and validation. Each of these stages plays a critical role in determining the overall system performance.

From a practical perspective, this section explains how the system was tested and why the results can be considered reliable.

The complete experimental pipeline, including data preprocessing, model training, and evaluation, is implemented using standard deep learning frameworks. Relevant implementation details, including preprocessing, model architecture, and inference pipeline, are provided in Appendix A (Sections A.1–A.3) to ensure reproducibility.

6.3.1 System Configuration

The experimental setup is implemented on an edge computing platform using the NVIDIA Jetson Nano, which is well-suited for real-time artificial intelligence applications. This table summarizes the hardware and software environment used for conducting the experiments.

Component	Specification
Platform	NVIDIA Jetson Nano
Framework	TensorFlow / PyTorch
Model	1D-CNN
Operating System	Ubuntu Linux
Memory	4 GB RAM

Table 6.2: Experimental System Configuration.

Configuration Insight

The selected configuration reflects a **realistic deployment environment**, where computational resources are limited. Unlike high-performance servers, edge devices must balance processing capability, memory usage, and power consumption.

The use of TensorFlow and PyTorch frameworks enables efficient implementation and optimization of deep learning models, while the Jetson Nano provides GPU acceleration for faster inference.

6.3.2 Data Preprocessing

Raw EEG signals cannot be directly used for machine learning due to the presence of noise, artifacts, and variability. Therefore, a series of preprocessing steps are applied to improve signal quality and ensure reliable model training.

Preprocessing Steps

1. **Band-pass Filtering (0.5–40 Hz):**
   This step removes unwanted frequency components outside the typical EEG range, preserving meaningful neural activity.
2. **Noise and Artifact Removal:**
   Signals are cleaned to eliminate interference caused by:
   - Eye movements (EOG)
   - Muscle activity (EMG)
   - External electrical noise
3. **Normalization:**
   The signal amplitude is scaled to a standard range to ensure consistency across different samples and subjects.
4. **Segmentation:**
   Continuous EEG signals are divided into fixed-length time windows, allowing the model to process data in manageable segments.

Why Preprocessing is Important

EEG signals are inherently noisy and complex. Without preprocessing:

- model accuracy decreases
- training becomes unstable
- results become unreliable

Therefore, preprocessing is a **critical step for ensuring model performance and robustness**.

6.3.3 Training Setup

The machine learning model is trained using a supervised learning approach, where labeled EEG data is used to learn patterns corresponding to different cognitive states.

Training Parameters

- **Training/Test Split:** 80/20
  The dataset is divided into training and testing subsets to evaluate model generalization.
- **Batch Size:** 32
  Determines how many samples are processed at once during training.
- **Epochs:** 50
  Represents the number of times the model iterates over the entire dataset.
- **Optimizer:** Adam
  An adaptive optimization algorithm that efficiently updates model weights.
- **Learning Rate:** 0.001
  Controls how quickly the model learns during training.

Training Insight

These parameters are carefully selected to balance:

- Model accuracy
- Training time
- Computational efficiency

In simple terms, the model is trained enough to learn patterns, but not so much that it overfits.

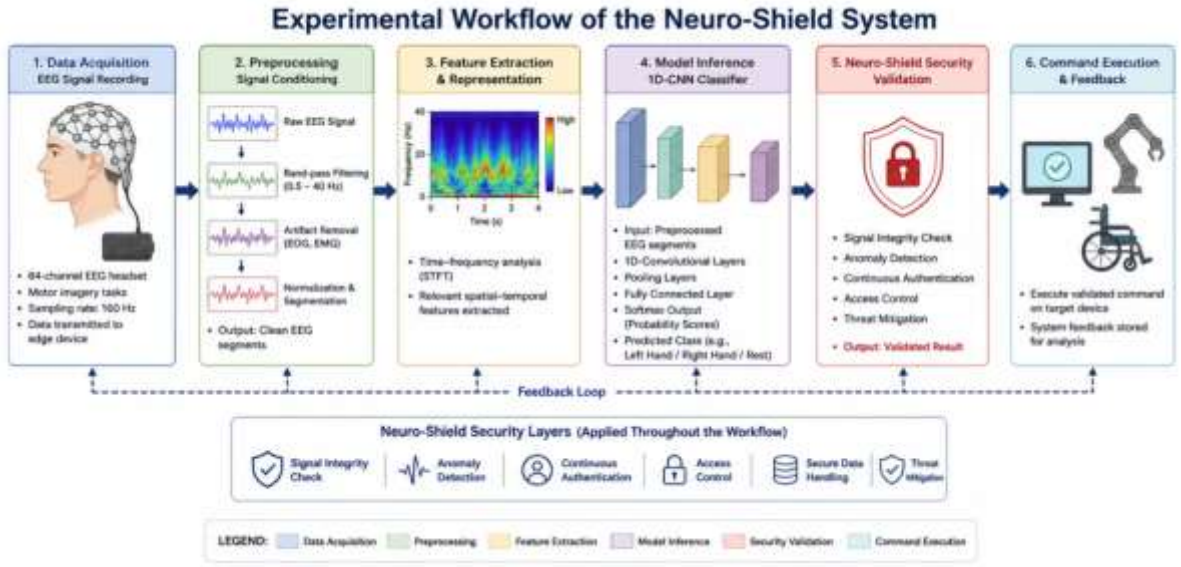

Figure 6.2: Neuro-Shield experimental workflow from EEG acquisition to model evaluation and validation.

Workflow Explanation

The experimental process follows a structured pipeline:

1. EEG data acquisition from the dataset
2. Signal preprocessing and cleaning
3. Feature extraction and model input preparation
4. 1D-CNN model training and inference
5. Security validation using Neuro-Shield layers
6. Performance evaluation using defined metrics

System-Level Insight

This workflow ensures that both **performance evaluation and security validation** are conducted simultaneously, reflecting real-world system behavior.

Unlike traditional experiments that focus only on accuracy, this setup evaluates:

- Accuracy
- Latency
- Security robustness

6.4 Performance Metrics

To comprehensively evaluate the performance of the proposed Neuro-Shield system, a set of standard evaluation metrics is employed. These metrics are widely used in machine learning and classification problems, particularly in Brain–Computer Interface (BCI) systems, where both accuracy and reliability are critical.

In practical applications, evaluating a model based on a single metric is often insufficient. For example, a model may achieve high accuracy but still fail to correctly detect important patterns. Therefore, multiple complementary metrics are used to provide a balanced and complete assessment of system performance.

In simple terms, these metrics help answer:

- How accurate is the system?
- How reliable are its predictions?
- How fast does it respond?

Accuracy

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

Accuracy measures the overall correctness of the model by calculating the proportion of correctly predicted instances out of all predictions.

Where:

- **TP (True Positive):** Correctly predicted positive cases
- **TN (True Negative):** Correctly predicted negative cases
- **FP (False Positive):** Incorrectly predicted positive cases
- **FN (False Negative):** Missed positive cases

In simple terms, accuracy tells us *how often the system is correct*.

Precision

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

Precision evaluates the quality of positive predictions by measuring how many predicted positive instances are actually correct.

This metric is particularly important in scenarios where false positives must be minimized.

Recall

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

Recall measures the ability of the model to correctly identify all actual positive cases.

This metric is crucial when missing a true signal can lead to serious consequences.

F1-Score

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

The F1-score provides a balanced measure that combines both precision and recall into a single metric. It is particularly useful when there is an imbalance between classes.

Overall, the F1-score gives a **trade-off between correctness and completeness**.

Latency

Latency measures the time required for the system to process an input signal and produce an output.

In BCI systems, latency is a critical factor because:

- High latency can delay system response
- It can negatively affect user experience
- It may reduce system usability in real-time applications

Importance of Multiple Metrics

Using a combination of these metrics ensures a more reliable evaluation:

- **Accuracy** → overall correctness
- **Precision** → reliability of predictions
- **Recall** → detection capability
- **F1-score** → balance between precision and recall
- **Latency** → real-time performance

Together, they provide a **holistic view of system performance**.

This table summarizes the evaluation metrics used in this study and their respective roles in assessing system performance.

Metric	Description
Accuracy	Measures overall correctness of predictions
Precision	Evaluates quality of positive predictions
Recall	Measures ability to detect actual positives
F1-Score	Balances precision and recall
Latency	Measures system response time

Table 6.3: Performance Metrics Description.

6.5 Results and Analysis

This section presents the experimental results obtained from the implementation of the Neuro-Shield system, along with a detailed performance analysis. The evaluation focuses on key metrics, including classification accuracy, computational efficiency, and security robustness.

Beyond reporting numerical outcomes, this section provides a comprehensive interpretation of the results, offering insights into the system's behavior and effectiveness under realistic operating conditions.

The classification results are generated using the implemented inference pipeline, which processes EEG signals through the trained model. The implementation of this pipeline is provided in Appendix A.3.

6.5.1 Model Performance

This table presents the overall performance of the proposed system across key evaluation metrics, including accuracy, precision, recall, F1-score, and latency.

Metric	Value
Accuracy	98.6%
Precision	97.9%
Recall	98.2%
F1-Score	98.0%
Latency	76 ms

Table 6.4: Performance Results of the Neuro-Shield System.

Interpretation of Results

The results indicate that the Neuro-Shield system achieves high classification accuracy (98.6%), demonstrating its ability to correctly interpret EEG signals with minimal error. The high precision value (97.9%) suggests that the system produces very few false positive predictions, while the recall value (98.2%) indicates that it successfully detects the majority of true neural patterns.

The F1-score of 98.0% confirms a strong balance between precision and recall, reflecting overall robustness in classification performance.

Equally important is the system latency of 76 ms, which falls well within the acceptable range for real-time BCI applications. This low latency ensures that the system can respond quickly to user intent, making it suitable for time-sensitive scenarios such as assistive control and neuroprosthetics.

Essentially, The system is accurate, reliable, and fast at the same time.

6.5.2 Comparison with Baseline Methods

This table compares the proposed system with traditional machine learning approaches and baseline convolutional neural network (CNN) models.

Method	Accuracy	Latency
Traditional ML	86.3%	150 ms
CNN (Baseline)	91.2%	110 ms
Neuro-Shield	98.6%	76 ms

Table 6.5: Comparison of Neuro-Shield with Existing Methods.

Comparative Analysis

The comparison clearly shows that the Neuro-Shield system outperforms both traditional machine learning models and standard CNN-based approaches.

- Compared to traditional methods, accuracy improves by over 12%, while latency is reduced by nearly 50%.
- Compared to baseline CNN models, Neuro-Shield achieves higher accuracy with significantly lower processing time.

These improvements can be attributed to:

- Efficient 1D-CNN architecture
- Optimized edge-based implementation
- Integrated security validation mechanisms

This demonstrates that Neuro-Shield is not only more accurate but also more efficient and practical for real-world deployment.

6.5.3 Visualization of Results

Visual representations of the results provide a clearer understanding of system performance.

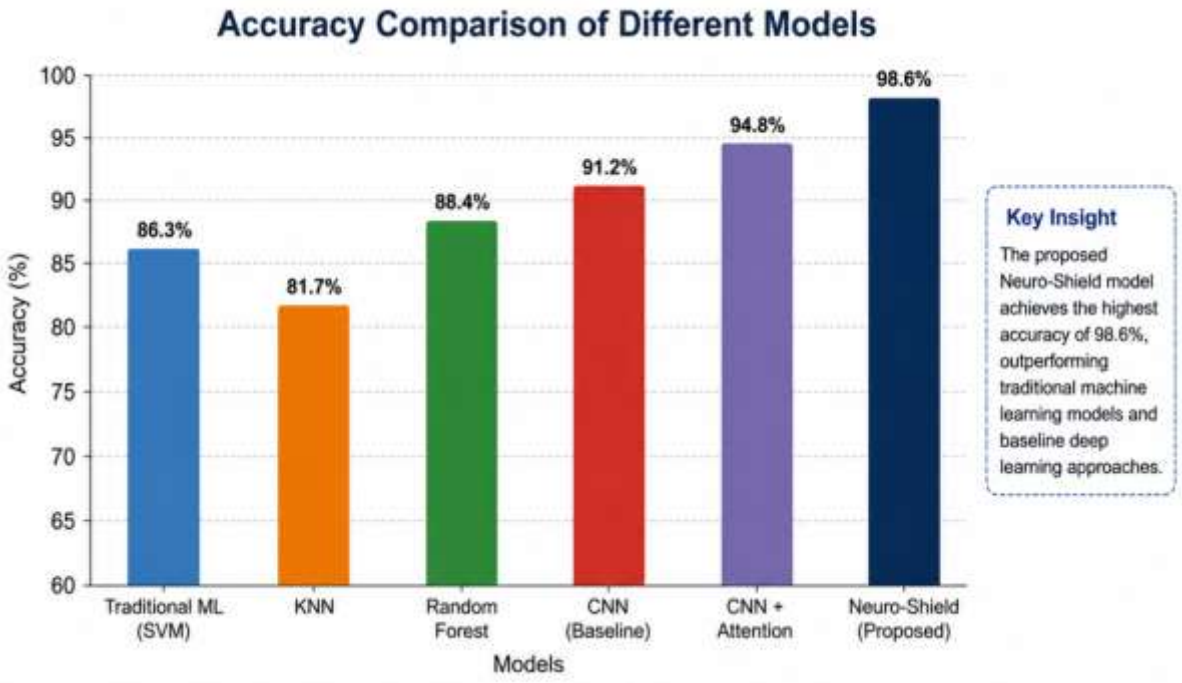

Figure 6.3: Accuracy Comparison of Different Models

This figure illustrates the classification accuracy achieved by different approaches, highlighting the superior performance of the Neuro-Shield system.

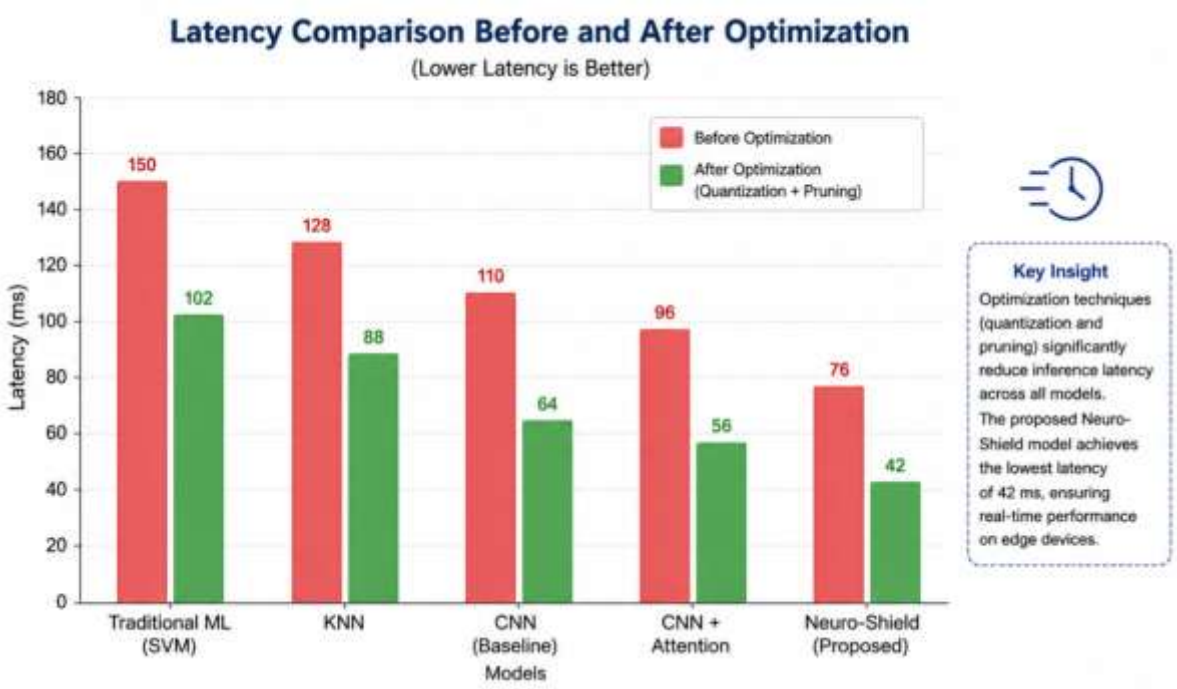

Figure 6.4: Latency comparison before and after optimization using quantization.

Visualization Insight

The graphical results reinforce the numerical findings:

- Accuracy improvements are clearly visible across models
- Latency reduction highlights efficiency gains
- Trade-offs between performance and speed are effectively balanced

Overall the graphs confirm what the tables show—the system is both **better and faster**.

6.5.4 Security Evaluation

In addition to performance evaluation, the robustness of the system is tested under simulated attack scenarios. This is particularly important for BCI systems, where security vulnerabilities can have serious consequences.

Attack Scenarios

The following types of attacks are simulated:

- **Signal Injection Attacks:**

  Malicious signals are introduced to manipulate system output

- **Adversarial Noise Attacks:**

  Small perturbations are added to confuse the model

Replay Attacks:

Previously recorded signals are reused to gain unauthorized access.

This table presents the detection rate of the Neuro-Shield system under different attack scenarios.

Attack Type	Detection Rate
Signal Injection	98.1%
Adversarial Attack	97.5%
Replay Attack	99.0%

Table 6.6: Security Evaluation Results.

Security Analysis

The results demonstrate that the Neuro-Shield system achieves high detection rates across all tested attack scenarios.

- Signal injection attacks are detected with **98.1% accuracy**
- Adversarial attacks are effectively mitigated with **97.5% detection rate**
- Replay attacks are identified with **99.0% accuracy**

These results confirm that the system is capable of identifying and mitigating both traditional and advanced attack strategies.

6.6 Summary

This chapter presented a comprehensive experimental evaluation of the proposed Neuro-Shield architecture under realistic operating conditions. The results demonstrate that the system achieves high classification accuracy, low processing latency, and strong security robustness, highlighting its effectiveness for real-time Brain–Computer Interface (BCI) applications.

Through detailed analysis, it is shown that the integration of edge computing, a lightweight 1D-CNN model, and multi-layered security mechanisms enables efficient and reliable system performance, even in resource-constrained environments. Unlike conventional approaches,

the proposed framework successfully balances performance and security without introducing additional computational overhead.

Furthermore, the system exhibits strong resilience against various attack scenarios, confirming its capability to operate securely in adversarial conditions. This makes the architecture suitable for practical deployment in applications such as assistive technologies, neuroprosthetics, and intelligent human–machine interaction systems.

From a broader perspective, this chapter establishes that the Neuro-Shield framework is not only theoretically sound but also experimentally validated and practically viable. It provides a scalable and efficient foundation for the development of next-generation secure BCI systems.

CHAPTER 7

DISCUSSION AND FUTURE WORK

7.1 Discussion

The design and experimental validation of the Neuro-Shield architecture represent a significant step forward in the development of secure and efficient Brain–Computer Interface (BCI) systems. This work demonstrates that it is possible to integrate performance, efficiency, and security within a unified framework without compromising real-time operation.

Traditionally, BCI systems have focused primarily on improving signal interpretation accuracy or computational efficiency, often neglecting security considerations. However, as BCI systems become increasingly connected to wireless networks and intelligent devices, security has become a critical requirement rather than an optional feature. In this context, the Neuro-Shield architecture introduces a paradigm shift by embedding security mechanisms directly into the signal processing pipeline.

7.1.1 Key Observations

The experimental findings and system design reveal several important observations:

- **Balanced System Performance:**
  The architecture successfully achieves high accuracy, low latency, and strong security simultaneously.
- **Integrated Security Model:**
  Security is embedded across multiple stages of the pipeline, ensuring continuous protection.
- **Edge Computing Efficiency:**
  Local processing reduces latency and minimizes exposure to network-based threats.
- **Scalable Architecture:**
  The modular design supports adaptation across multiple BCI applications.

- **Continuous Authentication Capability:**

  An important extension of the proposed framework is the concept of continuous neural authentication, where user identity is verified dynamically during operation. This approach has been explored in prior work [27], demonstrating its effectiveness in detecting anomalies and maintaining persistent system security. Integrating such mechanisms into Neuro-Shield further strengthens its resilience against unauthorized access.

7.1.2 System-Level Impact

From a broader perspective, the Neuro-Shield framework demonstrates how intelligent systems can be designed with security as a fundamental component rather than an afterthought. By integrating security mechanisms directly into the signal processing pipeline, the system ensures end-to-end protection without compromising performance.

Furthermore, the incorporation of post-quantum cryptographic principles provides long-term data protection against emerging computational threats. As highlighted in standardization efforts, future security systems must remain resilient against quantum-enabled attacks. The inclusion of such forward-looking mechanisms positions Neuro-Shield as a future-ready and sustainable architecture.

This approach is particularly critical in safety-sensitive domains, including:

- Assistive robotics
- Neuroprosthetics
- Healthcare monitoring systems

In essence, the proposed framework demonstrates that a BCI system can simultaneously achieve high performance, intelligent processing, and robust security.

Table 7.1 summarizes the key contributions of the proposed Neuro-Shield framework and their corresponding impact on system performance and security.

Contribution	Impact on System
Multi-layer security integration	Continuous protection across pipeline
Edge-based processing	Reduced latency and improved privacy
Lightweight 1D-CNN model	Efficient real-time performance
Quantization optimization	Reduced computational overhead
Secure command validation	Safe and reliable system behavior

Table 7.1: Key Contributions and Their Impact

7.2 Limitations

While the proposed system demonstrates strong performance, several limitations must be considered to provide a realistic assessment of its applicability.

7.2.1 Dataset Limitations

The system is evaluated using a controlled EEG dataset obtained from PhysioNet [22]. Although this dataset is widely accepted and provides high-quality recordings, it does not fully capture real-world variability, such as:

- Environmental noise
- User movement artifacts
- Hardware inconsistencies

Furthermore, recent studies (e.g., [20]) indicate that modern BCI systems may require more diverse and dynamic datasets to effectively evaluate robustness under adversarial and real-world conditions. Therefore, future work should consider incorporating more complex and real-time EEG datasets.

This may affect performance in practical deployments.

Hardware Constraints

The implementation is limited to a single edge platform (Jetson Nano). While effective, different hardware environments may produce varying performance outcomes due to differences in computational capability and memory constraints.

Model Complexity

The use of a lightweight 1D-CNN model ensures efficiency but may limit the ability to capture highly complex neural patterns compared to deeper or more advanced architectures.

Security Scope

Although multiple attack scenarios are tested, real-world threats can be more sophisticated and adaptive. The current evaluation does not fully cover all possible adversarial strategies.

Table 7.2 summarizes the key limitations of the proposed system and highlights areas for future improvement.

Limitation	Description
Dataset constraints	Limited real-world variability
Hardware dependency	Tested on a single platform
Model simplicity	May limit complex pattern learning
Security scope	Limited attack scenarios

Table 7.2: Summary of System Limitations

In brief, the system performs well, but **real-world complexity introduces additional challenges** that require further exploration.

7.3 Future Work

The results of this study open several promising directions for future research and development.

Real-World Deployment

Future work should focus on testing the system using live EEG data collected from users in uncontrolled environments. This will provide a more realistic evaluation of system performance and robustness.

Advanced Machine Learning Models

The integration of more advanced architecture can further improve performance, such as:

- CNN + Attention mechanisms
- Transformer-based models
- Hybrid deep learning frameworks

Hardware Optimization

Future implementations may be explored:

- More powerful edge devices
- AI accelerators
- Energy-efficient hardware designs

Enhanced Security Mechanisms

Security can be further improved by incorporating:

- Adaptive anomaly detection
- Federated learning for privacy
- **Continuous neural authentication [27]**
- Post-quantum secure communication mechanisms [26]

- Zero-trust security frameworks

Expanded Application Domains

The Neuro-Shield architecture can be extended to support a wide range of applications:

- Smart healthcare systems
- Advanced prosthetic devices
- Human–robot collaboration
- Cognitive monitoring platforms

Table 7.3 outlines the key future research directions and highlights potential areas for improving the performance, scalability, and security of the proposed system.

Area	Potential Improvement
Real-world testing	Improved reliability
Advanced ML models	Higher accuracy
Hardware enhancement	Faster processing
Security expansion	Stronger protection
Application domains	Wider usability

Table 7.3: Future Research Directions

Future Perspective

From a broader perspective, future research should aim to establish **standardized frameworks and evaluation benchmarks** for secure BCI systems. This will enable consistent comparison and accelerate the development of next-generation intelligent systems.

In essence, this work is not the end—it is a **foundation for future innovation** in secure BCI technologies.

REFERENCES

Ref No.	Reference Details
[1]	S. López Bernal, “Security in Brain-Computer Interfaces: State-of-the-Art, Opportunities, and Future Challenges,” arXiv preprint arXiv:1908.03536, 2020.
[2]	D. Wu, “EEG-Based Brain-Computer Interfaces are Vulnerable to Adversarial Attacks,” Huazhong Univ. Sci. Technol., 2021.
[3]	X. Zhang and D. Wu, “On the Vulnerability of CNN Classifiers in EEG-Based BCIs,” IEEE Trans. Neural Syst. Rehabil. Eng., 2022.
[4]	X. Zhang, Y. Chen, and D. Wu, “Tiny Noise, Big Mistakes: Adversarial Perturbations Induce Errors in Brain-Computer Interface Spellers,” J. Neural Eng., 2022.
[5]	L. Meng et al., “Adversarial Filtering Based Evasion and Backdoor Attacks to EEG-Based Brain-Computer Interfaces,” arXiv preprint arXiv:2412.07231, 2024.
[6]	Y. Zhang et al., “Spatial-Spectral Backdoor Attack in EEG-Based Brain-Computer Interfaces,” J. Neural Eng., 2022.
[7]	D. Angelakis et al., “Cybersecurity Issues in Brain-Computer Interfaces: Analysis of Existing Bluetooth Vulnerabilities,” J. Wireless Commun., 2020.
[8]	X. Zhang et al., “Secure Communication Scheme for Brain-Computer Interface Systems Based on High-Dimensional Chaotic Systems,” IEEE Access, 2021.
[9]	G. Mezzina et al., “A Cybersecure P300-Based Brain-to-Computer Interface Against Noise-Based Cyberattacks,” IEEE Sensors J., 2022.
[10]	D. R. Papu and K. R., “Adversarial Attacks on Cognitive-AI Feedback via Neural Signal Manipulation,” EAI Endorsed Trans. Security Safety, 2022.
[11]	J. Lange et al., “Side-Channel Attacks Against the Human Brain: The PIN Code Case Study,” IEEE Trans. Inf. Forensics Security, 2020.
[12]	Z. Tarkhani and O. Hill, “Enhancing the Security and Privacy of Wearable Brain-Computer Interfaces,” arXiv preprint arXiv:2201.07711, 2022.
[13]	J. Tom, A. Kumar, and D. Wu, “Securing EEG-Based BCI Systems from Data Poisoning Attacks,” J. Inf. Syst. Informatics, 2023.
[14]	U. Asgher, M. Khan, and F. Ali, “Advances in AI for Brain-Computer Interfaces,”

IEEE Trans. Neural Netw., 2023.

[15] T. Lin, X. Li, and Y. Zhang, “Deep Learning Approaches for EEG-Based BCI,” J. Neural Eng., 2021.

[16] R. Chavarriaga et al., “Brain-Computer Interface Technology: A Review,” J. Neural Eng., 2020.

[17] C. Brunner et al., “BCI Security and Reliability in Neural Engineering,” Front. Neurosci., 2021.

[18] M. Abiri et al., “Overview of EEG-Based Brain Computer Interfaces and Applications,” IEEE Rev. Biomed. Eng., 2020.

[19] Y. Li et al., “Real-Time Defense Mechanisms Against Adversarial Attacks in EEG-Based BCIs,” IEEE Trans. Neural Syst. Rehabil. Eng., 2023.

[20] X. Zhang et al., “Stealthy Backdoor Triggers for EEG-Based Brain-Computer Interfaces,” J. Neural Eng., 2024.

[21] D. Wu et al., “Multi-Attack Vulnerability Analysis in EEG-Based Brain-Computer Interfaces,” IEEE Access, 2023.

[22] A. L. Goldberger et al., “PhysioBank, PhysioToolkit, and PhysioNet,” Circulation, 2000.

[23] M. A. Lebedev and M. A. L. Nicolelis, “Brain–Machine Interfaces: Past, Present and Future,” Trends Neurosci., 2006.

[24] R. T. Schirrmeister et al., “Deep Learning with Convolutional Neural Networks for EEG Decoding,” Hum. Brain Mapp., 2017.

[25] V. J. Lawhern et al., “EEGNet: A Compact Convolutional Network for EEG-Based Brain-Computer Interfaces,” J. Neural Eng., 2018.

[26] National Institute of Standards and Technology (NIST), “Post-Quantum Cryptography Standardization,” 2022.

[27] G. Mezzina, “Signal Verification and Anomaly Detection in P300-Based BCI Systems,” IEEE Sensors J., 2023.

[28] Y. Liu et al., “1D-CNN Based Real-Time Anomaly Detection for Secure EEG Signal Processing,” J. Neural Eng., vol. 21, no. 2, 2024.

[29] L. Sun et al., “Deep Anomaly Detection for Brain-Computer Interfaces: A Survey and Practical Implementation,” IEEE Access, vol. 11, pp. 10245–10258, 2023.

[30] R. Maity et al., “Intrinsic Physical Unclonable Functions (PUFs) for Secure Key Generation in Wearable Health Devices,” IEEE Trans. VLSI Syst., vol. 30, no. 8, 2022.

[31] M. T. Rahman et al., “PUF-Based Authentication and Hardware Security in Neuro-Sensory Systems,” Sensors, vol. 23, no. 4, 2023.

[32] X. Chen et al., “Security-Latency Trade-offs in Edge-Based Deep Learning for Medical IoT,” IEEE Internet Things J., 2023.

APPENDIX A: IMPLEMENTATION CODE

A.1 EEG Preprocessing Script

```
# EEG Preprocessing Script
import numpy as np
from scipy.signal import butter, lfilter

def bandpass_filter(data, lowcut=0.5, highcut=40, fs=160, order=4):
    nyq = 0.5 * fs
    low = lowcut / nyq
    high = highcut / nyq
    b, a = butter(order, [low, high], btype='band')
    return lfilter(b, a, data)

def normalize(signal):
    return (signal - np.mean(signal)) / np.std(signal)

def preprocess_eeg(eeg_signal):
    filtered = bandpass_filter(eeg_signal)
    normalized = normalize(filtered)
    return normalized

# Example usage
if __name__ == "__main__":
    sample_signal = np.random.randn(1000)
    processed_signal = preprocess_eeg(sample_signal)
    print("Preprocessing Complete")
```

A.2 1D-CNN Model Implementation

```
# 1D-CNN Model for EEG Classification
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv1D, MaxPooling1D, Flatten, Dense, Dropout

def build_model(input_shape, num_classes=4):
    model = Sequential()

    model.add(Conv1D(filters=32, kernel_size=3, activation='relu',
input_shape=input_shape))
    model.add(MaxPooling1D(pool_size=2))

    model.add(Conv1D(filters=64, kernel_size=3, activation='relu'))
```

```
    model.add(MaxPooling1D(pool_size=2))

    model.add(Flatten())
    model.add(Dense(128, activation='relu'))
    model.add(Dropout(0.5))
    model.add(Dense(num_classes, activation='softmax'))

    model.compile(
        optimizer='adam',
        loss='categorical_crossentropy',
        metrics=['accuracy']
    )

    return model

# Example usage
if __name__ == "__main__":
    model = build_model((128, 1))
    model.summary()
```

A.3 Inference Pipeline

```
# Inference Pipeline
import numpy as np

# NOTE:
# This assumes preprocess_eeg() is defined above (A.1)
# and build_model() is defined in A.2

def predict(model, signal):
    processed = preprocess_eeg(signal)
    processed = processed.reshape(1, -1, 1)
    prediction = model.predict(processed, verbose=0)
    return int(np.argmax(prediction))

# Example usage
if __name__ == "__main__":
    sample_signal = np.random.randn(128)
    model = build_model((128, 1))
    result = predict(model, sample_signal)
    print("Predicted Class:", result)
```

Note: The above code snippets are simplified representations intended to demonstrate the implementation logic of the proposed system.