Securing Brain-to-Brain Communication Channels Using Adversarial Training on SSVEP EEG

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier xxxx

Securing Brain-to-Brain Communication Channels Using Adversarial Training on SSVEP EEG

HOSSEIN AHMADI[1], ALI KUHESTANI[2], (Member, IEEE), MOHAMMADREZA KESHAVARZI[3], and LUCA MESIN[1],

[1]Mathematical Biology and Physiology, Department of Electronics and Telecommunications, Politecnico di Torino, 10129, Turin, Italy (e-mail: hossein.ahmadi@polito.it; luca.mesin@polito.it)

[2]Communications and Electronics Department, Faculty of Electrical and Computer Engineering, Qom University of Technology, Qom 3718146645, Iran (e-mail: kuhestani@qut.ac.ir)

[3]Iran Telecommunication Research Center, ITRC, Tehran, Iran; mrkeshavarzi@itrc.ac.ir

Corresponding author: Hossein Ahmadi (e-mail: hossein.ahmadi@polito.it).

ABSTRACT

In this study, we investigate the effects of Adversarial Neural Network Training (ANNT) on the robustness and effectiveness of Brain-to-Brain Communication (B2B-C) systems using Steady-State Visually Evoked Potentials (SSVEP) EEG data. We utilized a combined Convolutional Neural Network-Temporal Convolutional Network (CNN-TCN) architecture to classify the data and assessed the system's resistance to various adversarial strategies, including Fast Gradient Sign Method (FGSM), Projected Gradient Descent (PGD), Basic Iterative Method (BIM), Carlini & Wagner (C&W), and Momentum Iterative Method (MIM). By analyzing publicly accessible datasets, specifically Lee2019_SSVEP and Nakanishi2015, we observed significant enhancements in both accuracy and AUC metrics when ANNT was applied. Specifically, the Lee2019_SSVEP dataset exhibited a 24% increase in accuracy and a 0.23-point improvement in AUC, while the Nakanishi2015 dataset demonstrated improvements of 9% and 0.07 points, respectively. Our results indicate that PGD posed the greatest challenge to the model, significantly reducing accuracy and AUC across various scenarios, whereas FGSM was the least impactful. These findings highlight ANNT's potential in fortifying the security and stability of B2B-C systems against diverse adversarial conditions.

INDEX TERMS Adversarial Neural Network Training, Brain-to-Brain Communication, Electroencephalogram, Adversarial Attacks, Neuro-engineering, Security Enhancement, Robustness, Machine Learning.

I. INTRODUCTION

BRAIN-to-Brain Communication (B2B-C) is an emerging field at the intersection of neuroscience and communication technology. It proposes a revolutionary shift in how information can be transferred between individuals. This field, still in its nascent stages, aims to enable direct B2B interaction, bypassing traditional forms of communication like speech and text. Such advancements can transform numerous domains, including healthcare, education, and everyday social interactions [1]. The complexity of B2B-C arises from the need to interpret and transmit neural signals between brains accurately [2]. The human brain is an incredibly intricate organ consisting of approximately 86 billion neurons, each forming thousands of synaptic connections. Translating the electrochemical signals from one brain into meaningful information that can be received and understood by another is a task fraught with challenges, requiring sophisticated technology and a deep understanding of neuroscience and signal processing [3].

Electroencephalography (EEG) signals are pivotal in this era of brain research due to their non-invasive nature and ability to provide real-time monitoring of brain activity. EEG captures the brain's electrical activity using sensors placed on the scalp, offering a window into the neural processes underlying cognition, perception, and motor functions. The complexity of EEG data lies in its low amplitude, high dimensionality, and noise, which necessitates advanced methods for effective data analysis and interpretation [4]. High-density EEG, in particular, offers a rich dataset analogous to a series of images evolving over time [5]. These data can be processed using methods from medical image analysis and Deep Learning (DL) to extract meaningful patterns and insights. Techniques commonly used in image processing have proven effective in handling EEG data's high dimensionality and complexity, enhancing the robustness and accuracy of B2B-C systems [6].

Integrating these advanced methodologies into EEG anal-

ysis opens new possibilities for improving the performance and security of B2B-C systems. For example, by employing DL techniques, researchers have significantly advanced the accuracy of EEG signal classification, contributing to more reliable and secure communication channels [7]. Furthermore, medical image analysis techniques have been adapted to process high-density EEG data, allowing for the detailed exploration of brain activity and its applications in various domains [8]. Given the high complexity of EEG data, dimensionality reduction techniques are often employed to distill the data into a more manageable form without losing critical information [9].

Within the realm of EEG, Steady-State Visually Evoked Potentials (SSVEPs) stand out for their robustness and reliability in various applications, including B2B-C [10]. SSVEPs are elicited by visual stimuli flickering at specific frequencies, producing brain responses that are relatively easy to detect and analyze [11]. This makes SSVEPs particularly useful for B2B-C, as they can provide consistent and high-quality signals that facilitate accurate interpretation and communication between brains [12].

Ensuring the security and robustness of B2B-C systems, particularly when transmitting EEG signals, is paramount due to the sensitive nature of neural data [13]. While offering a non-invasive and real-time glimpse into brain activity, EEG signals are highly susceptible to noise, interference, and malicious attacks [14]. The integrity of the transmitted data is critical; any compromise can lead to significant errors in interpretation and potentially harmful consequences. One of the main challenges lies in protecting these signals from adversarial attacks that can subtly alter the data, leading to miscommunication or data breaches [15]. Robustness must be built into the system to ensure that it can withstand such perturbations and continue to function accurately. This requires advanced Machine Learning (ML) techniques, such as Adversarial Neural Network Training (ANNT), which can develop models to recognize and resist these adversarial patterns. Additionally, secure transmission protocols must be developed to safeguard the data as it travels through wireless channels, preventing unauthorized access and ensuring the privacy and integrity of the communication [16]. As B2B-C technology evolves, addressing these security and robustness challenges will be essential to realizing its full potential and maintaining trust in its applications across healthcare, education, and other fields.

Enormous studies have been conducted on EEG analysis, particularly within Brain-Computer Interface (BCI) technology, significantly advancing our understanding and application. One extraordinary work is by Donchin and colleagues, who developed the P300-based BCI, a system that allows individuals to communicate without muscle activity by detecting the P300 wave, a component of EEG that occurs in response to decision-making processes [17]. Another notable study is by Wolpaw et al., who created the sensorimotor rhythm-based BCI, enabling users to control external devices by modulation of sensorimotor rhythms in their EEG signals [18]. Furthermore, Buzsáki's research on using EEG for understanding the brain's oscillatory activity has provided profound insights into how different brain waves are associated with various cognitive functions and states of consciousness [19]. In [20], a novel ensemble model was introduced, significantly enhancing Motor Imagery (MI) EEG classification by integrating multiple ML classifiers through a weighted stacking approach. Additionally, their study on the Weighted and Stacked Adaptive Integrated Ensemble Classifier (WS-AIEC) has shown superior performance in EEG signal classification, achieving exceptional accuracy and reliability across multiple datasets [21]. These outstanding works, among many others, highlight the versatility and potential of EEG analysis in advancing both theoretical neuroscience and practical applications in neurotechnology.

Despite the extensive research on various aspects of EEG signal analysis, studies focusing on using EEG in direct B2B-C remain relatively scarce. Only a limited number of investigations have addressed the critical security and robustness issues in this context. The challenge is further compounded when considering integrating interdisciplinary fields such as wireless communication, neuroscience, and artificial intelligence. This interdisciplinary approach is crucial for developing a comprehensive and secure B2B-C system, yet it is an area that has not been extensively explored. For instance, authors in [1] and [22] have done valuable foundational work in B2B-C but didn't discuss its critical aspect: security. On the other hand, studies like [23] and [24] have emphasized the need for security but have not ventured into applying ANNT, particularly in conjunction with EEG, for fortifying B2B-C systems. This gap underscores the importance of advancing research in this specific intersection to realize the full potential of B2B-C technology.

In our previous work [25], we aimed to enhance the robustness and security of B2B-C systems against adversarial attacks using Event-Related Potentials (ERP) EEG data through ANNT. By focusing on ERP and employing eight diverse datasets, our objectives were to rigorously evaluate the model's defense mechanisms against adversarial manipulations and optimize trial durations and sampling rates for maximum security. The results signified a notable advancement in system defense, evidenced by an average increase in adversarial accuracy by 17% and an improvement in the Area Under the Curve (AUC) by 0.12 points, demonstrating the effectiveness of our approach in strengthening B2B-C systems against sophisticated cyber threats.

This study extends our previous work [25] by focusing on SSVEP EEG data. This extension will allow us to explore the robustness and security of B2B-C systems in a different and highly reliable EEG paradigm, further enhancing such systems' practical applicability and resilience against adversarial attacks. Our contributions to this field include:

- **First-time comprehensive analysis of diverse adversarial attacks:** We evaluate the robustness of B2B-C systems against various attacks and their combinations.

To the best of our knowledge, this extensive evaluation is the first of its kind in the context of B2B-C systems.
- **Application of ANNT:** We leverage ANNT to improve the robustness and security of B2B-C systems based on SSVEP EEG data, systematically evaluating the model's ability to resist various adversarial patterns.
- **Optimization of critical parameters:** We perform a detailed analysis of key factors such as sampling rate and number of classes to identify optimal conditions that enhance the performance and security of SSVEP-based B2B-C systems.
- **Contribution to secure neural communication technologies:** Our findings provide valuable insights into developing more secure and reliable neural communication technologies, paving the way for practical and trustworthy B2B-C applications in various domains, including healthcare, education, and social interactions.

The organization of this paper is structured to provide a comprehensive overview of our research methodology, findings, and implications. Following this Introduction, the Methodology section details the data description, signal processing techniques, and algorithms used in our study, highlighting the specific methods employed to preprocess and analyze SSVEP EEG data. Additionally, we introduce the different attacks used and the rationale behind choosing each one. In the Results section, we present our findings, including the impact of various parameters on the performance and security of the B2B-C system. The Discussion section discusses and interprets our results, offering insights into our research's practical implications and potential applications. Finally, the Conclusion summarizes our study, emphasizing the significance of our contributions and outlining potential future research directions.

II. METHODOLOGY

A. DATASETS

This study utilized two SSVEP EEG datasets: Nakanishi2015 [26] and Lee2019_SSVEP [27]. Both datasets offer comprehensive EEG recordings but differ in the number of subjects, channels, classes, trials/class, trial duration, and sampling rates, providing a robust basis for evaluating our methods. The details of these datasets are summarized in Table 1.

The Nakanishi2015 dataset includes 12 classes corresponding to different flicker frequencies: 9.25, 11.25, 13.25, 9.75, 11.75, 13.75, 10.25, 12.25, 14.25, 10.75, 12.75, and 14.75 Hz. The Lee2019_SSVEP dataset comprises 4 classes corresponding to flicker frequencies of 5.45, 6.67, 8.57, and 12 Hz.

B. ADVERSARIAL ATTACKS

In this study, we evaluate the robustness and security of our model against various adversarial attacks. We have selected five attacks, each representing a unique approach to perturbing EEG data. By analyzing the effect of each attack individually and in combination, we aim to obtain a comprehensive evaluation of our proposed model's resilience to adversarial manipulations. Below, we detail each attack and our rationale for their inclusion in this study.

- **Fast Gradient Sign Method (FGSM):** FGSM is a straightforward and efficient attack that perturbs the input data by adding noise proportional to the gradient of the loss function with respect to the input [28]. The perturbation is calculated as:

$$\mathbf{x}' = \mathbf{x} + \epsilon \cdot \text{sign}(\nabla_{\mathbf{x}} J(\theta, \mathbf{x}, y)) \tag{1}$$

  where ϵ is a small constant, $\mathbf{x}$ is the input data, J is the loss function, θ represents the model's set of parameters, and y is the true label. FGSM is widely used due to its simplicity and effectiveness in generating adversarial examples. It provides a baseline for evaluating model robustness against adversarial attacks.
- **Projected Gradient Descent (PGD):** PGD is an iterative version of FGSM and is considered one of the strongest first-order adversaries [29]. It applies FGSM iteratively, with each step followed by a projection onto the ϵ-ball around the original input. The iterative nature allows for more refined and potent perturbations. The update rule is:

$$\mathbf{x}'^{(t+1)} = \text{Proj}_{\mathbf{x}+\epsilon}\left(\mathbf{x}'^{(t)} + \alpha \cdot \text{sign}(\nabla_{\mathbf{x}} J(\theta, \mathbf{x}', y))\right) \tag{2}$$

  where α is the step size. PGD's iterative refinement makes it a critical attack for stress-testing model defenses.
- **Basic Iterative Method (BIM):** BIM, also known as iterative FGSM, further enhances FGSM by applying it multiple times with small perturbations at each step [30]. This iterative approach helps in crafting more precise adversarial examples. The iterative process can be expressed as:

$$\mathbf{x}'^{(t+1)} = \mathbf{x}'^{(t)} + \alpha \cdot \text{sign}(\nabla_{\mathbf{x}} J(\theta, \mathbf{x}', y)) \tag{3}$$

  with a clipping step to ensure the perturbations remain within the ϵ-ball. BIM allows for a more granular evaluation of the model's robustness compared to single-step methods.
- **Carlini & Wagner (C&W)**: The C&W attack is an optimization-based method that seeks to find the smallest perturbation that causes a misclassification [31]. It is known for generating very subtle and often undetectable perturbations. The objective is to maximize the likelihood of misclassification while keeping the perturbation minimal. The optimization problem is formulated as:

$$\min_{\delta} \|\delta\|_p + c \cdot f(\mathbf{x} + \delta) \tag{4}$$

  where:
  - -- δ is the perturbation added to the input data,
  - -- $\|\delta\|_p$ denotes the L_p norm (commonly the L_2 norm), which measures the size of the perturbation,
  - -- c is a constant that controls the trade-off between the perturbation size and the likelihood of misclassification,

TABLE 1: Datasets

Dataset	Subjects	Channels	Classes	Trials / Class	Trial Duration	Sampling rate	Sessions
Nakanishi2015 [26]	9	8	12	15	4.15s	256Hz	1
Lee2019_SSVEP [27]	54	62	4	50	4s	1000Hz	2

-- $f(\mathbf{x}+\delta)$ is a function that measures how much the perturbed input $\mathbf{x}+\delta$ deviates from the true label, effectively increasing as the model becomes more confident in a wrong prediction.

The goal of the C&W attack is to find a perturbation δ that both maximizes $f(\mathbf{x}+\delta)$ (making the input more likely to be misclassified) and minimizes $\|\delta\|_p$ (keeping the perturbation small and imperceptible).

- **Momentum Iterative Method (MIM):** MIM integrates momentum into the iterative attack process, enhancing the attack's effectiveness by stabilizing the update direction [32]. The method accumulates a velocity vector that helps in escaping poor local maxima:

$$\mathbf{g}^{(t+1)} = \mu \cdot \mathbf{g}^{(t)} + \frac{\nabla_{\mathbf{x}} J(\theta, \mathbf{x}', y)}{\|\nabla_{\mathbf{x}} J(\theta, \mathbf{x}', y)\|_1} \tag{5}$$

$$\mathbf{x}'^{(t+1)} = \mathbf{x}'^{(t)} + \alpha \cdot \text{sign}(\mathbf{g}^{(t+1)}) \tag{6}$$

where μ is the decay factor for the momentum term. MIM's use of momentum makes it a powerful and effective method for generating adversarial examples.

The rationale behind choosing these attacks is based on their diverse methodologies and effectiveness in generating adversarial examples. FGSM and BIM provide a basic understanding of the model's vulnerability to gradient-based attacks ([28] and [30]), while PGD, being an iterative and refined version of FGSM, offers insights into the model's robustness under stronger adversaries [29]. The C&W attack, known for its subtle perturbations, challenges the model's ability to detect and defend against sophisticated adversarial inputs [31]. With its momentum-based approach, MIM ensures a comprehensive evaluation of the model's robustness by simulating more realistic adversarial conditions [32]. By analyzing the effects of each attack individually and in various combinations, we can thoroughly evaluate our proposed model's performance against different types of adversarial threats. This comprehensive approach allows us to understand the strengths and weaknesses of our B2B-C system, ensuring its resilience in practical applications.

Table 2 lists the parameters and their specific values for each adversarial attack method in our study.

Before delving into the detailed results, we illustrate the impact of two specific attacks—FGSM and PGD—on a randomly selected epoch, channel, and subject from the Nakanishi2015 dataset. This preliminary analysis, depicted in Figure 1, visually explains how these adversarial methods alter EEG signals and highlight the necessity for robust defense mechanisms.

TABLE 2: Parameters and specific values for each adversarial attack method.

Adversarial Attack	Parameter	Value
FGSM	ϵ (perturbation size)	0.01
PGD	ϵ (perturbation size) α (step size) t (iterations)	0.01 0.001 50
BIM	ϵ (perturbation size) α (step size) t (iterations)	0.01 0.001 50
C&W	δ (perturbation size) c (trade-off constant) κ (confidence) t (iterations) L_p norm	0.01 0.001 10 1000 L_2
MIM	ϵ (perturbation size) α (step size) μ (momentum decay factor) t (iterations)	0.01 0.001 0.5 50

C. SYSTEM MODEL

This study employs a DL model based on a Convolutional Neural Network and Temporal Convolutional Network (CNN-TCN) architecture. The system model, illustrated in Figure 2, integrates several critical components and steps to ensure comprehensive data processing and model evaluation. Below is a concise overview of the key stages:

- **Data Acquisition and Preprocessing:**
  - -- We start by loading the SSVEP EEG data from our chosen datasets, Nakanishi2015 and Lee2019_SSVEP.
  - -- The data undergoes preprocessing, including normalization, filtering, and reshaping, as detailed in the preprocessing section. This ensures the EEG signals are clean and formatted correctly for further analysis.
- **Feature Extraction:**
  - -- The first layer is a Conv2D layer with 32 filters and a 3×3 kernel size, followed by a ReLU activation function.
  - -- A MaxPooling2D layer with a 2×2 pool size is used to reduce the spatial dimensions.
  - -- Another Conv2D layer with 64 filters and a 3×3 kernel size, also with ReLU activation, extracts more complex features.
  - -- The reshaped data is then passed through a TCN layer with 64 filters and a 3×3 kernel size to capture temporal dependencies in the EEG signals.
  - -- Finally, a Dense layer with units equal to the number of classes (12 for Nakanishi2015 and 4 for Lee2019_SSVEP) and a softmax activation function classify the signals.

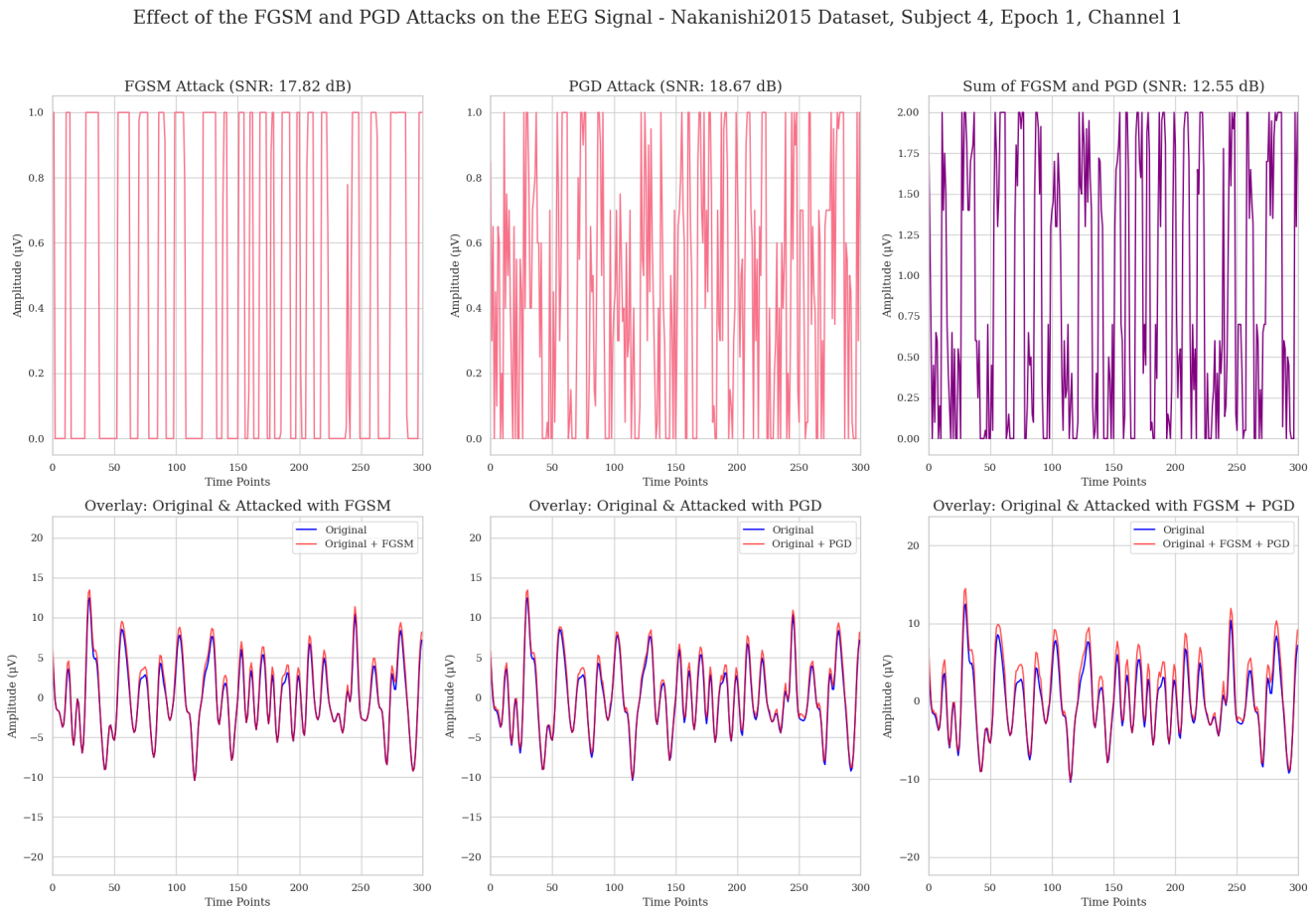


FIGURE 1: Comparison of FGSM and PGD Adversarial Attacks on EEG Signals from the Nakanishi2015 Dataset (Subject 4, Epoch 1, Channel 1). The top row shows the perturbations introduced by FGSM (left), PGD (middle), and the combined FGSM + PGD (right) attacks, along with their respective Signal-to-Noise Ratios (SNRs). The bottom row overlays the original EEG signal with the signals after each attack, demonstrating the distortion caused by the adversarial noise. The y-axis represents amplitude in microvolts (μV), indicating the intensity of the EEG signals.

- **Adversarial Attacks and Training:**
  - -- Adversarial examples are generated using attacks like FGSM, PGD, BIM, C&W, and MIM to simulate potential threats. The perturbed data is created by applying the specified attack or combination of attacks to the clean data. This ensures that the model is exposed to the specific characteristics of each attack during training.
  - -- In scenarios involving more than one attack, the perturbed data results from a combination of the individual attacks, producing unique perturbations. These combinations create a new, distinct adversarial scenario encompassing characteristics from each constituent attack. Consequently, we do not use specific attacks for training and others for testing; rather, the combination acts as a new attack scenario.
  - -- The model is trained using ANNT, which enhances the model's resilience against these adversarial patterns. ANNT involves training the model on both clean and perturbed data, thereby improving its ability to correctly classify inputs and maintain high performance even in the presence of adversarial attacks. This approach aims to make the model robust across various adversarial scenarios, including those involving complex attack combinations.
- **Model Evaluation:**
  - -- The model's performance is evaluated under three scenarios: classification of clean data, classification of adversarially attacked data, and classification of attacked data using ANNT.
  - -- Performance metrics such as accuracy and AUC are used to assess the model's robustness across different scenarios.
- **Comparison and Analysis:**
  - -- The model's performance is compared across the different scenarios to determine the optimal conditions under which ANNT enhances the security and

robustness of SSVEP-based B2B-C systems.

D. PREPROCESSING

The preprocessing of EEG data is crucial for ensuring the accuracy and effectiveness of our model. SSVEP EEG data is a great candidate for B2B-C systems due to its high quality; however, handling and managing this data requires careful consideration, especially in preparation and preprocessing. Therefore, we have implemented rigorous preprocessing steps to ensure the data is clean, standardized, and formatted for advanced analysis. The following steps outline the detailed process for preparing the SSVEP EEG data for our analysis.

- **Loading the Dataset:** The first step involves loading the datasets. This is done by accessing the data files containing the EEG recordings, metadata, and necessary labels. For both the Nakanishi2015 and Lee2019_SSVEP datasets, we use appropriate functions to load the data into our working environment

$$\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{N}, \quad \mathbf{Y} = \{y_i\}_{i=1}^{N} \tag{7}$$

  where $\mathbf{X}$ is the set of EEG data and $\mathbf{Y}$ is the set of corresponding labels.

- **Fetching Data:** Once the datasets are loaded, we fetch the raw EEG data, associated labels, and metadata. This includes extracting the dataset's EEG signal values, trial information, and class labels

$$\mathbf{X}_{raw} = \text{fetch\_data}(\mathbf{X}), \quad \mathbf{Y}_{raw} = \text{fetch\_labels}(\mathbf{Y}) \tag{8}$$

- **Creating MNE Info Structure:** Using the MNE-Python library, we create an info structure that contains metadata about the EEG data, such as the names of the channels, the sampling rate, and the type of data (e.g., EEG channels for recording brain activity and stimulation channels for recording external stimuli). This info structure is essential for further processing and analysis using MNE functions.

- **Reshaping Data for MNE RawArray:** The EEG data is reshaped to fit the format required by the MNE RawArray object. This involves organizing the data into a 2D array where each row represents a channel, and each column represents a time point

$$\mathbf{X}_{reshaped} = \text{reshape}(\mathbf{X}_{raw}) \tag{9}$$

- **Setting Montage:** We set the montage, which defines the spatial arrangement of the EEG electrodes on the scalp. The standard 10-20 system ensures accurate spatial representation of the EEG data.

- **Setting Common Average Reference:** A common average reference (CAR) is applied to the EEG data. This involves re-referencing the data by subtracting the average of all channels from each channel. The purpose of CAR is to reduce the influence of common mode and improve signal quality

$$\mathbf{X}_{CAR} = \mathbf{X}_{reshaped} - \frac{1}{C}\sum_{c=1}^{C}\mathbf{X}_{reshaped}[c,:] \tag{10}$$

  where $\mathbf{X}_{reshaped}[c,:]$ denotes the EEG data from the c-th channel across all time points. The notation (':') indicates that all elements along the second dimension (time points) are included. C represents the total number of channels. This operation computes the average signal across all channels and subtracts it from each channel's signal, resulting in the re-referenced data $\mathbf{X}_{CAR}$.
- **Applying Bandpass Filter:** A bandpass filter is applied to the EEG data to retain only the frequencies of interest. For SSVEP data, this typically involves filtering the data to keep frequencies within the range of the SSVEP stimuli (4-16 Hz). This step helps in removing noise and irrelevant frequency components.

$$\mathbf{X}_{filtered} = \text{bandpass\_filter}(\mathbf{X}_{CAR}, 4\text{ Hz}, 16\text{ Hz}) \tag{11}$$

  where 'bandpass_filter' represents the operation of applying a bandpass filter to the data. Specifically, it filters the data $\mathbf{X}_{CAR}$ to retain frequencies between 4 Hz and 16 Hz, which correspond to the typical frequency range of SSVEP stimuli.
- **Normalizing the EEG Data:** Normalization is performed to standardize the EEG data, ensuring all signals have a mean of zero and a standard deviation of one. This step reduces variability due to different measurement scales and improves the performance of subsequent analysis methods

$$\mathbf{X}_{normalized} = \frac{\mathbf{X}_{filtered} - \mu_{\mathbf{X}}}{\sigma_{\mathbf{X}}} \tag{12}$$

  where $\mu_{\mathbf{X}}$ and $\sigma_{\mathbf{X}}$ are the mean and standard deviation of the filtered data respectively.

- **Creating Event Array:** An event array is created to mark specific events in the EEG data, such as the onset of a stimulus. This array is essential for segmenting the continuous EEG data into epochs corresponding to individual trials

$$\text{events} = \text{create\_events}(\mathbf{X}_{normalized}, \mathbf{Y}_{raw}) \tag{13}$$

  where 'create_events' denotes the process used to identify and record the timing of significant events (e.g., stimuli presentation) within the normalized EEG data $\mathbf{X}_{normalized}$, based on the raw labels $\mathbf{Y}_{raw}$.
- **Creating Epochs:** The continuous EEG data is segmented into epochs, time-locked segments around the events of interest (the presentation of SSVEP stimuli).

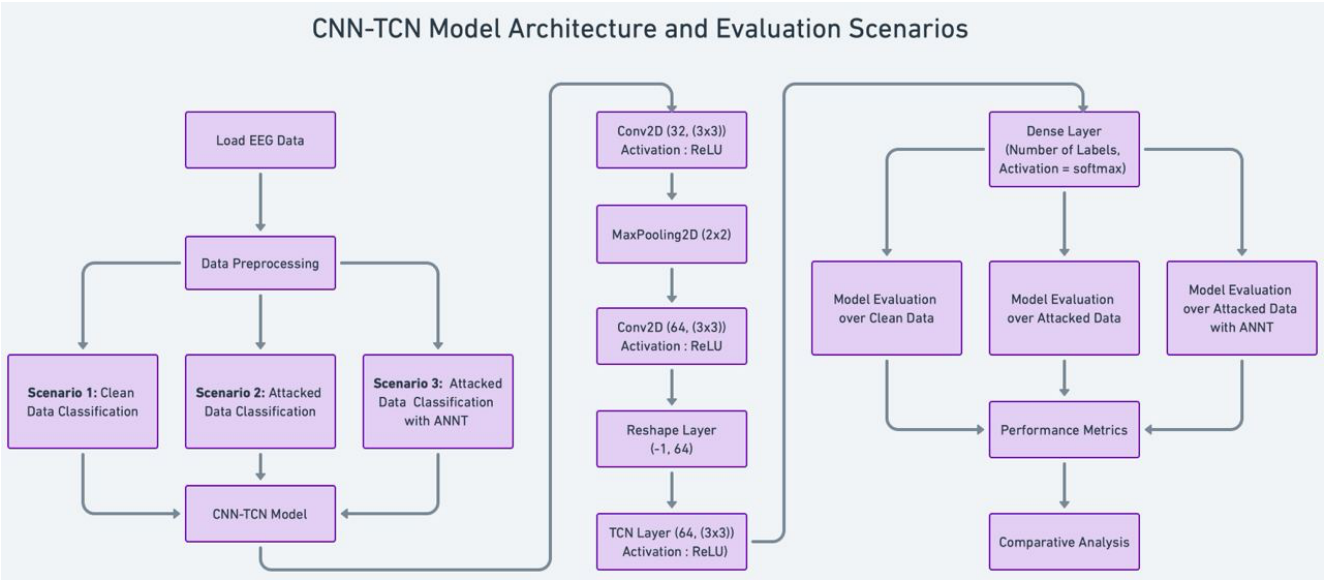


FIGURE 2: CNN-TCN Model Architecture and Evaluation Scenarios. This figure outlines the overall model architecture, showing the flow from loading EEG data, preprocessing, and CNN-TCN model layers to evaluating clean and attacked data scenarios, with and without ANNT [25].

Each epoch corresponds to a single trial and contains the EEG data from a specified time window around the event

$$\text{epochs} = \text{create\_epochs}(\mathbf{X}_{normalized}, \text{events}) \quad (14)$$

where 'create_epochs' refers to the function that segments the continuous EEG data into distinct time periods (epochs) around the events, as defined by the 'events' array.

- **Creating Reference Signals:** Reference signals for the SSVEP stimuli are created based on the known flicker frequencies. These reference signals are used for subsequent analysis and classification of the SSVEP responses

$$\mathbf{R} = \text{create\_reference\_signals}(\text{frequencies}) \quad (15)$$

where 'create_reference_signals' indicates the function used to generate reference signals corresponding to the SSVEP stimuli's known frequencies.

- **Reshaping Data for CNN-TCN:** The EEG data epochs are reshaped into a format suitable for input to the CNN-TCN. This involves organizing the data into 3D arrays where each dimension represents time points, channels, and epochs

$$\mathbf{X}_{CNN-TCN} = \text{reshape\_for\_cnn\_tcn}(\text{epochs}) \quad (16)$$

where 'reshape_for_cnn_tcn' is the function that restructures the segmented EEG data ('epochs') into a format compatible with the CNN-TCN model.

- **Mapping Frequency Labels to Numerical Indices:** The frequency labels associated with the SSVEP stimuli are mapped to numerical indices to facilitate their use in DL algorithms. Each unique frequency is assigned a distinct numerical index

$$\mathbf{Y}_{indices} = \text{map\_labels}(\mathbf{Y}_{raw}) \quad (17)$$

where 'map_labels' describes the process of converting the raw frequency labels ($\mathbf{Y}_{raw}$) into numerical indices ($\mathbf{Y}_{indices}$). This conversion is necessary for training the model.

- **Converting Labels to One-Hot Encoding:** The numerical labels are converted to one-hot encoding, a binary representation of categorical data. This step is necessary for training the CNN-TCN, allowing the network to output probabilities for each class

$$\mathbf{Y}_{one\_hot} = \text{one\_hot\_encoding}(\mathbf{Y}_{indices}) \quad (18)$$

where 'one_hot_encoding' is the function that converts the numerical labels ($\mathbf{Y}_{indices}$) into a one-hot encoded format ($\mathbf{Y}_{one\_hot}$).

- **Splitting Data into Training and Testing/Validation Sets:** Finally, the preprocessed data is split into training (80%) and testing/validation (20%) sets. This ensures that the model is trained on one subset of the data and evaluated on another unseen subset, enabling an unbiased assessment of its performance

$$\mathbf{X}_{train}, \mathbf{X}_{test}, \mathbf{Y}_{train}, \mathbf{Y}_{test} = \text{split\_data}(\mathbf{X}_{CNN-TCN}, \mathbf{Y}_{one\_hot}) \quad (19)$$

where 'split_data' refers to the function that divides the processed data into training and testing/validation sets. $\mathbf{X}_{train}$ and $\mathbf{Y}_{train}$ represent the input data and labels for training, while $\mathbf{X}_{test}$ and $\mathbf{Y}_{test}$ are used for evaluation.

Following these preprocessing steps ensures that the EEG data is clean, standardized, and formatted correctly for analysis using advanced DL techniques.

E. MATHEMATICAL FORMULATION

In this section, we provide the mathematical formulation of the key steps involved in our system model, from data loading and preprocessing to adding perturbations, training the CNN-TCN model, and evaluating its performance.

- **Data Loading and Preprocessing**
  First, we load and preprocess the SSVEP EEG datasets described in the previous section to ensure they are ready for analysis.

$$\mathbf{X}_{\text{preprocessed}} = \text{preprocess}(\mathbf{X}_{\text{raw}}) \tag{20}$$

- **Adding Perturbations**
  Next, we introduce adversarial perturbations to the preprocessed data to evaluate the model's robustness. The general formula for an adversarial attack is:

$$\mathbf{x}' = \mathbf{X}_{\text{preprocessed}} + \delta, \tag{21}$$

  where δ represents the perturbation added to the original input $\mathbf{X}_{\text{preprocessed}}$. This perturbation can be generated using any of the five attacks (FGSM, PGD, BIM, C&W, MIM) or their combinations.

- **Training the CNN-TCN Model**
  The preprocessed and potentially perturbed data is then used to train the CNN-TCN model

$$\theta^* = \arg\min_{\theta} \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} J(\theta, \mathbf{X}^{(i)}_{\text{preprocessed}}, y_i) \tag{22}$$

  where θ^* represents the optimized vector parameters of the model after training, θ represents the model's set of parameters, N_{test} is the number of test samples, J is the loss function, $\mathbf{X}^{(i)}_{\text{preprocessed}}$ is the input data, and y_i are the labels.

- **Model Evaluation Scenarios**
  The model's performance is evaluated across three scenarios:
  -- Scenario 1: Clean Data Classification:

$$\hat{y} = \arg\max(\text{CNN-TCN}(\mathbf{X}_{\text{preprocessed}})) \tag{23}$$

  In this scenario, the model classifies clean, unperturbed data to assess its baseline performance.
  -- Scenario 2: Attacked Data Classification:

$$\hat{y'} = \arg\max(\text{CNN-TCN}(\mathbf{x}')) \tag{24}$$

  In this scenario, the model's performance is evaluated on perturbed data without ANNT. This scenario helps assess how well the model can classify inputs that have been perturbed by adversarial attacks without specific training to handle such perturbations.
  -- Scenario 3: Attacked Data Classification with ANNT:

$$\hat{y''} = \arg\max(\text{CNN-TCN}_{\text{ANNT}}(\mathbf{x}')) \tag{25}$$

  In this scenario, the model's robustness is tested using adversarially perturbed data, similar to Scenario 2. However, the difference lies in the training process: the model, $\text{CNN-TCN}_{\text{ANNT}}$, has been trained using ANNT. This scenario evaluates the effectiveness of ANNT in improving the model's resilience to adversarial attacks.
- **Performance Metrics**
  Finally, we assess the model's performance using accuracy and AUC metrics:
  -- Accuracy:

$$\text{Accuracy} = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} \mathbf{1}(\hat{y}_i = y_i) \tag{26}$$

  where $\hat{y}_i$ is the predicted label for the i-th sample.
  -- AUC:

$$\text{AUC} = \int_0^1 \text{TPR}(x)\, dx \tag{27}$$

  where TPR is the true positive rate.

F. COMPUTATIONAL TIME EVALUATION

In addition to accuracy and AUC metrics, we also evaluated the computational time required for each scenario. Given that the computational time varies significantly depending on the hardware characteristics of each computer used, presenting absolute times would not provide a consistent basis for comparison. Therefore, we normalized the computational times to provide a hardware-independent relative measure.

The normalization process involved recording the actual computational times for all scenarios and then expressing these times as a percentage of the minimum time observed. The normalization formula is:

$$S_i = \frac{T_i}{T_{\text{min}}} \times 100 \tag{28}$$

where T_i is the computational time for the i-th scenario, and T_{min} is the minimum computational time observed. This formula expresses each scenario's computational time as a percentage of the minimum observed time. This method allows us to present a standardized comparison independent of specific hardware configurations, providing a clear and quantitative measure of the computational demands for each scenario. This approach allows us to easily identify scenarios with higher computational costs and compare the efficiency of different methods.

By following these steps, we thoroughly evaluate the robustness and security of our B2B-C system against various adversarial attacks. The analysis includes assessing the model's performance on clean data, perturbed data without ANNT, and perturbed data with ANNT, providing a comprehensive view of its effectiveness and resilience. Additionally, the normalized computational time offers insights into the efficiency of each scenario, highlighting the computational cost associated with different attack combinations.

III. RESULTS

In this section, we present our findings based on accuracy, AUC, and computational time for each scenario to demonstrate the effectiveness and robustness of our proposed model against various adversarial attacks. First, we comprehensively evaluate our proposed model's performance across various scenarios. Table 3 summarizes the results of our experiments, highlighting the accuracy, AUC, and computational time for each attack scenario with and without the application of ANNT. This table overviews how our model improves accuracy and AUC and the computational time required under different attack conditions. Detailed explanations of these findings will be discussed in the following section.

To better understand the effects of ANNT and the behavior of the attacks on the signal, we present the Receiver Operating Characteristic (ROC) curves for the weakest and strongest individual attacks on both datasets in Figures 3, 4, 5, and 6. While showing the visual representation for all scenarios is impractical due to their extensive number, these selected figures provide critical insights. They illustrate how ANNT enhances the robustness of our model against the most and least challenging adversarial attacks individually, thereby demonstrating the overall effectiveness of our approach.

To further elaborate on the impact of ANNT and the behavior of the adversarial attacks on the signal, we present additional analyses in Figures 7, 8, 9 and Table 4.

Figure 7 consists of bar charts comparing the accuracy and AUC improvements achieved by applying ANNT across different adversarial attack scenarios. The results consistently show higher improvements for the Lee2019_SSVEP dataset, indicating the greater enhancement in both accuracy and AUC due to ANNT.

Table 4 provides a detailed breakdown of the least and most effective attack scenarios for the Nakanishi2015 and Lee2019_SSVEP datasets. The metrics include accuracy with and without ANNT, accuracy and AUC improvements, and computational time for different attack combinations. We have not included the scenario with five attacks, as there is only one such scenario, making it meaningless to compare it with itself.

Figure 8 displays the distribution of accuracy and AUC improvements across different attack scenarios. For both datasets, the plots indicate that as the number of combined attacks increases, the improvements in accuracy and AUC due to ANNT also increase. The Lee2019_SSVEP dataset generally shows larger improvements and a higher baseline improvement level than the Nakanishi2015 dataset, although there is some variability in the results.

Figure 9 shows the computational time required for different adversarial attack scenarios. The computational time is generally higher for the Lee2019_SSVEP dataset, reflecting the increased complexity and size of this dataset.

IV. DISCUSSION

A. INDIVIDUAL ATTACKS COMPARISON

In this study, we observed varying impacts of different adversarial attacks on the Nakanishi2015 and Lee2019_SSVEP datasets as detailed in Table 3. Further analysis in Table 4 revealed that PGD consistently caused the most significant decrease in accuracy and AUC, demonstrating its strength as an adversarial method. This attack, known for its iterative refinement process, applies perturbations that effectively reduce the model's performance, highlighting its effectiveness in compromising its robustness [29].

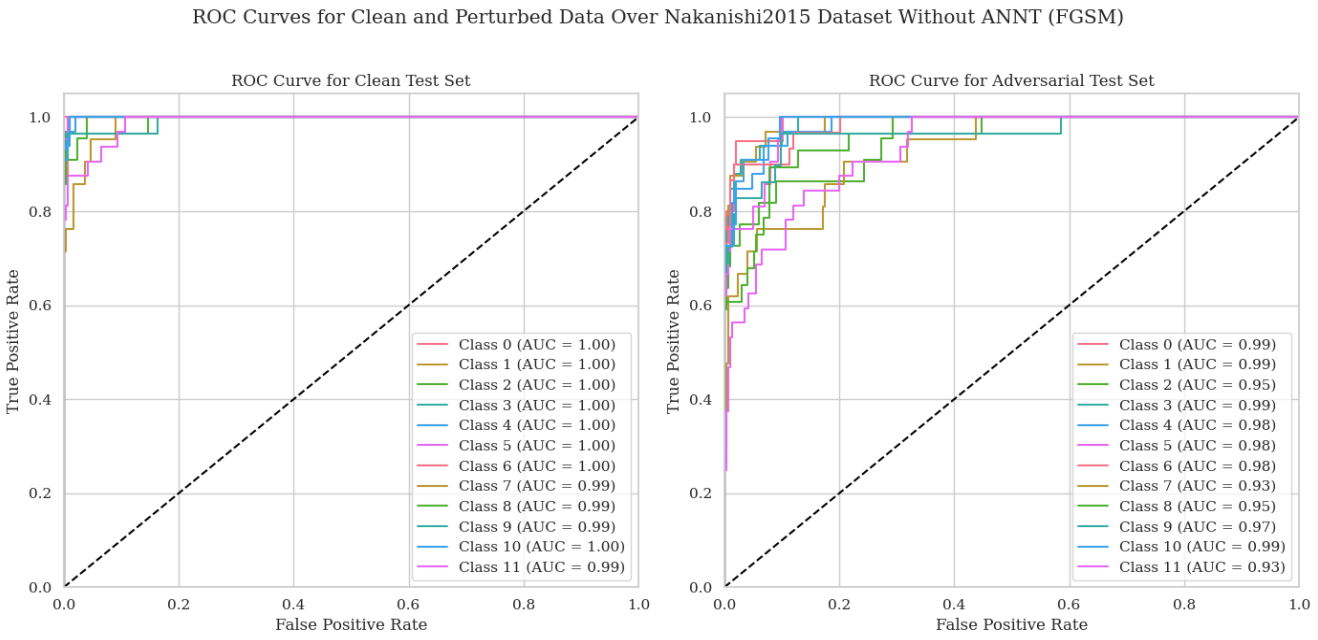
FGSM, on the other hand, was consistently the weakest attack across both datasets. It resulted in the least reduction in accuracy and AUC, maintaining relatively high values even under adversarial conditions. This outcome aligns with FGSM's simpler and less potent mechanism than more sophisticated attacks like PGD or BIM [28]. Figures 3 and 4 illustrate this point, as the ROC curves show relatively minor drops in performance, indicating FGSM's limited impact on the model's robustness. The effectiveness of PGD also can be observed in Figures 5 and 6, where the ROC curves for the adversarial test sets show substantial performance degradation compared to clean test sets, even with ANNT applied.

BIM, being an iterative method similar to PGD, also showed strong negative impacts, though slightly less than PGD. Its effectiveness underscores the role of iterative gradient-based methods in adversarial attacks.

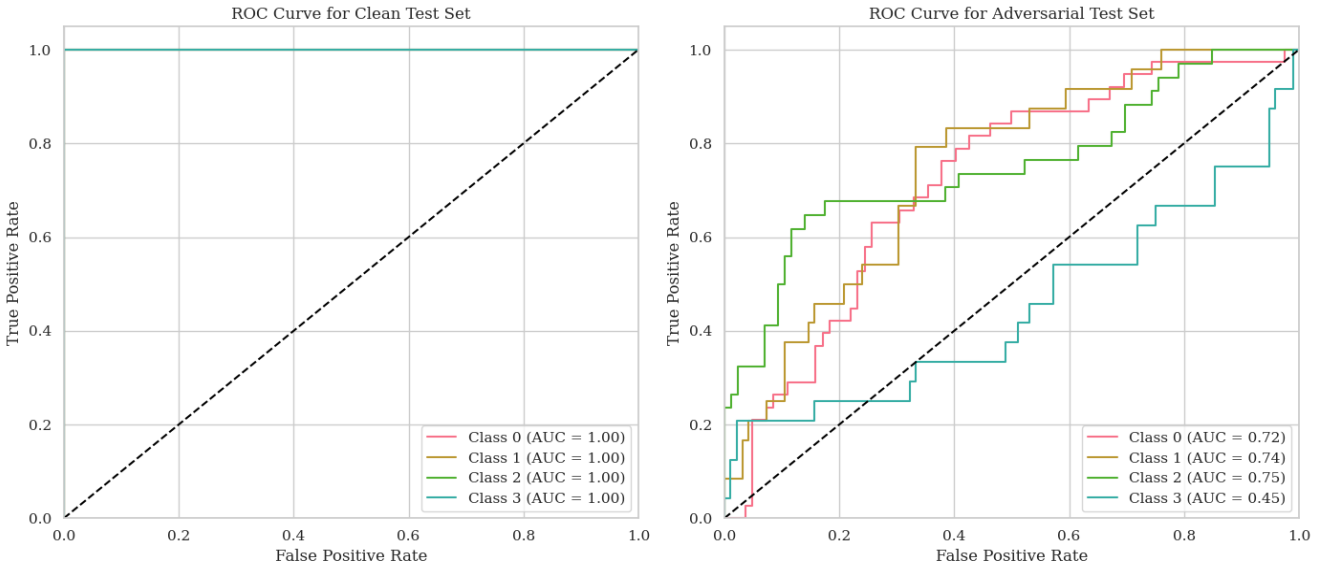
While C&W shows less impact in the Lee2019_SSVEP dataset, it is more impactful in combination scenarios and with the Nakanishi2015 dataset. Its characteristic of creating minimal and often imperceptible perturbations means that it may not drastically reduce performance metrics when used alone but can be effective when combined with other attacks, particularly those like BIM and PGD, that exploit different aspects of model vulnerabilities.

MIM, with its momentum-based approach, showed strong adversarial effects, particularly in combination with other attacks. This method's ability to maintain attack direction and avoid oscillations made it particularly challenging for the model to defend against.

The ROC curves in Figure 3 for the clean test set in the Lee2019_SSVEP dataset reveal near-perfect classification performance for all four classes, with AUC values of 1.0. This outcome can be attributed to the high signal-to-noise ratio of SSVEP signals, which are frequency-locked to external visual stimuli and therefore easier to distinguish compared to paradigms like MI or resting-state EEG. Additionally, the dataset includes only four classes, reducing the complexity of the classification task. The robust preprocessing pipeline



(a) ROC Curves for Nakanishi2015 dataset without ANNT (FGSM).

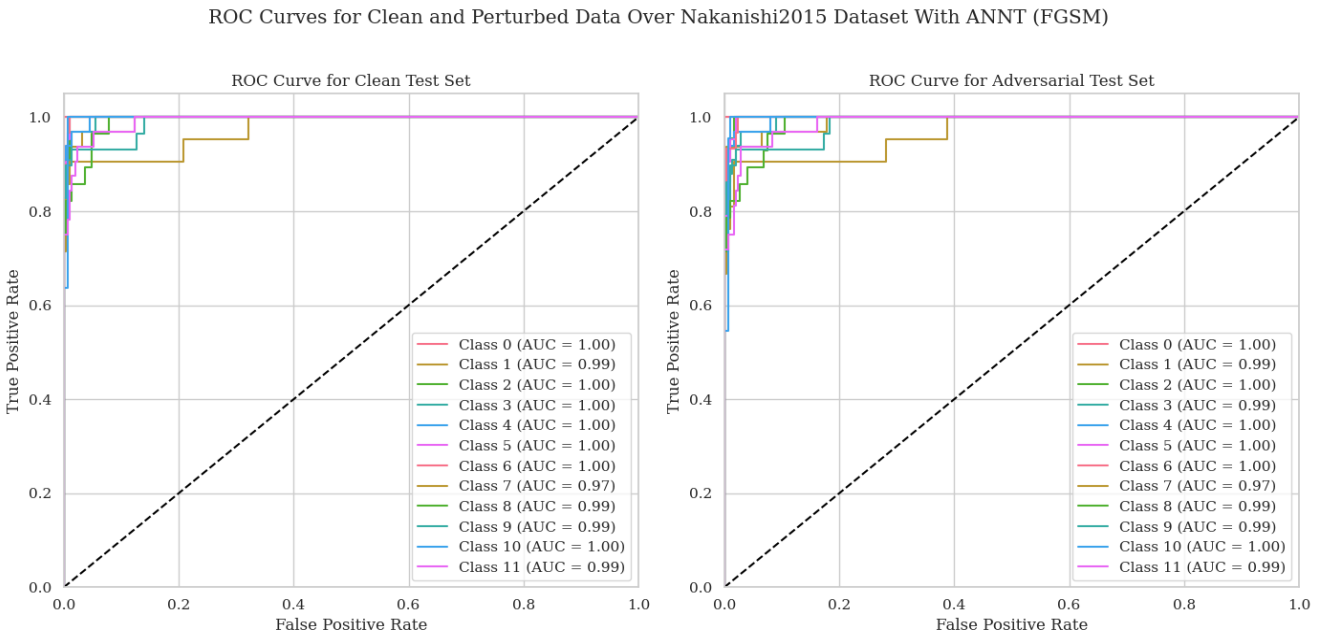


(b) ROC Curves for Lee2019_SSVEP dataset without ANNT (FGSM).

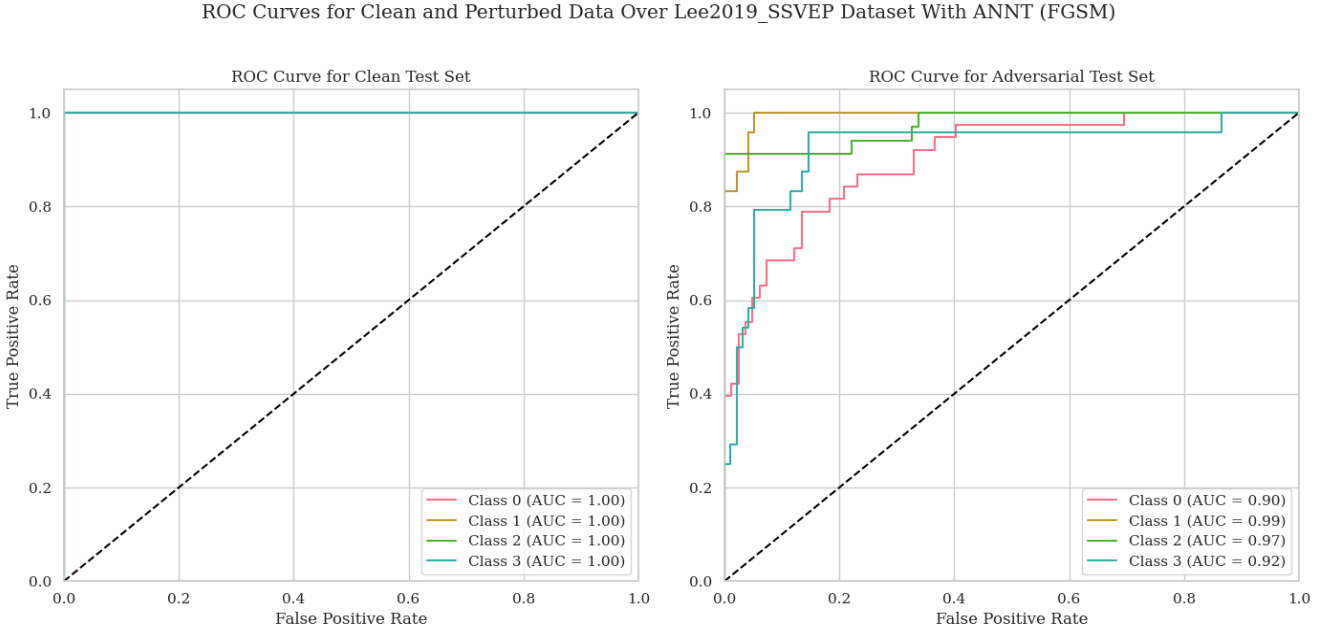
FIGURE 3: ROC curves for clean and perturbed data without ANNT (FGSM). The top row presents the ROC curves for the Nakanishi2015 dataset, while the bottom row shows the ROC curves for the Lee2019_SSVEP dataset. In the clean test set (left panels), the Lee2019_SSVEP dataset shows perfect classification performance across all classes, reflecting the high signal-to-noise ratio of SSVEP signals, a simplified four-class classification task, and robust preprocessing. In the adversarial test set (right panels), both datasets show reduced AUC values under FGSM attack, highlighting vulnerabilities when adversarial training is not applied.

further enhances the data quality, contributing to this result. Finally, the CNN-TCN architecture effectively leverages the temporal and spatial characteristics of SSVEP signals to ensure highly accurate classification.

In contrast, the adversarial test set shows notable performance degradation, particularly in the Lee2019_SSVEP



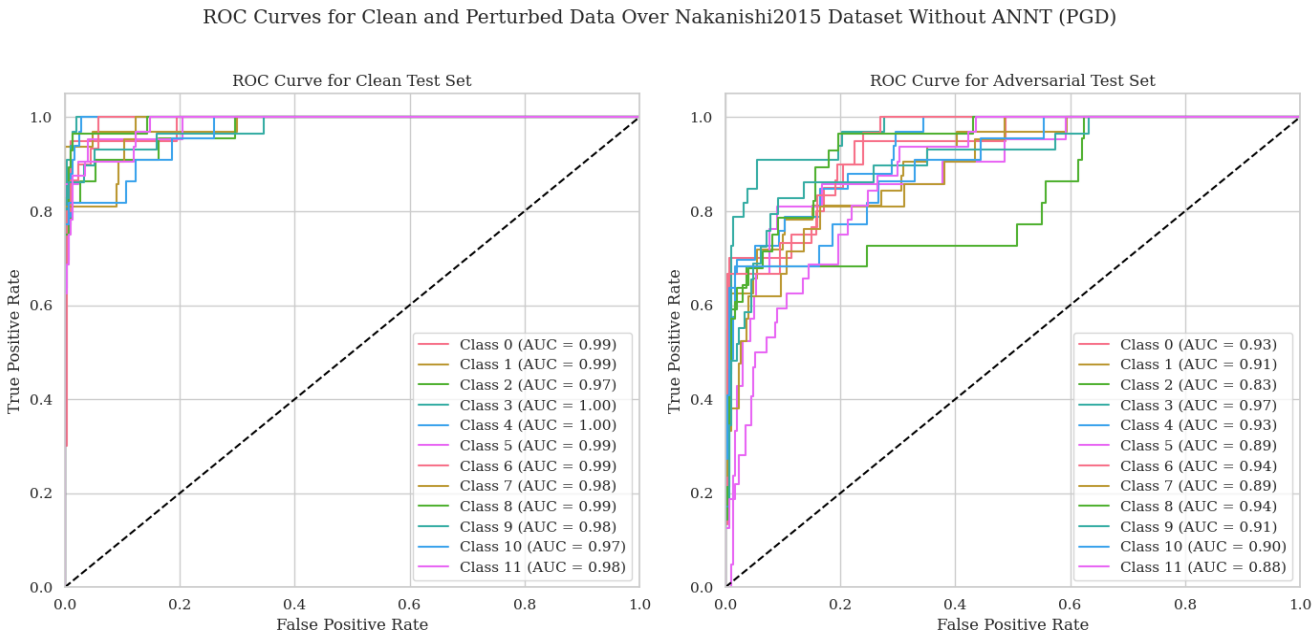
(a) ROC Curves for Nakanishi2015 dataset with ANNT (FGSM).



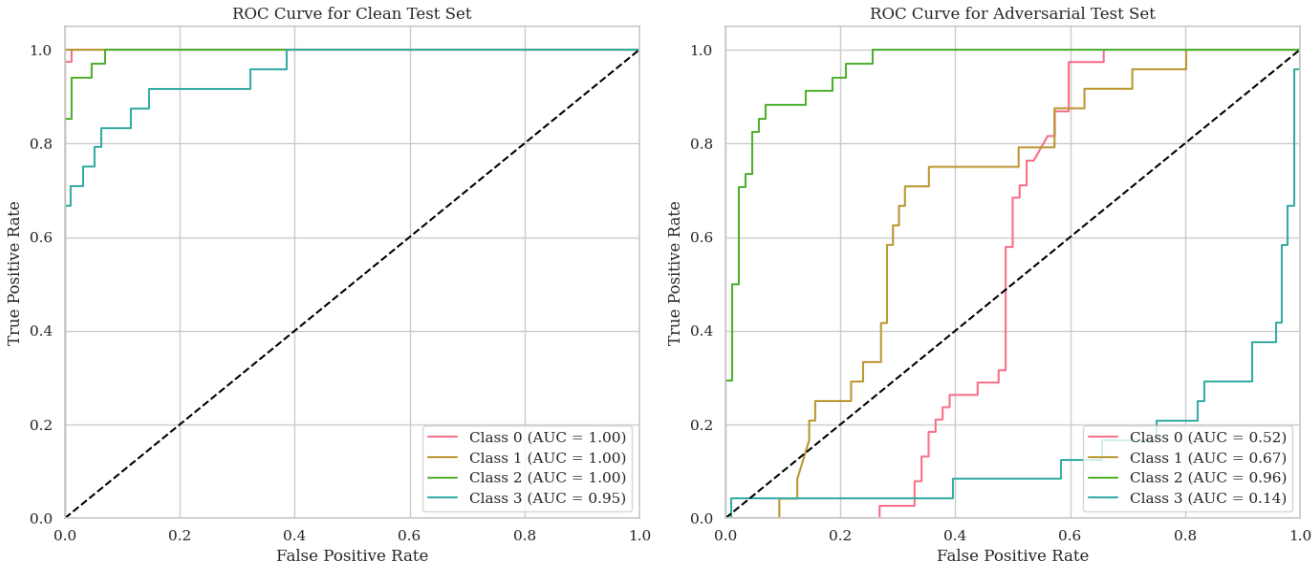
(b) ROC Curves for Lee2019_SSVEP dataset with ANNT (FGSM).

FIGURE 4: ROC curves for clean and perturbed data with ANNT (FGSM). The left subfigure presents the ROC curves for the clean test sets, while the right subfigure shows the ROC curves for the adversarial test sets attacked with FGSM. The AUC values for each class are indicated, demonstrating the impact of the FGSM attack on the model's performance and the effectiveness of ANNT in mitigating this impact.

dataset, where AUC values drop significantly for some classes (e.g., 0.45 for Class 3). This degradation highlights the susceptibility of the model to adversarial perturbations in the absence of ANNT. The stark contrast between clean and perturbed performance underscores the necessity of adversarial training for securing SSVEP-based systems, as the model is otherwise vulnerable to even simple attacks like FGSM.



(a) ROC Curves for Nakanishi2015 dataset without ANNT (PGD).

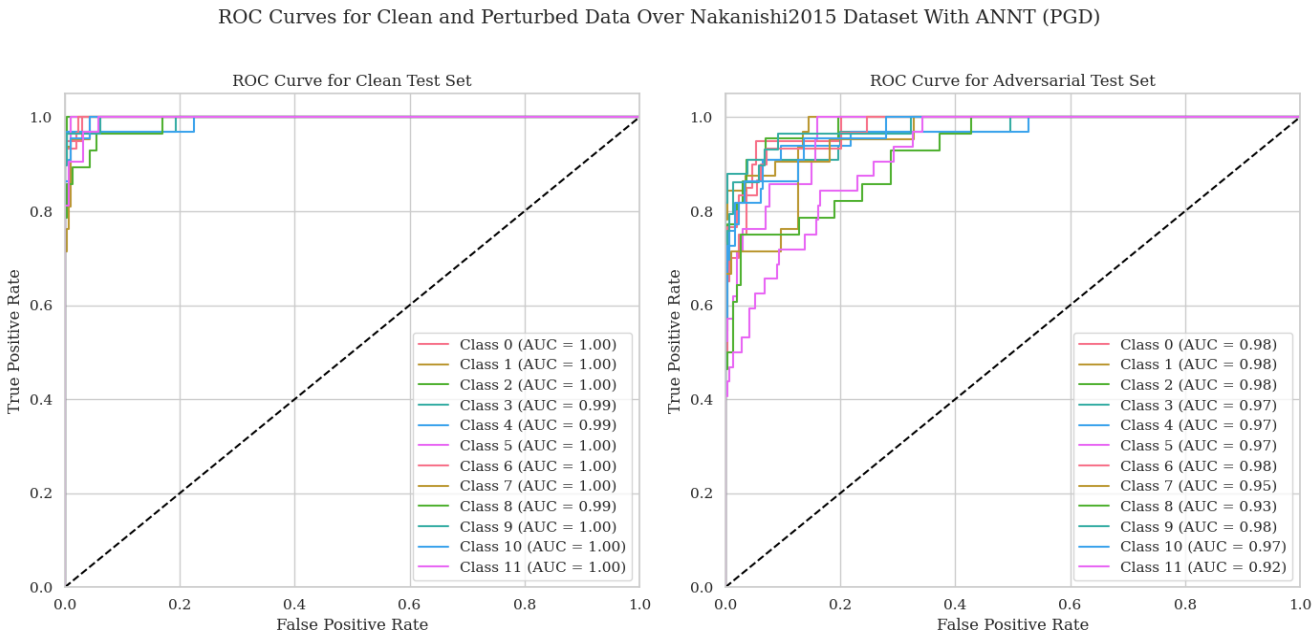


(b) ROC Curves for Lee2019_SSVEP dataset without ANNT (PGD).

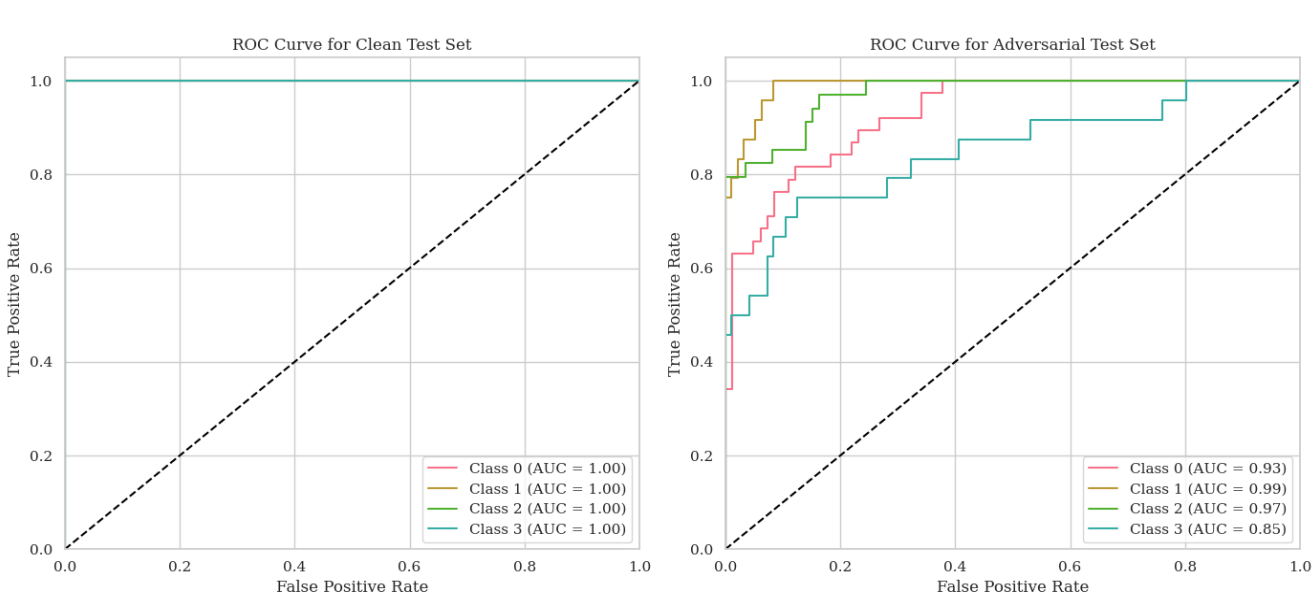
FIGURE 5: ROC curves for clean and perturbed data without ANNT (PGD). The top subfigure presents the ROC curves for the Nakanishi2015 dataset, with the left graph showing the performance on the clean test set and the right graph displaying the performance on the adversarial test set. Similarly, the bottom subfigure presents the ROC curves for the Lee2019_SSVEP dataset, with the left graph showing the clean test set performance and the right graph displaying the adversarial test set performance. These ROC curves illustrate the trade-off between true positive and false positive rates, providing a detailed view of the model's ability to distinguish between classes under clean and perturbed conditions without applying ANNT.

B. PERFORMANCE IMPROVEMENT WITH INCREASING NUMBER OF ATTACKS

As illustrated in Table 3, model performance evaluation across different attack scenarios with and without ANNT reveals that introducing multiple attacks generally leads to a greater decrease in accuracy and AUC. However, it is also evident that the application of ANNT significantly improves



(a) ROC Curves for Nakanishi2015 dataset with ANNT (PGD).



(b) ROC Curves for Lee2019_SSVEP dataset with ANNT (PGD).

FIGURE 6: ROC curves for clean and perturbed data with ANNT (PGD). The top subfigure presents the ROC curves for the Nakanishi2015 dataset, with the left graph showing the performance on the clean test set and the right graph displaying the performance on the adversarial test set. Similarly, the bottom subfigure presents the ROC curves for the Lee2019_SSVEP dataset, with the left graph showing the clean test set performance and the right graph displaying the adversarial test set performance. These ROC curves illustrate the trade-off between true positive and false positive rates, providing a detailed view of the model's ability to distinguish between classes under clean and perturbed conditions with ANNT applied.

these metrics. This trend is further supported by the data from Table 4, which provide a breakdown of the metrics for the Nakanishi2015 and Lee2019_SSVEP datasets.

Figure 7 illustrates the improvements in accuracy and AUC across various attack scenarios as the number of combined attacks increases. This consistent upward trend demonstrates

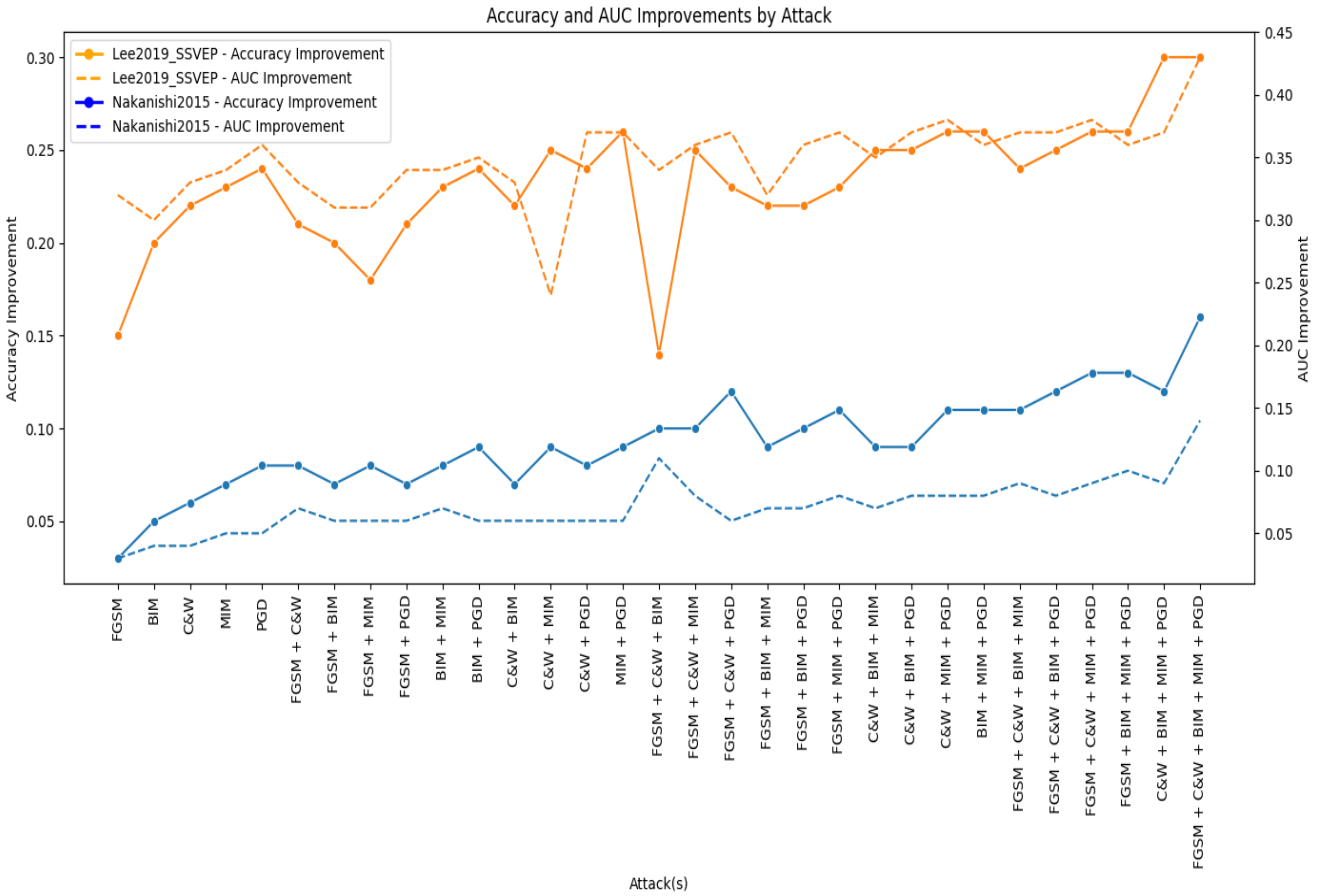
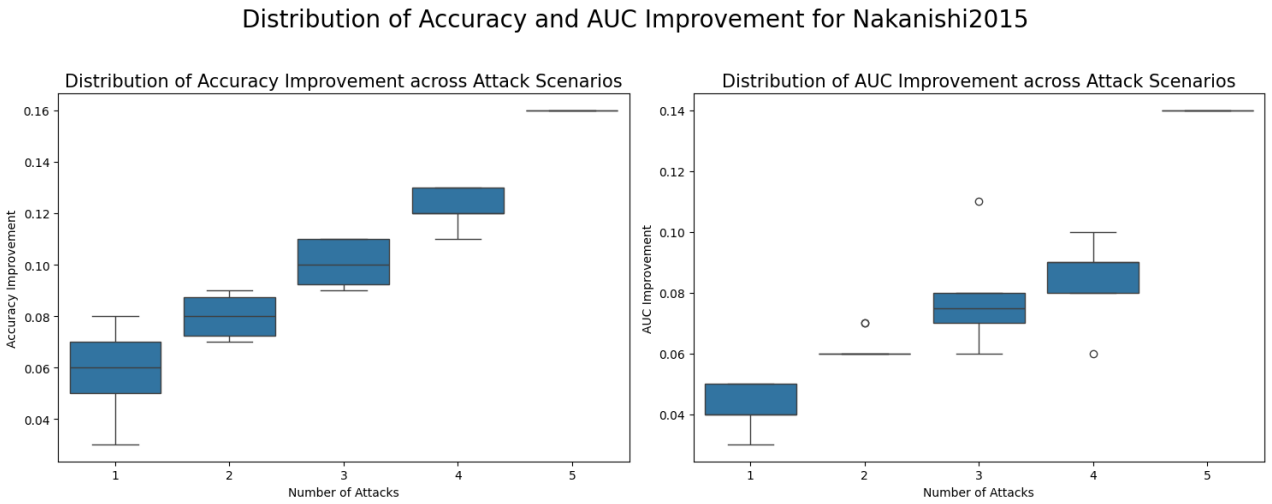


FIGURE 7: This figure illustrates the accuracy and AUC improvements achieved by applying ANNT across various adversarial attack scenarios. The Lee2019_SSVEP dataset consistently demonstrates higher enhancements in both accuracy and AUC compared to the Nakanishi2015 dataset, highlighting the effectiveness of ANNT in counteracting adversarial attacks.

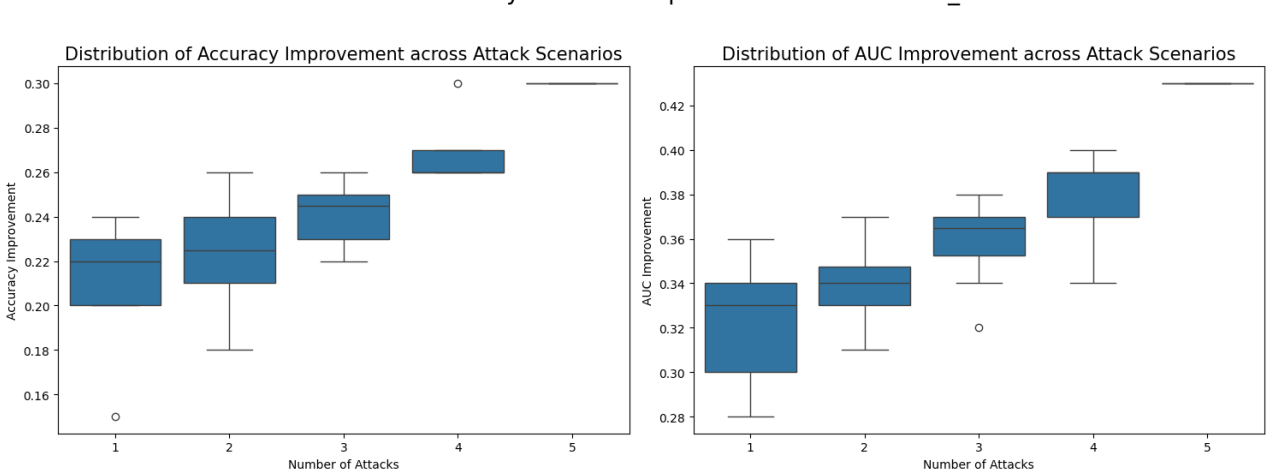
the enhanced capability of ANNT to strengthen model robustness. In Figure 8, the distribution of these improvements is presented, further emphasizing how ANNT's effectiveness becomes more pronounced with a wider range of adversarial strategies. This pattern highlights ANNT's ability to adapt and counter multiple combined attacks, a trend that can be attributed to several key factors:

- **Increased Data for Learning:** When combining multiple attacks, the adversarial dataset becomes more diverse. This diversity presents the model with a wider range of perturbations during training, effectively increasing the data the model can learn from. DL models typically perform better with more data, allowing them to generalize to unseen scenarios better. Therefore, by exposing the model to various attack strategies, ANNT enhances the model's ability to learn robust features resistant to adversarial perturbations.
- **Learning to Defend Against Multiple Attacks:** Combining multiple attacks forces the model to learn defenses against different perturbations. For instance, PGD, a strong attack, can push the model's decision boundary in a specific way, while a weaker attack like FGSM may do so differently. The model learns a more generalized defense strategy by encountering both in training, covering a broader range of possible adversarial examples. This comprehensive learning process strengthens the model's robustness, particularly when facing weaker attacks within the combined scenarios.
- **Targeting Weak Attacks More Effectively:** When attacks are presented in isolation, the model may not have sufficient exposure to learn robust defenses against them, especially for weaker attacks like FGSM or C&W (as indicated by their higher accuracy and AUC in Table 4. However, in a combined attack scenario, these weak attacks are reinforced by stronger attacks, leading the model to develop more effective countermeasures. This results in greater overall improvements in performance metrics when ANNT is applied, as the model has effectively 'learned' from the stronger perturbations how to handle the weaker ones better.

The observed performance improvements with ANNT, particularly the 24% increase in accuracy and a 0.23-point improvement in AUC for the Lee2019_SSVEP dataset, can be attributed to several key factors. First, the Lee2019_SSVEP



(a) Distribution of Accuracy and AUC Improvement for Nakanishi2015.



(b) Distribution of Accuracy and AUC Improvement for Lee2019_SSVEP.

FIGURE 8: These box plots display the accuracy and AUC improvements distribution across different attack scenarios. For both datasets, the plots indicate that as the number of combined attacks increases, the improvements in accuracy and AUC due to ANNT also increase. The Lee2019_SSVEP dataset generally shows larger improvements and a higher baseline improvement level than the Nakanishi2015 dataset, although there is some variability in the results.

dataset has a higher sampling rate of 1000 Hz, providing richer temporal resolution and capturing fine-grained neural patterns. This high granularity allows the CNN-TCN model, combined with ANNT, to extract more robust and detailed features that are critical in distinguishing between adversarially perturbed and clean signals. The enhanced feature extraction capabilities improve the model's ability to generalize across both clean and adversarially perturbed data, leading to significant gains in classification performance.

Additionally, the Lee2019_SSVEP dataset's larger subject pool (54 subjects compared to 9 in the Nakanishi2015 dataset) introduces more variability and diversity in the training data. This diversity, when coupled with ANNT's ability to adaptively learn from a broad spectrum of perturbations, enhances the model's robustness to adversarial attacks. Conversely, the Nakanishi2015 dataset exhibits a lower temporal resolution due to its sampling rate of 256 Hz and contains fewer trials per class. These limitations reduce the richness of extracted features and the variability in training data, which may explain the relatively smaller improvements of 9% in accuracy and 0.07 points in AUC.

Another contributing factor to the disparity in improvements is the number of classes. The Lee2019_SSVEP dataset contains only 4 classes, simplifying the classification task compared to the Nakanishi2015 dataset's 12 classes. This difference allows the model to focus its learning capacity more effectively on refining its defenses against adversarial perturbations in the simpler classification scenario.

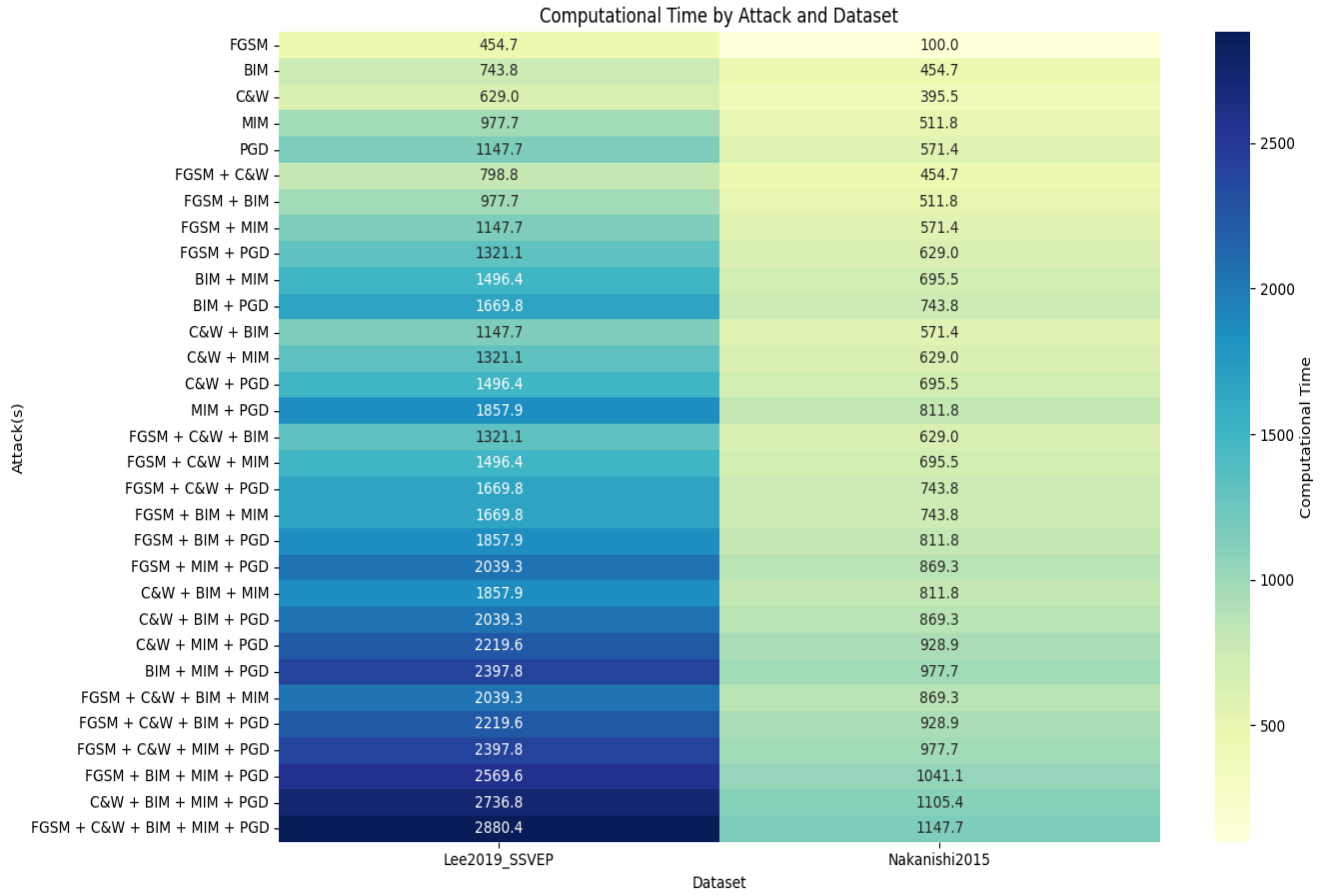


FIGURE 9: This heatmap illustrates the normalized computational time (expressed as a percentage of the minimum time observed) required for different combinations of adversarial attacks. The Lee2019_SSVEP dataset generally exhibits higher computational times than the Nakanishi2015 dataset, highlighting the impact of dataset-specific characteristics on processing requirements.

In summary, the model's enhanced learning opportunities can explain the trend that ANNT leads to greater performance improvements with an increasing number of combined attacks. The combination of diverse attacks increases the amount and diversity of training data and forces the model to generalize its defenses across a wider spectrum of adversarial examples. This comprehensive exposure allows the model to develop a more robust and versatile defense mechanism, making it better equipped to handle various types of perturbations, whether weak or strong. This detailed understanding of the interplay between attack combinations and model defense mechanisms underscores the importance of comprehensive adversarial training in enhancing the robustness of ML models.

C. COMPARISON ACROSS DATASETS

The impact of these attacks varied between the two datasets, influenced by their characteristics. The Lee2019_SSVEP dataset, with its higher sampling rate of 1000Hz and more complex setup, generally exhibited higher improvements in both accuracy and AUC after applying ANNT. Conversely, the Nakanishi2015 dataset, with a lower sampling rate of 256Hz and a more limited number of trials/class, showed smaller improvements. The lower temporal resolution restricted the richness of the extracted features, making the model less robust against adversarial attacks. However, the dataset's greater number of classes (12 classes versus 4 in the Lee2019_SSVEP) posed a more challenging classification task, which could also account for the different levels of improvements observed.

In our previous work [25], we found that the sampling rate was the most crucial parameter among dataset characteristics, primarily because it directly influences the temporal resolution of the data. This higher temporal resolution allows for more precise alignment and potentially better signal quality, crucial for identifying subtle patterns in neural signals, even within the filtered frequency range.

Thus, this research not only reaffirms our previous findings about the paramount importance of sampling rate but also extends them by demonstrating the nuanced effects of dataset characteristics on the effectiveness of adversarial neural network training. This further solidifies the role of high temporal

resolution in enhancing the robustness and security of B2B-C systems against adversarial threats despite the filtering that limits the frequency range of interest.

D. COMPUTATIONAL TIME TRENDS AND OPTIMIZATION STRATEGIES

Figure 9 provides a comprehensive view of the computational time required for different attack scenarios across both datasets. The chart shows an upward trend in computational time as the number of combined attacks increases, highlighting a critical challenge associated with ANNT.

The computational time grows significantly as the number of attacks increases, primarily due to the complexity and variety of adversarial examples the model must handle. With each additional attack, the model's training process becomes more demanding as it must learn to defend against a broader range of perturbations and process a more extensive dataset of adversarial examples. This increased data complexity requires more computational resources, leading to longer training times.

Lee2019_SSVEP dataset shows a steeper increase in computational time, especially as the number of combined attacks rises. The high sampling rate of 1000Hz contributes to this trend, as it provides finer temporal resolution, resulting in more detailed and larger data points. This increased data granularity requires more computation for feature extraction and model training, extending the training time significantly when multiple attacks are involved.

In contrast, the Nakanishi2015 dataset exhibits a more gradual increase in computational time. With a lower sampling rate of 256Hz, the data is less granular, leading to fewer data points and less complexity in feature extraction. As a result, the computational load is relatively lighter, even as the number of combined attacks increases. However, despite the less steep trend, the absolute computational time still rises due to the additional complexity of handling multiple adversarial attacks.

The increased computational time with ANNT, especially with multiple attacks, poses a significant challenge. While ANNT enhances the model's robustness against various adversarial threats, it also imposes a substantial computational burden. This trade-off is one of the most pressing issues in implementing ANNT, as the benefits of improved model security and performance must be weighed against the practical limitations of increased computational costs.

The computational demand can become a bottleneck, especially in resource-constrained environments or when scaling up models for larger datasets or more complex tasks. To address these computational challenges, we propose several optimization strategies. Firstly, efficient adversarial example generation techniques, such as single-step or low-complexity iterative methods, can be prioritized during training to reduce computation without compromising robustness. Secondly, leveraging parallel processing on GPUs or distributed computing frameworks can significantly decrease training time. Thirdly, adaptive adversarial training approaches can be explored, wherein adversarial examples are selectively generated based on their impact on model robustness, reducing redundant computations. Finally, optimizing model architecture by pruning or quantization can further decrease computational requirements while maintaining accuracy and robustness. These strategies collectively pave the way for scaling ANNT to large datasets and more complex applications in practical deployments.

E. PRACTICAL IMPLICATIONS AND DEPLOYMENT CHALLENGES OF ANNT IN B2B-C SYSTEMS

The integration of ANNT into B2B-C systems offers significant advancements in robustness and security, enabling the potential for real-world applications in critical domains. For instance, healthcare can benefit from secure and reliable neural communication channels for neurorehabilitation and post-stroke motor training, ensuring accurate signal interpretation even under adversarial conditions. Similarly, in education and collaborative environments, B2B-C systems fortified with ANNT can facilitate seamless exchange of neural information, paving the way for innovative learning techniques and enhanced human collaboration. By mitigating adversarial risks, ANNT ensures system reliability in environments where data integrity is crucial, such as in high-stakes decision-making scenarios in defense or emergency response systems.

However, transitioning these advancements to practical deployments presents several challenges. Firstly, the computational cost associated with ANNT, particularly when addressing multiple adversarial attacks, remains a bottleneck, necessitating high-performance computing infrastructure. This limitation can be significant in real-time applications where low latency is critical. Additionally, the complexity of configuring ANNT models to handle diverse datasets and attack combinations may require domain-specific customizations, increasing development time and costs. Furthermore, ethical considerations surrounding the collection, storage, and transmission of neural data introduce regulatory hurdles that must be addressed to ensure compliance and user privacy. Overcoming these challenges will require collaborative efforts across neuroscience, ML, and system engineering to optimize computational efficiency, enhance adaptability, and establish robust data governance frameworks for B2B-C systems.

F. COMPARISON WITH RELATED TECHNOLOGIES

The application of adversarial training to SSVEP EEG signals represents a significant advancement over traditional robustness-enhancing methods and related approaches in the context of secure neural communication. Unlike regularization techniques such as dropout or L2 regularization, which improve generalization to unseen clean data, adversarial training directly targets vulnerabilities to adversarial perturbations, a critical aspect for real-world deployment of secure B2B-C systems.

Compared to ensemble learning approaches or noise filtering techniques, adversarial training uniquely equips the

model to recognize and defend against maliciously crafted adversarial examples, ensuring robustness in scenarios where signal integrity is paramount. Specifically, the high-frequency resolution and frequency-locked nature of SSVEP signals make them particularly susceptible to adversarial perturbations, yet these same characteristics provide a foundation for adversarial training to develop more resilient feature representations.

Another related method is data augmentation, which is commonly used to enhance model performance by synthetically increasing the variability of training data. While data augmentation can improve model generalization to natural variations in the data, it is less effective against adversarial examples specifically designed to exploit model vulnerabilities. Adversarial training, by contrast, proactively exposes the model to these perturbations during training, enabling it to identify and mitigate them during inference.

Additionally, denoising-based approaches have been proposed to enhance the robustness of EEG systems by filtering out noise and irrelevant frequency components. While effective in improving the signal-to-noise ratio, these methods do not address adversarial perturbations specifically, as adversarial noise is often designed to mimic relevant signal features. Adversarial training complements denoising techniques by directly targeting adversarial patterns, providing an added layer of robustness.

Our findings demonstrate that adversarial training on SSVEP signals not only enhances robustness across a wide range of adversarial scenarios but also ensures scalability to increasingly complex attack combinations. These contributions make adversarial training a pivotal advancement in securing neural communication systems, offering significant advantages over traditional robustness-enhancing methods and related technologies.

G. FUTURE RESEARCH DIRECTIONS:

For future research, several critical areas warrant exploration to enhance the robustness and efficiency of B2B-C systems:

- **Optimizing Computational Time:** This study identified one of the most pressing challenges: the high computational time associated with combined attacks. Future work should focus on developing optimized algorithms and techniques that reduce computational overhead without compromising the effectiveness of ANNT.
- **Larger and More Diverse Datasets:** Future research should focus on larger and more diverse datasets to validate these findings and understand the limitations of ANNT. Additionally, integrating multimodal EEG data from different paradigms could improve the generalizability and robustness of B2B-C systems.
- **Real-World Applications:** Implementing ANNT in real-world B2B-C systems will be crucial to understanding its practical implications and any challenges. Specific focus areas include neurorehabilitation, cognitive training, and secure communication in collaborative environments.
- **Normalization Limitations:** While reducing variability and improving analysis performance, the normalization process may lead to the loss of information regarding differences in EEG amplitude distribution on the scalp. Future research could explore advanced techniques to preserve this information while achieving the desired standardization.
- **Exploring Additional Attacks:** Future work should consider and apply alternative adversarial techniques, such as generative adversarial perturbations or physically realizable attacks, to further test the robustness and security of the B2B-C systems. Investigating countermeasures against these attacks will provide a more comprehensive understanding of system vulnerabilities and defenses.
- **Semantic Features Extraction:** Another promising avenue for future research is the exploration of semantic features in EEG signals. Semantic features could offer significant advantages given their relevance to B2B-C and potential resilience against adversarial attacks. These features are also well-suited for semantic communication, a field we intend to explore in depth in our next study. By leveraging semantic features, we aim to develop systems that are robust and capable of more sophisticated and meaningful communication.
- **Incorporating Additional EEG Modalities:** Future studies should investigate incorporating additional EEG modalities, such as high-density EEG or intracranial EEG, to explore their potential in improving robustness and accuracy in adversarial scenarios. These modalities could provide richer, higher-resolution data for more effective neural communication.

V. CONCLUSION

Our study highlights the substantial improvements in robustness and performance achieved by applying ANNT to B2B-C systems using SSVEP EEG data. The comprehensive analysis of various attack scenarios demonstrates ANNT's effectiveness in enhancing model resilience against adversarial attacks, contributing to developing more secure and reliable neural communication technologies. These advancements pave the way for practical applications in healthcare, education, and social interactions. Our findings, consistent with previous research [25], confirm a significant improvement in adversarial accuracy and AUC, with notable gains of 24% in accuracy and 0.23 points in AUC for the Lee2019_SSVEP dataset, and 9% in accuracy and 0.07 points in AUC for the Nakanishi2015 dataset. This reproducibility underscores ANNT's potential to enhance the security of B2B-C systems across different datasets and conditions. Additionally, our detailed examination of dataset characteristics emphasizes the importance of considering factors such as sampling rate in model training. Beyond adversarial attacks, future research should also explore other threats like jamming and eavesdropping to develop comprehensive security frameworks. In summary, our findings affirm the significant role of ANNT

in improving B2B-C systems' accuracy and robustness, encouraging further exploration of this promising area to enhance the security and efficiency of neural communication technologies.

REFERENCES

[1] Grau, C., Ginhoux, R., Riera, A., Nguyen, T. L., Chauvat, H., Berg, M., & Amengual, J. L. (2014). Conscious Brain-to-Brain Communication in Humans Using Non-Invasive Technologies. *PLOS ONE*, **9**(8), e105225. https://doi.org/10.1371/journal.pone.0105225

[2] Lebedev, M., Kunicki, C., Wang, J., & Nicolelis, M. A. (2013). A Brain-to-Brain Interface for Real-Time Sharing of Sensorimotor Information. *Scientific Reports* **2013**, *3*, 1-10. https://doi.org/10.1038/srep01319

[3] Hindley, N., Sanchez Avila, A., & Henstridge, C. (2023). Bringing synapses into focus: Recent advances in synaptic imaging and mass-spectrometry for studying synaptopathy. *Frontiers in Synaptic Neuroscience*, **15**, 1130198. https://doi.org/10.3389/fnsyn.2023.1130198

[4] Saeidi, M., Karwowski, W., Farahani, F. V., Fiok, K., Taiar, R., & Hancock, P. A. (2021). Neural Decoding of EEG Signals with Machine Learning: A Systematic Review. *Brain Sciences*, **11**(11), 1525. https://doi.org/10.3390/brainsci11111525

[5] Dattola, S., Morabito, F. C., Mammone, N., & La Foresta, F. Findings about LORETA Applied to High-Density EEG—A Review. *Electronics*, **9**(4), 660. https://doi.org/10.3390/electronics9040660

[6] Fares, A., Zhong, S. & Jiang, J. EEG-based image classification via a region-level stacked bi-directional deep learning framework. *BMC Medical Informatics and Decision Making*, **19**(Suppl 6), 268 (2019). https://doi.org/10.1186/s12911-019-0967-9

[7] Walther, D., Viehweg, J., Haueisen, J. & Mäder, P. (2023). A systematic comparison of deep learning methods for EEG time series analysis. *Frontiers in Neuroinformatics*, **17**:1067095. doi: 10.3389/fninf.2023.1067095

[8] Rong, J., Sun, R., Guo, Y., & He, B. (2023). Effects of EEG Electrode Numbers on Deep Learning-Based Source Imaging. In: Liu, F., Zhang, Y., Kuai, H., Stephen, E.P., Wang, H. (eds) *Brain Informatics. BI 2023. Lecture Notes in Computer Science()*, vol 13974. Springer, Cham. https://doi.org/10.1007/978-3-031-43075-6_11

[9] Orban, M., Elsamanty, M., Guo, K., Zhang, S., & Yang, H. (2022). A Review of Brain Activity and EEG-Based Brain–Computer Interfaces for Rehabilitation Application. *Bioengineering*, **9**(12), 768. https://doi.org/10.3390/bioengineering9120768

[10] Müller-Putz, G. R., Scherer, R., Brauneis, C., & Pfurtscheller, G. (2005). Steady-state visual evoked potential (SSVEP)-based communication: impact of harmonic frequency components. *Journal of Neural Engineering*, **2**(4), 123. https://doi.org/10.1088/1741-2560/2/4/008

[11] Zhang, Y., Xu, P., Liu, T., Hu, J., Zhang, R., & Yao, D. (2012). Multiple Frequencies Sequential Coding for SSVEP-Based Brain-Computer Interface. *PLOS ONE*, **7**(3), e29519. https://doi.org/10.1371/journal.pone.0029519

[12] Diez, P. F., Mut, V. A., Avila Perona, E. M. et al. (2011). Asynchronous BCI control using high-frequency SSVEP. *Journal of NeuroEngineering and Rehabilitation*, **8**, 39. https://doi.org/10.1186/1743-0003-8-39

[13] Sergio, L. B., et al. (2021). Security in Brain-Computer Interfaces: State of the Art, Opportunities, and Future Challenges. *ACM Computing Surveys*, **54**(1), Article 11. https://doi.org/10.1145/3427376

[14] Yoo, S., Kim, H., Filandrianos, E., Taghados, S. J., & Park, S. (2013). Non-Invasive Brain-to-Brain Interface (BBI): Establishing Functional Links between Two Brains. *PLOS ONE*, **8**(4), e60410. https://doi.org/10.1371/journal.pone.0060410

[15] Dongrui, W., et al. (2023). Adversarial attacks and defenses in physiological computing: a systematic review. *National Science Open*, **2**(1). https://doi.org/10.1360/nso/20220023

[16] Adesina, D., Hsieh, C.-C., Sagduyu, Y. E., & Qian, L. (2023). Adversarial Machine Learning in Wireless Communications Using RF Data: A Review. *IEEE Communications Surveys & Tutorials*, **25**(1), 77-100. https://doi.org/10.1109/COMST.2022.3205184

[17] Donchin, E., Spencer, K. M., & Wijesinghe, R. (2000). The mental prosthesis: assessing the speed of a P300-based brain-computer interface. *IEEE Transactions on Rehabilitation Engineering*, **8**(2), 174-179. https://doi.org/10.1109/86.847808

[18] Wolpaw, J. R., & McFarland, D. J. (2004). Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proceedings of the National Academy of Sciences*, **101**(51), 17849-17854. https://doi.org/10.1073/pnas.0403504101

[19] Buzsáki, G. (2006). Rhythms of the Brain. *Oxford University Press*. https://doi.org/10.1093/acprof:oso/9780195301069.001.0001

[20] Ahmadi, H., & Mesin, L. (2024). Enhancing Motor Imagery Electroencephalography Classification with a Correlation-Optimized Weighted Stacking Ensemble Model. *Electronics*, **13**(6), 1033. https://doi.org/10.3390/electronics13061033

[21] Ahmadi, H., & Mesin, L. (2024). Enhancing MI EEG Signal Classification with a Novel Weighted and Stacked Adaptive Integrated Ensemble Model: A Multi-Dataset Approach. *IEEE Access*, https://doi: 10.1109/ACCESS.2024.3434654.

[22] Rao, R. P., Stocco, A., Bryan, M., Sarma, D., Youngquist, T. M., Wu, J., & Prat, C. S. (2014). A Direct Brain-to-Brain Interface in Humans. *PLOS ONE*, **9**(11), e111332. https://doi.org/10.1371/journal.pone.0111332

[23] Rajesh, S., Paul, V., Menon, V. G., Jacob, S., & Vinod, P. (2020). Secure Brain-to-Brain Communication With Edge Computing for Assisting Post-Stroke Paralyzed Patients. *IEEE Internet of Things Journal*, **7**(4), 2531-2538. https://doi.org/10.1109/JIOT.2019.2951405

[24] Ajmeria, R., Sharma, N., Kalyani, N., Iyer, B., Kamal, M. K., & Pathan, M. (2023). A Critical Survey of EEG-Based BCI Systems for Applications in Industrial Internet of Things. *IEEE Communications Surveys & Tutorials*, **25**(1), 184-212. https://doi.org/10.1109/COMST.2022.3232576

[25] Ahmadi, H., Kuhestani, A., & Mesin, L. (2024). Adversarial Neural Network Training for Secure and Robust Brain-to-Brain Communication. *IEEE Access*, **12**, 39450-39469. https://doi.org/10.1109/ACCESS.2024.3376657

[26] Nakanishi, M., Wang, Y., Wang, T., & Jung, P. (2015). A Comparison Study of Canonical Correlation Analysis Based Methods for Detecting Steady-State Visual Evoked Potentials. *PLOS ONE*, **10**(10), e0140703. https://doi.org/10.1371/journal.pone.0140703

[27] Lee, M., Kwon, O., Kim, Y., Kim, H., Lee, Y., Williamson, J., Fazli, S., & Lee, S. (2019). EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy. *GigaScience*, **8**(5). https://doi.org/10.1093/gigascience/giz002

[28] Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and Harnessing Adversarial Examples. *ArXiv*, **/abs/1412.6572**. https://arxiv.org/abs/1412.6572

[29] Madry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2017). Towards Deep Learning Models Resistant to Adversarial Attacks. *ArXiv*, **/abs/1706.06083**. https://arxiv.org/abs/1706.06083

[30] Kurakin, A., Goodfellow, I., & Bengio, S. (2016). Adversarial examples in the physical world. *ArXiv*, **/abs/1607.02533**. https://arxiv.org/abs/1607.02533

[31] Carlini, N., & Wagner, D. (2017). Towards Evaluating the Robustness of Neural Networks. *2017 IEEE Symposium on Security and Privacy (SP)*, **pp. 39-57**. doi: 10.1109/SP.2017.49.

[32] Dong, Y., Liao, F., Pang, T., Su, H., Zhu, J., Hu, X., & Li, J. (2017). Boosting Adversarial Attacks with Momentum. *ArXiv*, **/abs/1710.06081**. https://arxiv.org/abs/1710.06081

HOSSEIN AHMADI received his B.S. in Electronics Engineering from Kurdistan University, Sanandaj, Iran, and his M.S. in Telecommunications Engineering from Amirkabir University of Technology, Tehran, Iran, in 2010 and 2016, respectively. He is a Ph.D. candidate at Politecnico di Torino, Italy, specializing in Electrical, Electronics, and Communications Engineering. His research focuses on Brain-to-Brain Communication, Semantic Communication, and Signal Processing.

ALI KUHESTANI (Member, IEEE) received the Ph.D. degree in electrical engineering from the Amirkabir University of Technology, Tehran, Iran, in 2017. He is an Associate Professor in the Communications and Electronics Department, Faculty of Electrical and Computer Engineering, Qom University of Technology, Iran. His research interests include physical-layer security of wireless communications, the Internet of Things, millimeter-wave communication, massive MIMO systems, and space-time coding.

MOHAMMADREZA KESHAVARZI received his Ph.D. degree in electrical engineering from the Amirkabir University of Technology, Tehran, Iran, in 2009. Since 2012, he has been with the Iran Telecommunication Research Center (ITRC) in Tehran, Iran as an assistant professor. His research interests include physical-layer security of wireless communications, Internet of Things, and MIMO systems.

LUCA MESIN received the graduate degree in electronics engineering in 1999, and the Ph.D. degree in applied mathematics from Politecnico di Torino, Italy, in 2003. He is an associate professor of biomedical engineering and a supervisor of the Mathematical Biology and Physiology Group, Department of Electronics and Telecommunications, Politecnico di Torino. His research interests include biomedical image and signal processing and mathematical modeling.

TABLE 3: Evaluation of Model Performance and Computational Time Across Different Scenarios with and without ANNT

Attack(s)	Dataset	Acc. No ANNT	Acc. ANNT	Acc. Imp.	AUC No ANNT	AUC ANNT	AUC Imp.	Comp. Time
FGSM	A	0.88	0.91	0.03	0.96	0.99	0.03	100
	B	0.65	0.80	0.15	0.66	0.94	0.32	454.68
BIM	A	0.83	0.88	0.05	0.93	0.97	0.04	454.68
	B	0.60	0.80	0.20	0.65	0.95	0.30	743.75
C&W	A	0.85	0.91	0.06	0.91	0.95	0.04	395.54
	B	0.63	0.85	0.22	0.65	0.98	0.33	628.96
MIM	A	0.82	0.89	0.07	0.92	0.97	0.05	511.79
	B	0.58	0.81	0.23	0.63	0.97	0.34	977.71
PGD	A	0.78	0.86	0.08	0.91	0.96	0.05	571.43
	B	0.55	0.79	0.24	0.57	0.93	0.36	1147.68
FGSM + C&W	A	0.84	0.92	0.08	0.91	0.98	0.07	454.68
	B	0.62	0.83	0.21	0.64	0.97	0.33	798.75
FGSM + BIM	A	0.83	0.90	0.07	0.93	0.99	0.06	511.79
	B	0.61	0.81	0.30	0.63	0.94	0.31	977.71
FGSM + MIM	A	0.83	0.91	0.08	0.92	0.98	0.06	571.43
	B	0.60	0.78	0.18	0.63	0.94	0.31	1147.68
FGSM + PGD	A	0.81	0.88	0.07	0.91	0.97	0.06	628.96
	B	0.58	0.79	0.21	0.60	0.94	0.34	1321.07
C&W + BIM	A	0.82	0.89	0.07	0.90	0.96	0.06	571.43
	B	0.60	0.82	0.22	0.63	0.96	0.33	1147.68
C&W + MIM	A	0.81	0.90	0.09	0.90	0.96	0.06	628.96
	B	0.58	0.83	0.25	0.62	0.96	0.24	1321.07
C&W + PGD	A	0.80	0.88	0.08	0.89	0.95	0.06	695.54
	B	0.57	0.81	0.24	0.59	0.96	0.37	1496.43
BIM + MIM	A	0.80	0.88	0.08	0.90	0.97	0.07	695.54
	B	0.57	0.80	0.23	0.62	0.96	0.34	1496.43
BIM + PGD	A	0.78	0.87	0.09	0.90	0.96	0.06	743.75
	B	0.55	0.79	0.24	0.59	0.94	0.35	1669.82
MIM + PGD	A	0.78	0.87	0.09	0.90	0.96	0.06	811.79
	B	0.54	0.80	0.26	0.58	0.95	0.37	1857.86
FGSM + C&W + BIM	A	0.79	0.89	0.10	0.85	0.96	0.11	628.96
	B	0.57	0.81	0.14	0.59	0.93	0.34	1321.07
FGSM + C&W + MIM	A	0.78	0.88	0.10	0.87	0.95	0.08	695.54
	B	0.56	0.81	0.25	0.58	0.94	0.36	1496.43
FGSM + C&W + PGD	A	0.76	0.87	0.12	0.87	0.93	0.06	743.75
	B	0.55	0.78	0.23	0.55	0.92	0.37	1669.82
FGSM + BIM + MIM	A	0.77	0.86	0.09	0.87	0.94	0.07	743.75
	B	0.55	0.77	0.22	0.60	0.92	0.32	1669.82
FGSM + BIM + PGD	A	0.75	0.85	0.10	0.87	0.94	0.07	811.79
	B	0.54	0.76	0.22	0.55	0.91	0.36	1857.86
FGSM + MIM + PGD	A	0.75	0.86	0.11	0.86	0.94	0.08	869.29
	B	0.53	0.76	0.23	0.55	0.92	0.37	2039.29
C&W + BIM + MIM	A	0.77	0.86	0.09	0.86	0.93	0.07	811.79
	B	0.54	0.79	0.25	0.58	0.93	0.35	1857.86
C&W + BIM + PGD	A	0.75	0.84	0.09	0.85	0.93	0.08	869.29
	B	0.53	0.78	0.25	0.55	0.92	0.37	2039.29
C&W + MIM + PGD	A	0.74	0.85	0.11	0.85	0.93	0.08	928.93
	B	0.52	0.78	0.26	0.54	0.92	0.38	2219.64
BIM + MIM + PGD	A	0.74	0.85	0.11	0.86	0.94	0.08	977.71
	B	0.51	0.77	0.26	0.55	0.91	0.36	2397.82
FGSM + C&W + BIM + MIM	A	0.74	0.85	0.11	0.83	0.92	0.09	869.29
	B	0.53	0.77	0.24	0.55	0.91	0.36	2039.29
FGSM + C&W + BIM + PGD	A	0.72	0.84	0.12	0.83	0.91	0.08	928.93
	B	0.51	0.76	0.25	0.51	0.90	0.39	2219.64
FGSM + C&W + MIM + PGD	A	0.72	0.85	0.13	0.82	0.91	0.09	977.71
	B	0.50	0.76	0.26	0.51	0.90	0.39	2397.82
FGSM + BIM + MIM + PGD	A	0.70	0.83	0.13	0.82	0.92	0.10	1041.07
	B	0.49	0.75	0.26	0.51	0.90	0.39	2569.64
C&W + BIM + MIM + PGD	A	0.71	0.83	0.12	0.82	0.91	0.09	1105.36
	B	0.46	0.76	0.30	0.51	0.88	0.37	2736.79
FGSM + C&W + BIM + MIM + PGD	A	0.63	0.79	0.16	0.73	0.87	0.14	1147.68
	B	0.41	0.71	0.30	0.42	0.85	0.43	2880.36

A = Nakanishi2015, **B** = Lee2019_SSVEP, **Acc. No ANNT** = Accuracy without ANNT, **Acc. ANNT** = Accuracy with ANNT, **Acc. Imp.** = Accuracy Improvement, **AUC No ANNT** = AUC without ANNT, **AUC ANNT** = AUC with ANNT, **AUC Imp.** = AUC Improvement, **Comp. Time** = Computational Time.

TABLE 4: Metrics and Attack Scenarios for Both Datasets

Dataset	Metric	Scenario	Least Effective Attack	Most Effective Attack
Nakanishi2015	Accuracy with ANNT	1 Attack	FGSM (0.91)	PGD (0.86)
		2 Attacks	FGSM + C&W (0.92)	BIM + PGD (0.87)
		3 Attacks	FGSM + C&W + BIM (0.89)	C&W + BIM + PGD (0.84)
		4 Attacks	FGSM + C&W + BIM + MIM (0.85)	FGSM + BIM + MIM + PGD (0.83)
	AUC with ANNT	1 Attack	FGSM (0.99)	C&W (0.95)
		2 Attacks	FGSM + BIM (0.99)	C&W + PGD (0.95)
		3 Attacks	FGSM + C&W + BIM (0.96)	FGSM + C&W + PGD (0.93)
		4 Attacks	FGSM + BIM + MIM + PGD (0.92)	FGSM + C&W + BIM + MIM (0.91)
	Accuracy Improvement	1 Attack	FGSM (0.03)	PGD (0.08)
		2 Attacks	FGSM + PGD (0.08)	C&W + MIM (0.09)
		3 Attacks	FGSM + C&W + MIM (0.10)	FGSM + C&W + PGD (0.11)
		4 Attacks	FGSM + C&W + BIM + MIM (0.11)	FGSM + C&W + MIM + PGD (0.13)
	AUC Improvement	1 Attack	FGSM (0.03)	MIM (0.05)
		2 Attacks	FGSM + BIM (0.06)	FGSM + C&W (0.07)
		3 Attacks	FGSM + C&W + BIM (0.07)	FGSM + C&W + PGD (0.11)
		4 Attacks	FGSM + C&W + BIM + MIM (0.09)	FGSM + BIM + MIM + PGD (0.10)
	Computational Time	1 Attack	FGSM (100)	PGD (571.43)
		2 Attacks	FGSM + C&W (454.68)	MIM + PGD (811.79)
		3 Attacks	FGSM + C&W + BIM (628.96)	BIM + MIM + PGD (977.71)
		4 Attacks	FGSM + C&W + BIM + MIM (869.29)	C&W + BIM + MIM + PGD (1105.36)
Lee2019_SSVEP	Accuracy with ANNT	1 Attack	C&W (0.85)	PGD (0.79)
		2 Attacks	FGSM + C&W (0.83)	FGSM + MIM (0.78)
		3 Attacks	FGSM + C&W + BIM (0.81)	FGSM + BIM + PGD (0.76)
		4 Attacks	FGSM + C&W + BIM + MIM (0.78)	FGSM + BIM + MIM + PGD (0.75)
	AUC with ANNT	1 Attack	C&W (0.98)	PGD (0.93)
		2 Attacks	FGSM + C&W (0.97)	FGSM + BIM (0.94)
		3 Attacks	FGSM + C&W + MIM (0.94)	FGSM + BIM + PGD (0.91)
		4 Attacks	FGSM + C&W + BIM + MIM (0.94)	C&W + BIM + MIM + PGD (0.88)
	Accuracy Improvement	1 Attack	FGSM (0.15)	PGD (0.24)
		2 Attacks	FGSM + C&W (0.21)	C&W + MIM (0.25)
		3 Attacks	FGSM + C&W + MIM (0.25)	C&W + BIM + PGD (0.30)
		4 Attacks	FGSM + C&W + BIM + MIM (0.30)	C&W + BIM + MIM + PGD (0.43)
	AUC Improvement	1 Attack	FGSM (0.28)	PGD (0.36)
		2 Attacks	FGSM + C&W (0.33)	FGSM + BIM (0.39)
		3 Attacks	FGSM + C&W + MIM (0.35)	FGSM + BIM + PGD (0.43)
		4 Attacks	FGSM + C&W + BIM + MIM (0.36)	C&W + BIM + MIM + PGD (0.45)
	Computational Time	1 Attack	FGSM (454.68)	PGD (1147.68)
		2 Attacks	FGSM + C&W (798.75)	MIM + PGD (1857.86)
		3 Attacks	FGSM + C&W + BIM (1321.07)	BIM + MIM + PGD (2397.82)
		4 Attacks	FGSM + C&W + BIM + MIM (2039.29)	C&W + BIM + MIM + PGD (2736.79)