Adversarial Neural Network Training for Secure and Robust Brain-to-Brain Communication

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 26 February 2024, accepted 9 March 2024, date of publication 13 March 2024, date of current version 19 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3376657

RESEARCH ARTICLE

Adversarial Neural Network Training for Secure and Robust Brain-to-Brain Communication

HOSSEIN AHMADI[1], ALI KUHESTANI[2], (Member, IEEE), AND LUCA MESIN[1]
[1]Mathematical Biology and Physiology, Department of Electronics and Telecommunications, Politecnico di Torino, 10129 Turin, Italy
[2]Communications and Electronics Department, Faculty of Electrical and Computer Engineering, Qom University of Technology, Qom 3718146645, Iran

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

ABSTRACT In the rapidly evolving domain of brain-to-brain communication, safeguarding the transmission of information against adversarial threats is paramount. This study introduces an advanced approach to enhance the resilience and security of brain-to-brain communication systems utilizing electroencephalogram data against such threats through adversarial neural network training. Concentrating on event-related potentials and employing a diverse collection of eight datasets, our research rigorously evaluates and optimizes the system's defense mechanisms against adversarial manipulations. We specifically target the optimization of trial durations and sampling rates to bolster system security. Our findings reveal a marked improvement in the system's defensive capabilities, demonstrated by a significant increase in adversarial accuracy by 17% and enhancement in the area under the receiver operating characteristic curve by 0.12 points. These results underscore the efficacy of our approach in fortifying brain-to-brain communication systems against sophisticated cyber threats, marking a significant step forward in the secure and robust transmission of neural signals.

INDEX TERMS Adversarial accuracy, adversarial attacks, adversarial neural network training, brain-to-brain communication, electroencephalogram, event-related potentials, neuro-engineering, security enhancement.

I. INTRODUCTION

Brain-to-Brain Communication (B2B-C) represents a paradigm shift in neuroscience and neuro-engineering, with the potential to redefine our understanding of cognitive processes and interpersonal communication [1]. While various methods exist for facilitating B2B-C, this study focuses on Electroencephalogram (EEG)–based systems due to their unique noise sensitivity and security challenges. However, the sensitivity of EEG signals to noise and distortion presents significant challenges, especially when considering the security implications of such a communication [2].

Machine learning (ML), particularly neural networks and Deep Learning (DL), has shown promise in handling the complexities of high-dimensional, noisy EEG data [3]. Yet, adversarial attacks, characterized by subtle perturbations, can lead traditional machine learning models astray [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Ghulam Muhammad.

The advent of adversarial attacks presents a formidable challenge to the security of B2B-C systems, emphasizing the need for robust defense mechanisms. As highlighted in [5], adversarial examples can significantly compromise the integrity of EEG-based Brain-Computer Interfaces (BCIs), undermining their reliability and safety. The comprehensive review by [6] highlights significant security challenges facing physiological computing systems, including B2B-C networks, underscoring the urgent need for robust adversarial training and defense mechanisms. The burgeoning field of adversarial machine learning offers promising avenues for securing complex communication systems against sophisticated cyber threats. As explored in [7], implementing adversarial ML techniques within the consumer Internet of Things (IoT) landscape, particularly in smart healthcare, underscores the effectiveness of these methods in enhancing security measures. Recent advancements in adversarial ML have illuminated the susceptibility of EEG-based BCIs to adversarial attacks, a vulnerability that extends to the broader

domain of B2B-C systems. The work of Jung et al. [8] on Generative Perturbation Networks (GPN) underscores the critical necessity for incorporating adversarial robustness into developing these systems.

Wireless communication, integral to the practical implementation of B2B-C, introduces another layer of complexity. The transmission of EEG data over wireless channels can be susceptible to noise, interference, and potential security breaches, emphasizing the need for robust and secure communication methodologies [9].

While there is extensive literature on EEG analysis and wireless communication individually, few works explore the intersection of the two, and the domain of B2B-C remains in its infancy. A conspicuous gap emerges when considering the intersection of Adversarial Neural Networks Training (ANNT), EEG analysis, wireless communication, and B2B-C. This paper ventures into this relatively uncharted territory, aiming to fortify B2B-C systems against adversarial threats using ANNT. Table 1 underscores this observation, highlighting that while some exploration of B2B-C exists, its convergence with ANNT and wireless communication remains untouched.

For instance, works by Grau et al. [1] and Rao et al. [10] have laid foundational concepts in B2B-C but did not delve into the security aspects. On the other hand, studies like those by Rajesh et al. [11] and Ajmeria et al. [12] have emphasized the need for security but have not ventured into the application of ANNT, particularly in conjunction with EEG, for fortifying B2B-C systems.

Interestingly, the work of Brocal et al. [13] stands as the only example in the surveyed literature where ANNT is utilized. However, their study diverges significantly from our work. In their approach, ANNT is deployed primarily to enhance a neurohaptic interface through Generative Adversarial Networks (GANs), aiming to create tangible patterns for transmitting emotions and thoughts.

By preparing our system to resist adversarial threats, we aim to ensure the integrity and reliability of communications across varying environments. This distinct application of adversarial training highlights the versatility and expansive potential of ANNT in neural network research.

In this study, we significantly advance the security and robustness of B2B-C systems through the innovative application of ANNT. Our contributions are twofold: First, we demonstrate the critical role of Sampling Rate over Trial Duration in enhancing ANNT's effectiveness, offering a novel insight into optimizing B2B-C systems against adversarial threats. Second, we propose a framework for future research directions, emphasizing the need to explore a wider range of adversarial scenarios and their implications for neural signal processing and cognitive tasks. The remainder of this paper is organized as follows: Section II details our methodology and experimental design, Section III presents our results, Section IV discusses the implications of our findings, and we conclude in Section V with a summary and future research directions.

II. METHODOLOGY AND EXPERIMENT DESIGN

A. PROBLEM DEFINITION

This study uses EEG data capturing Event-Related Potential (ERP), specifically focusing on the P300 paradigm. The primary goal is to classify EEG samples into Non-target events (class 1) and Target events (class 2). The P300 signal is particularly pivotal in cognitive neuroscience due to its pronounced neural response, occurring approximately 300 milliseconds after stimulus onset [14]. This response is robust and reliable across different subjects and uniquely representative of cognitive processes like attention and decision-making, which renders it highly suitable for B2B-C [15].

The choice of the P300 paradigm over other EEG signals stems from several key advantages. Firstly, the P300 signal's distinctiveness lies in its clear and measurable response, which is less susceptible to variability and noise than other EEG components [16]. This makes it an ideal candidate for accurate and efficient B2B-C systems. Secondly, its association with communication-related cognitive processes - such as target recognition and differentiation - makes it inherently suited for transmitting meaningful information in a B2B-C setup [17]. Finally, the practical aspects of using P300, including its non-invasiveness and minimal requirement for subject training, further add to its feasibility and applicability in real-world B2B-C scenarios [46].

Our methodology comprises four main components.

- **Utilising a Convolutional Neural Network (CNN).** Our approach employs CNNs to extract spatial features from multi-channel EEG data, which is pivotal in analyzing ERP signals. In the P300 paradigm, the differentiation between Target and Non-target events is significantly based on the spatial distribution of brain activity. Target events typically elicit distinct spatial patterns in EEG signals, characterized by notable activations in specific brain regions. CNNs are particularly adept at capturing these unique spatial signatures due to their ability to process and learn from multi-dimensional data [18].
- **Employing Temporal Convolutional Networks (TCN).** To capture the temporal dependencies in EEG signals, we utilize TCNs. These are preferred over traditional Long Short-Term Memory (LSTM) models owing to their parallel processing capabilities and stable gradient flow, both crucial for managing the time-sensitive nature of ERP data [19]. The P300 component's identification heavily relies on its timing post-stimulus. TCNs, with their efficient handling of sequential data, can precisely track and learn these time-dependent patterns, ensuring a more accurate and temporally coherent analysis of the P300 signals. Additionally, TCNs reduce the likelihood of overfitting and speed up the training process, making our model more efficient and robust in real-time B2B-C applications.

TABLE 1. Summary of reviewed references.

Ref	Year	Main Contributions	Security	ANNT
Lebedev et al. [25]	2013	Real-time BTBI[1] in animals using ICMS[2]; exchange, processing, storage of information	-	-
Yoo et al. [29]	2013	Non-invasive BBI[3] using FUS[4] and BCI[5]; human-to-rat communication; potential applications	-	-
Yu et al., [21]	2014	Human brain signal-controlled rat robot for complex navigational tasks	-	-
Grau et al., [1]	2014	Demonstration of non-invasive communication between human brains; a combination of BCIs and CBIs[6]; proof-of-principle for future technologies	-	-
Rao et al., [10]	2014	Development of direct brain-to-brain interface (see note 2); successful experiment with computer game; evidence for non-invasive communication	-	-
Chiuffa et al., [28]	2015	Development of Brainet with interconnected rat brains; solving computational problems; exploration of organic computers	-	-
RaviKumar et al., [20]	2015	Non-invasive brain-to-brain communication via EEG and TMS[7]	-	-
Toppi et al., [22]	2015	Application of graph theory to brain-to-brain connectivity analysis via EEG hyperscanning.	-	-
Micek et al., [26]	2016	Novel architecture for B2B[8] communication using brain synchronization and EOG[9] signals	-	-
Dingemanse, [27]	2017	Conceptual exploration of brain-to-brain interfaces (see note 2); comparison to language; progression from BMIs[10]	-	-
Jiang et al., [34]	2019	Development of BrainNet interface; successful collaborative experiment; insights into trust and reliability	-	-
Lu et al., [33]	2020	Development of optical B2BI; demonstration of precise control; insight into neural information transmission	-	-
Moioli et al., [9]	2021	An interdisciplinary perspective on neurosciences and wireless communications; practical applications like brain-controlled vehicles	-	-
Willett et al., [30]	2021	Development of BCI for imagined handwriting; achievement of high typing speeds; opening new communication possibilities	-	-
Nam et al., [31]	2021	Systematic review of B2BI technology; insight into collaboration models; emphasis on potential in neuro ergonomics	-	-
Hekmatmanesh et al., [32]	2022	Development of BCI algorithm for vehicle control; incorporation of boosting technique; comparison of different algorithms	-	-
Ma et al., [24]	2022	EBCM[9] paradigm for brain-controlled communication; its capabilities; and explores its potential AI[10] integration for advanced systems	-	-
Wang et al., [13]	2020	Neurohaptic interface using EEG and GAN for tangible brain-to-brain emotion communication	-	✓
Hameed et al., [23]	2020	EEG-based alphabet classification for silent thought communication	-	-
Rajesh et al., [11]	2020	Development of secure BBI with encryption; assistance for post-stroke paralyzed patients; successful testing	✓	-
Sergio et al., [2]	2021	Overview of security in BCIs; identification of security threats; discussion of potential solutions; insights into future directions	✓	-
Ajrawi et al., [35]	2021	Emphasis on cybersecurity in BCIs; design-theoretical framework using RFID[11]; introduction of BCIIS[12]	✓	-
Tarkhani et al., [36]	2022	Detailed analysis of BCI vulnerabilities; introduction and evaluation of Argus; emphasis on security measures	✓	-
Ajmeria et al., [12]	2023	Critical survey of EEG-based BCI; identification of use cases and challenges; deployment study; emphasis on human intuition and reliability	✓	-
Brocal et al., [37]	2023	Exploration of BCI risks; a framework for BCI safety and security; emphasis on challenges and considerations	✓	-
Hossain et al., [3]	2023	Comprehensive review of deep learning in EEG-based BCI; guidance for future studies	-	-
This Study	**2024**	**Comprehensive exploration of ANNT in B2B-C; robust algorithm development; security and reliability insights.**	✓	✓

[1]**BTBI**: Brain-to-Brain Interface
[2]**ICMS**: Intracortical Microstimulation
[3]**BBI**: Brain-to-Brain Interface
[4]**FUS**: Focused Ultrasound
[5]**BCI**: Brain Computer Interface
[6]**CBI**: Computer Brain Interface
[7]**TMS**: Transcranial Magnetic Stimulation
[8]**B2B**: Brain-to-Brain
[9]**EOG**: Electrooculography
[10]**BMI**: Brain Machine Interface
[11]**EBCM**: Electromagnetic Brain-Computer-Metasurface
[11]**AI**: Artificial Intelligence
[12]**RFID**: Radio Frequency Identification
[13]**BCIIS**: BCI Identification System
Note 1: The acronyms used in this table are as they appear in the original references cited. They are retained here for fidelity to the source material and to maintain the context in which they were originally used.
Note 2: The term "brain-to-brain interfaces" is spelled out to avoid confusion with the acronym "BBI," which is used in a different context in the cited work by Yoo et al. [29].

- **Simulating Adversarial Attack on the Wireless Channel.** We simulate an adversarial attack using the Fast Gradient Sign Method (FGSM) to test the robustness of our model against potential security threats. FGSM, known for its efficiency in generating adversarial examples, helps us understand how subtle perturbations can mislead the model.
- **Applying ANNT.** To fortify our model against adversarial threats, we integrate ANNT. This approach involves training the model on both original and adversarially perturbed data. By doing so, the model learns to recognize and resist adversarial patterns, enhancing its ability to maintain high accuracy and reliability, even in adversarial perturbations.

B. SYSTEM MODEL

Figure 1 delineates the architecture of our CNN-TCN model, charting the journey from EEG data acquisition to adversarial resilience training. At the outset, our model harnesses multi-dimensional EEG data, capturing the nuanced electrical activities of the brain. This data undergoes meticulous preprocessing, a step crucial for isolating the P300 signal—a marker of cognitive acuity and focus.

Our architecture's ingenuity lies in its dual-layered approach to feature extraction. The CNN layer adeptly isolates spatial features, those distinct neural fingerprints indicative of target or non-target responses. Simultaneously, the TCN layer tracks the temporal evolution of the EEG signals, capturing the precise timing of the P300 wave's emergence post-stimulus.

To assess the robustness of our model against digital adversaries, we inject the system with adversarially modified data via the FGSM technique, thereby mimicking potential security breaches. The incorporation of ANNT is a strategic counter, training the model to recognize and resist these perturbations, thereby bolstering its defense mechanisms.

Evaluative measures are then rigorously applied across three distinct scenarios: unaltered data classification, adversarially attacked data classification, and attacked data classification post-ANNT integration. These scenarios test the model's accuracy and resilience, ensuring the integrity of B2B-C in the face of adversarial onslaughts.

C. DATA COLLECTION AND PREPROCESSING

We used various ERP EEG datasets, each with varying characteristics regarding the number of subjects, channels, trials per class, trial durations, sampling rates, and sessions. The datasets employed are summarised in Table 2.

The raw EEG signals from these datasets underwent preprocessing, including applying a band-pass filter within the 0.1-30 Hz range to effectively isolate the ERP signals' relevant features, especially the P300 component. Following filtering, the EEG signals were segmented into epochs, with each epoch's duration precisely matching the trial durations listed for each dataset in Table 2. This alignment ensures comprehensive capture of the ERP responses, particularly the P300 component, across all datasets. The epoch lengths thus directly correspond to the trials' durations, ranging from 0.8 seconds to 1.2 seconds, depending on the dataset specifics. This methodological choice underpins our data's consistency and accuracy in capturing the essential ERP features for analysis. In parallel, we utilized a K-fold cross-validation strategy for data splitting in our experiments, selecting K=10 as the optimal balance between computational efficiency and model accuracy after testing ranges from 5 to 15 folds. This approach ensures rigorous evaluation and utilization of each data point for training and validation, minimizing the risk of overfitting while maintaining computational manageability.

D. MATHEMATICAL FORMULATION

The methods employed in this study are mathematically represented, elucidating the interplay between neural networks, adversarial perturbation, training, and model evaluation.

- **Transmitter (Clean Data).** The clean EEG data is represented as a 3D tensor:

$$X \in \mathbb{R}^{n_{\text{trials}} \times n_{\text{channels}} \times n_{\text{time points}}} \tag{1}$$

where n_{trials} represents number of ERP trials, n_{channels} the EEG channels, and $n_{\text{time points}}$ the time samples in each trial. The corresponding labels are represented as a vector $y \in \{1, 2\}^{n_{\text{trials}}}$, with 1 and 2 denoting the two classes in our binary classification problem (Target and Non-target, respectively).

- **Wireless Channel (Adversarial Perturbation).** The EEG data, segmented into epochs, is subjected to adversarial perturbations using the FGSM, resulting in adversarial noise N, to simulate an attacked wireless channel, reflecting realistic adversarial conditions in EEG data transmission. The attacked data X' is represented as:

$$X' = X + N \tag{2}$$

where

$$N = \varepsilon \cdot \text{sign}(\nabla_X J(\theta, X, y)) \tag{3}$$

ε being a small constant set to 0.05 in our implementation of FGSM, ∇_X represents the gradient with respect to X, θ indicates the current parameters of the model used in the generation of adversarial noise, and J is the categorical cross-entropy loss function chosen for its effectiveness in handling classification tasks with one-hot encoded labels. Categorical cross-entropy loss measures the discrepancy between the predicted probabilities and the actual class labels. It is particularly suitable for our ERP signal classification task, where accurate probability estimation for each class (Target and Non-Target) is crucial.

While our experiment specifically employs the FGSM to generate an adversarial attack, the represented mathematical formulation of N in Equation (3) can

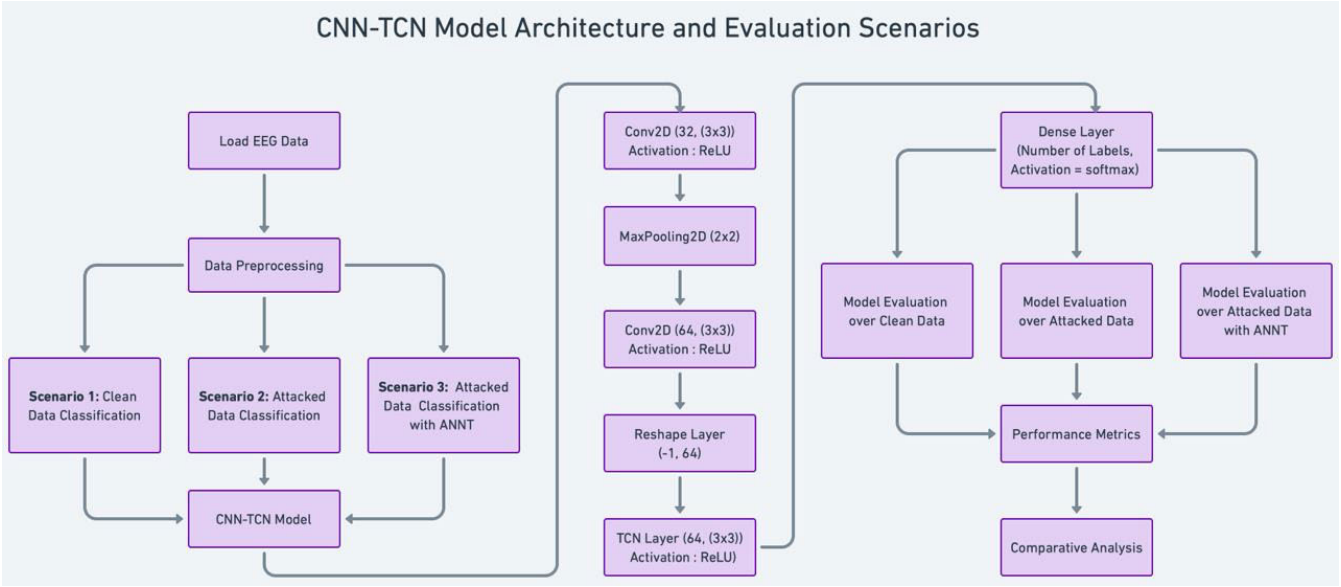

FIGURE 1. CNN-TCN Model Architecture and Evaluation Scenarios in our B2B-C system. This schematic illustrates the data flow from EEG signal acquisition through preprocessing, feature extraction with CNN and TCN, and classification, highlighting the robustness afforded by ANNT.

TABLE 2. Overview of the datasets employed in this study.

Name	Reference	#Subject	#Channel	#Trials/class NT: Non-Target; T: Target	Trials Duration	Sampling Rate	#Sessions
BI2012	[38]	25	16	640 NT / 128 T	1 s	128 Hz	2
BNCI2014_008	[45], [46]	8	8	3500 NT / 700 T	1 s	256 Hz	1
BNCI2014_009	[47]	10	16	1440 NT / 288 T	0.8 s	256 Hz	3
BI2013a	[39]–[41]	24	16	3200 NT / 640 T	1 s	512 Hz	(1-7)8 | (8-24)1
BI2014b	[42]	38	32	5 NT / 1 T	1 s	512 Hz	3
BI2015a	[43]	43	32	5 NT / 1 T	1 s	512 Hz	3
BI2015b	[44]	44	32	5 NT / 1 T	1 s	512 Hz	1
Sosulski2019	[48]–[50]	13	31	75 NT / 15 T	1.2 s	1000 Hz	3

conceptually be extended to encompass a variety of noises and perturbations, including different adversarial attack strategies. This extension is grounded in the general concept that adversarial attacks, regardless of their specific type, introduce perturbations to the input data intended to mislead the model. Thus, while equation (3) is derived from the FGSM approach, it can be seen as a broader representation of adversarial perturbations. The mathematical representation of our adversarial approach culminates in creating perturbed data that challenges the model's robustness. Figure 2 provides a comparative analysis of EEG signals before and after applying an FGSM attack to demonstrate the effect of such perturbations visually.

Upon examining Figure 2, one might be deceived by the seemingly identical nature of the clean and perturbed EEG signals. Although visually minimal and represented by a power value substantially lower than the original signal, the perturbation seriously threatens the classification performance. This discrepancy highlights the attack's potency: a minute alteration in the input can lead to a disproportionate degradation in the model's performance.

- **ANNT (Robustness Improvement).** The ANNT model is trained also on the perturbed data X', where X' has been modified with adversarial noise N. This training process can be mathematically framed as an optimization problem:

$$\theta^* = \arg\min_\theta E[J(\theta, X', y)] \tag{4}$$

where $E[\cdot]$ denotes the expectation (i.e., average over the training samples), and θ^* represents the optimized vector parameters of the model after training.

- **Receiver (Model Evaluation).** The model's performance is evaluated using both clean test data X_{test} and adversarially attacked data X'_{test}, with corresponding labels y_{test}. The evaluation metrics include Accuracy and Area Under the Curve (AUC). The Accuracy is calculated as:

$$\text{Accuracy} = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} I(y_{\text{test}}[i] = y_{\text{pred}}[i]) \tag{5}$$

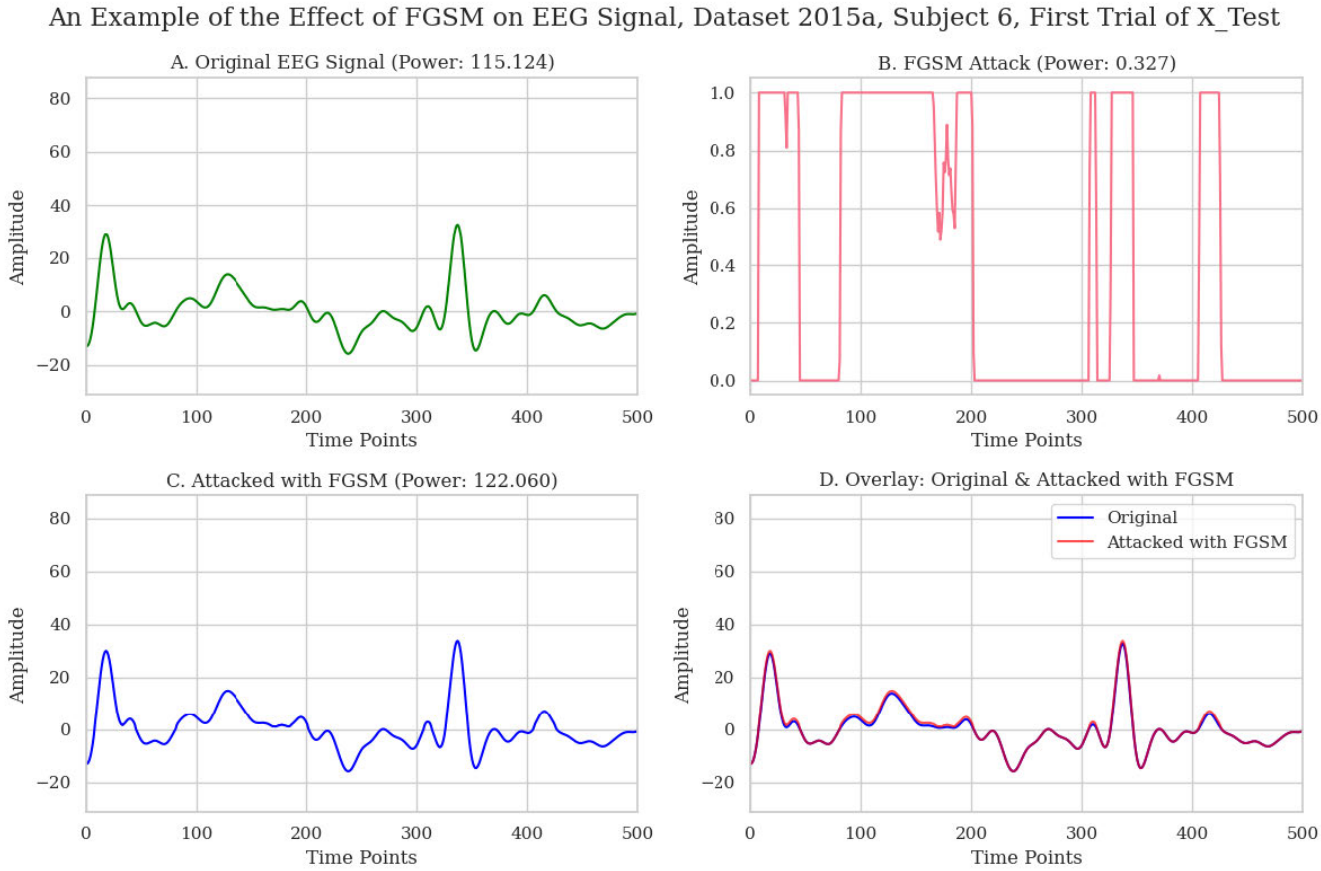

FIGURE 2. Example of the effect of the FGSM adversarial attack on EEG signals. A) The original EEG signal was without any perturbation. B) Adversarial perturbations generated by the FGSM attack. C) EEG signal after being subjected to the FGSM attack. D) Comparison of the clean original EEG and the signal post-attack.

where n_{test} is the number of test samples, y_{pred} is the vector of the model's predictions on the test data, and $I(.)$ is the indicator function, which equals 1 if $y_{\text{test}}[i] = y_{\text{pred}}[i]$ and 0 otherwise. Additionally, the Receiver Operating Characteristic (ROC) curves are plotted for clean and attacked scenarios to understand the model's performance across different threshold settings comprehensively.

E. WIRELESS CHANNEL SIMULATION AND EXPERIMENTAL DESIGN

This subsection outlines the simulation of a wireless channel for a B2B-C system and the experimental design, focusing on using adversarial attacks to evaluate and enhance model robustness.

Model Robustness. A key objective is to assess the robustness of our CNN-TCN model against adversarial attacks. These attacks are simulated to reflect potential real-world threats to the B2B-C system, testing the model's robustness under adversarial conditions.

Realistic Assessment. The simulation aims to understand the impact of adversarial attacks on B2B-C systems in real-world scenarios. We focus on the implications these attacks could have on the integrity and performance of such systems.

The experimental structure addresses these objectives through the following components:

- **Baseline Experiment.** This experiment is a control by evaluating the CNN-TCN model's performance on clean, unperturbed EEG data. It establishes a baseline for comparison with adversarially challenged scenarios.
- **Adversarial Attack Simulation.** We simulate adversarial conditions using the FGSM to create perturbed EEG data. This approach mimics potential adversarial attacks the system might encounter, allowing us to assess how the CNN-TCN model copes with such perturbations.
- **ANNT Experiment.** The model undergoes ANNT using clean and FGSM-perturbed EEG data. This process aims to improve the model's ability to withstand adversarial attacks, enhancing the security and robustness of the B2B-C system.

The experiments utilize the previously described data split, training on designated sets, and evaluation on testing sets. The adversarial training involves multiple iterations, refining the model's performance against adversarial examples generated by FGSM until satisfactory robustness is achieved.

Upon delving into our approach's theoretical and mathematical underpinnings, Figure 3 illustrates a comprehensive workflow mapping the experimental journey from EEG data acquisition to the nuanced application of ANNT. This figure captures the intricate process beginning with EEG data collection from 'Brain 1', meticulously transitioning through stages of systematic preprocessing to optimize the data for analysis. A significant emphasis is placed on introducing adversarial perturbations via the FGSM attack, a method chosen for its effectiveness in challenging the model's resilience, thereby evaluating its robustness against potential threats. This is followed by the dual-path processing of data through CNN and TCN networks, which are instrumental in feature extraction, laying the groundwork for the robustness enhancement provided by ANNT.

Each phase of the workflow is delineated by distinctive color codes, illustrating the transition from preprocessing (highlighted in purple) to adversarial example generation and evaluation (in red) and from training with clean data (in yellow) to the comprehensive evaluation of the model performance with ANNT (in green). This visual demarcation aids in understanding the workflow's complexity and the strategic interplay of various components aimed at securing a robust B2B-C framework. Furthermore, using dashed and straight lines distinguishes between the flow of evaluation and training labels and the progression of data processing steps.

An extensive evaluation phase quantitatively assesses the model's performance through accuracy, ROC, and AUC metrics, facilitating a direct comparison across clean and adversarially attacked scenarios. This meticulous evaluation underscores the effectiveness of our workflow in enhancing security and reliability in the B2B-C systems, acting not only as a procedural guide for replication but also highlighting the critical interplay among the various stages in fostering a secure framework.

F. ADVERSARIAL DATA INTEGRATION AND BIAS MITIGATION IN ANNT

Addressing the critical aspects of model evaluation, our strategic approach distributes the attacked dataset between the training and evaluation phases while mitigating bias towards specific adversarial attacks. Our methodology recognizes the importance of robust and generalizable models in B2B-C systems and ensures balanced exposure to clean and adversarially perturbed data. Through a meticulously designed alternating training regime, the model encounters various adversarial examples generated using the FGSM with a predetermined epsilon value. This enhances the model's resilience to adversarial perturbations, preventing overfitting clean data or developing a bias towards specific adversarial attacks. Incorporating adversarial examples in the training and testing phases allows for a rigorous assessment of the model's performance under realistic adversarial conditions. This ensures its effectiveness and security in real-world applications, underlining our commitment to advancing the security and reliability of B2B-C systems and addressing potential vulnerabilities in the face of adversarial threats.

G. STATISTICAL ANALYSIS OF MODEL PERFORMANCE

Our study used advanced statistical methods to evaluate the ANNT model's performance rigorously. A key component of our analysis was an Analysis of Variance (ANOVA), which aimed to assess the influence of several factors, including 'Condition', on model performance metrics such as Accuracy and AUC.

- **ANOVA Analysis.** The ANOVA, with Accuracy and AUC as dependent variables, examined factors such as 'Condition,' 'Trials Duration (s),' and 'Sampling Rate (Hz),' along with their interaction effects. The 'Condition' factor encapsulates different scenarios under which the data was analyzed, including 'Accuracy w/o ANNT (Clean),' 'Accuracy w/ ANNT (Clean),' 'Accuracy w/o ANNT (Attacked),' and 'Accuracy w/ ANNT (Attacked).' This analysis aimed to quantify each factor's impact on model performance.
- **Residual Analysis.** Residuals from the models were evaluated through Q-Q plots to verify the normality assumption, a critical aspect of ANOVA.
- **Variance Homogeneity.** To ensure the reliability of our ANOVA results, a Breusch-Pagan test was conducted to confirm variance homogeneity.

H. GENERALISING MODEL ROBUSTNESS THROUGH DIVERSE ADVERSARIAL ATTACKS

Following the initial evaluation phase, we enhanced our model's generalizability by subjecting it to three additional types of attacks. This approach affirms the model's security and robustness, ensuring its resilience across diverse scenarios. Given the multitude of potential attacks, our selection process prioritized those compatible with our unique context: the ERP EEG signal analysis in a simulated wireless B2B-C setup.

For instance, the Carlini & Wagner (CW) attack, despite its sophistication in minimizing detectable perturbations, may not be ideally suited due to its computational demand and lesser relevance to the specific robustness requirements of ERP EEG signals. Similarly, Universal Adversarial Perturbations (UAP), while revealing systemic weaknesses, could overlook the intricacies of B2B-C signals.

Thus, we meticulously selected attacks that align with both our data characteristics and model requirements:

- **Projected Gradient Descent (PGD):** This iterative enhancement of the FGSM introduces small, strategic perturbations over multiple steps. It offers a comprehensive assessment by fine-tuning the perturbations, making it particularly effective for mimicking real-world disruptions in B2B-C.
- **DeepFool:** By iteratively identifying the minimal perturbation needed to alter the model's prediction, DeepFool provides precise insights into model robustness. Its emphasis on minimal interference is congruent with the

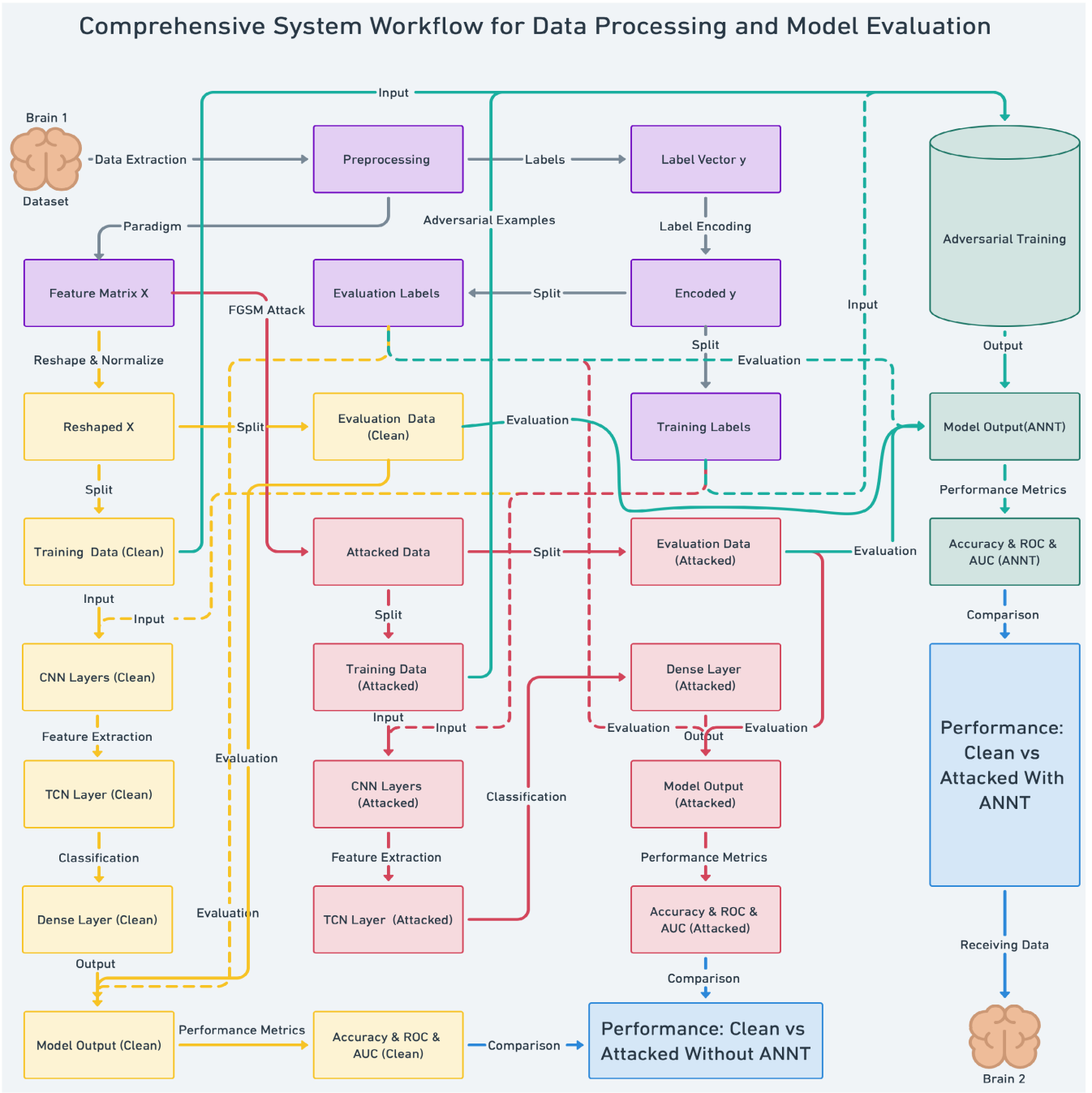

FIGURE 3. **Comprehensive Workflow of Data Processing and Model Evaluation: systematic progression from EEG data acquisition to model assessment. The workflow includes preprocessing (purple), the generation of adversarial examples and their evaluation (red), training with clean data (yellow), and the evaluation of model performance with ANNT (green). Each color highlights a specific phase in the process, detailing steps such as adversarial perturbation, feature extraction, classification, and the robustness conferred by ANNT. Dashed lines represent the flow of evaluation and training labels, while straight lines indicate the progression of data processing steps. The workflow compares clean vs attacked performance metrics with and without ANNT (blue).**

subtle nature of ERP EEG signal variations, making it highly relevant for our analysis.

- **Jacobian Saliency Map Attack (JSMA):** Focusing on the input's most impactful features for misclassification, JSMA tailors adversarial examples by altering specific features. While its precision is valuable for pinpointing critical signal components, the method's computational intensity could pose challenges for high-dimensional EEG data analysis.

These selected attacks are tailored to evaluate and enhance our model's resilience effectively, considering the specific nuances of our data and the simulated B2B-C environment.

In testing our model against these new attacks, we have broadened our evaluation metrics to include accuracy and AUC, which were used for assessing the model's performance against FGSM attacks, and precision, F1 Score, and Kappa metrics. This expanded set of metrics provides a more nuanced and generalized understanding of our findings, ensuring a comprehensive evaluation of the model's robustness across different adversarial scenarios.

III. RESULTS

This section delineates the empirical findings of our experiments, which focus on assessing the efficacy of ANNT on different EEG datasets under clean and adversarial conditions.

The model's performance was first evaluated across various datasets. As illustrated in Figure 4, the results provide a comparative analysis of accuracy and AUC metrics on clean and attacked data, with and without ANNT, indicating the general enhancement of model robustness by ANNT application.

The key observations and their interpretations are as follows. The datasets exhibited varied responses to adversarial attacks, with ANNT generally leading to an improvement in the robustness of the model as indicated by both accuracy and AUC values. Notably, dataset BI2013a demonstrated the most significant improvement in adversarial accuracy with the application of ANNT, while BNCI2014_009 showed the most substantial enhancement in AUC. These variations suggest that certain characteristics intrinsic to each dataset may differentially influence the effectiveness of ANNT (as detailed in the Discussion section).

While ROC curves were generated for all datasets to assess the model's discriminative ability, Figure 5 selectively showcases the curves for four datasets that highlight the differential impact of ANNT. The top row with ROC curves for BNCI2014_009 and BI2015a demonstrates ANNT's significant positive effect, manifesting in notable AUC improvements and suggesting enhanced defense capabilities in B2B-C systems. In contrast, the bottom row with BI2012 and BNCI2014_008 shows less pronounced improvements, indicating that the efficacy of ANNT may be contingent on specific dataset characteristics. This disparity emphasizes the need for tailored adversarial training approaches to optimize the robustness of B2B-C systems against adversarial threats.

To further dissect the influence of individual dataset characteristics on model performance, we conducted a normalized association analysis between performance metrics and dataset attributes, as depicted in Figure 6. This analysis revealed that Trial Duration and Sampling Rate are particularly impactful characteristics. To elucidate their specific effects, we performed a scatter analysis for both Trial Duration and Sampling Rate under various conditions, demonstrated in Figure 7. This granular view decisively illustrates how ANNT influences accuracy and AUC in the presence of both clean and adversarial data, underscoring the nuanced interplay between dataset features and model robustness.

The visual analyses presented in the preceding figures suggest a tangible influence of Trial Duration and Sampling Rate on model performance. To substantiate these observations with statistical evidence, we conducted an ANOVA, presented in Table 3. This statistical test quantifies the contribution of these dataset characteristics to the variance in performance metrics. The F-statistic and p-values, detailed in the table, provide a preliminary indication of the factors' impact.

Following the ANOVA, we conducted the Breusch-Pagan test to evaluate the residuals' homoscedasticity for accuracy and AUC metrics under various conditions. This step was crucial to ensure that the variance of the residuals met the assumptions required for the validity of the ANOVA analysis. The results of this test, detailed in Table 4, indicate that the variance across different groups remained consistent, supporting the reliability of our ANOVA findings.

To ensure the validity of the ANOVA test, we complemented the analysis with diagnostic checks. A boxplot in Figure 8 was constructed to visually assess the distribution of performance metrics, aiding in detecting outliers and evaluating group variance homogeneity (as supported by the Breusch-Pagan test in Table 4). Additionally, a Q-Q plot in Figure 9 was generated to thoroughly examine the residuals' normality. The Q-Q plot reveals how well the residuals correspond to a theoretical normal distribution, which is a central assumption for the validity of ANOVA. In the plot, the quantiles of the residuals are plotted against the expected quantiles of a normal distribution. The alignment of these points with the reference line (red line in Figure 9) indicates normality. The more closely the points adhere to this line, particularly in the center of the plot, the more evidence we have that the residuals are normally distributed. Minor deviations, especially in the tails of the distribution, can be acceptable but should not be systematic or extreme, as this could suggest non-normality and potential violations of ANOVA assumptions. In our analysis, the residuals largely conformed to the red line, substantiating the assumption of normality and thus supporting the validity of our ANOVA results. These findings, combined with the ANOVA results, provide a comprehensive understanding of the data and underscore the reliability of our statistical inferences.

Furthermore, we present the outcomes of applying three distinct adversarial attacks to our model: DeepFool, PGD, and JSMA. To ensure a comprehensive yet focused evaluation of our model's robustness and security, we strategically selected two datasets, BI2013a and BI2015b, from the original eight datasets used for the FGSM attack analysis. The selection of these datasets was guided by their contrasting complexity levels, as illustrated in Table 6. BI2013a, with the highest model complexity among the datasets, provides a rigorous test environment to evaluate the model's resilience under computationally intensive scenarios. On the other hand, BI2015b, representing one of the datasets with the lowest

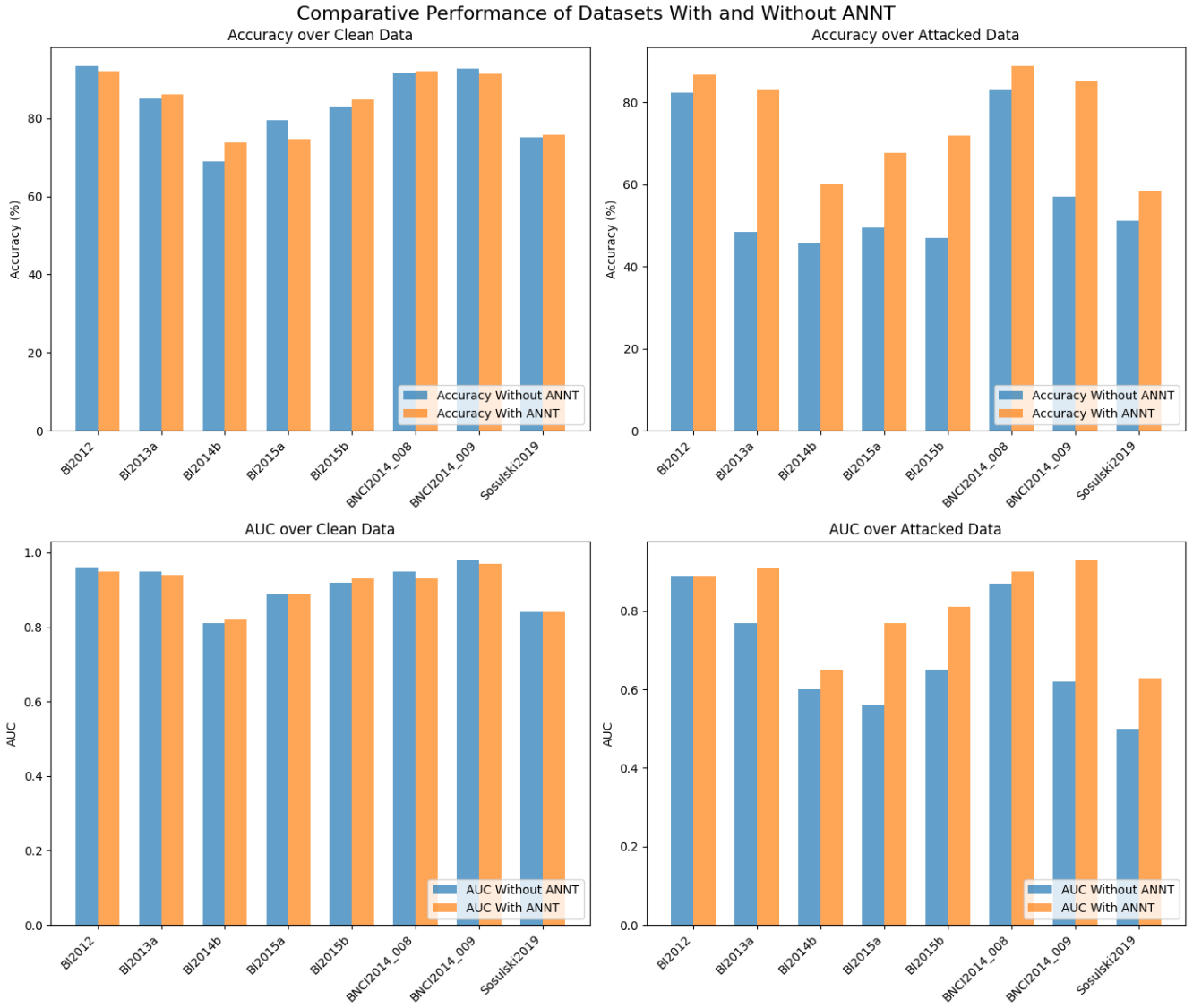

FIGURE 4. Comparative Performance of Datasets with and without ANNT. The top panels display the accuracy for each dataset on clean (left) and attacked (right) data, while the bottom panels show the AUC for the same conditions.

complexity, allows us to assess the model's performance under more lenient conditions. This deliberate choice of datasets with significantly different complexity levels enables a nuanced understanding of our model's robustness and security across various attack scenarios. As outlined in Table 5, the results demonstrate the model's performance variations under different attack scenarios. By focusing on two key datasets, this targeted approach enables us to provide detailed insights into the model's resilience against a broader spectrum of adversarial attacks, thereby underscoring its robustness and security.

Finally, we explore the computational complexity of the proposed model, with a particular focus on the adversarial perturbation process, which is implemented using the FGSM. The computational requirement of FGSM primarily depends on calculating the derivative of the loss function with respect to the input data, denoted by X. The complexity of this process is inherently linked to the model's architecture, including the number of layers (L), the number of neurons in each layer (H), and the dimensions of the input data (M). The computational complexity is thus represented mathematically as:

$$O(L \cdot H^2 \cdot M) \tag{6}$$

highlighting the reliance on gradient computation and matrix operations, which are fundamental to FGSM's execution. The term "O(.)" signifies the Big O notation, which mathematically expresses the upper limit of an algorithm's complexity. This notation is crucial for illustrating the worst-case scenario regarding execution time or space requirements, functioning

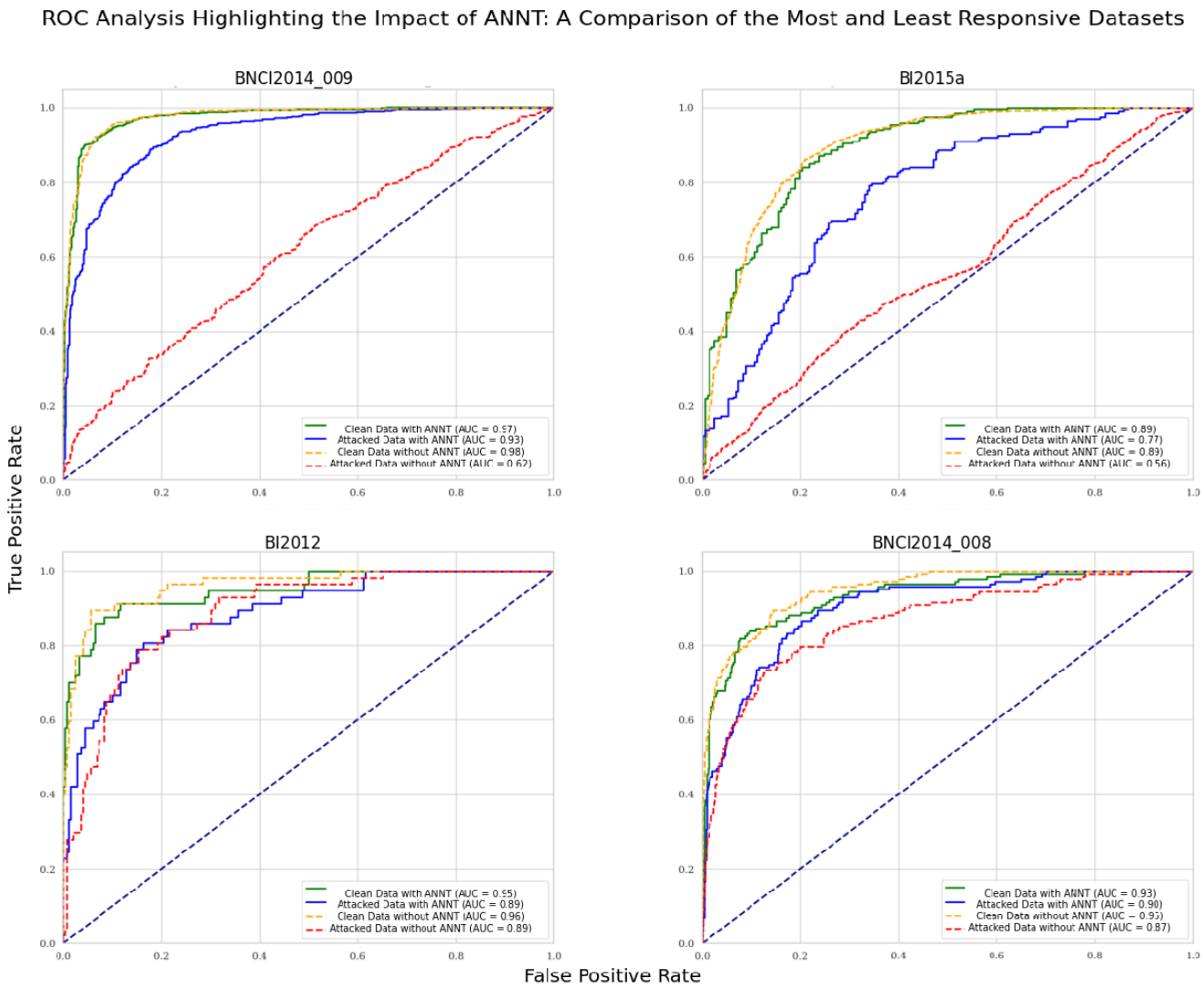

FIGURE 5. Differential Impact of ANNT on Dataset Performance: ROC curves for datasets where ANNT had a significant positive effect (top row) contrasted with those where its impact was less pronounced (bottom row), showcasing the varying degrees of enhanced robustness provided by ANNT under adversarial conditions.

as a comparative measure of algorithm efficiency, especially for large-scale inputs.

The architecture of our model includes layers such as Conv2D, MaxPooling2D, Reshape, and TCN, culminating in a Dense layer, as depicted in Figure 1. Given that the convolution operations within TCN layers differ from those in traditional fully connected layers, a non-linear increase in H mirrors the convolutional operations' multiplication activities.

Upon examining the system model illustrated in Figure 1, we pinpoint essential parameters for complexity analysis, including the number of computational layers and neurons per layer.

- L (The number of computational layers): This includes 2 Conv2D layers, 1 MaxPooling2D layer, and 1 TCN layer, making a total of 4 computational layers (excluding the input and reshape layers).
- H (Neurons per layer): This consists of 32 filters in the first Conv2D layer and 64 filters in both the second Conv2D layer and the TCN layer. The highest filter count, 64, is used for H.

Assuming M represents the size of the input data, its dimensions are determined by the aggregate product of trials (n_{trials}), channels (n_{channels}), and time points ($n_{\text{time points}}$) per trial. These parameters are derived from the dataset details provided in Table 2, enabling an accurate computation of the model's complexity for different datasets, as shown in Table 6.

This comprehensive analysis highlights the computational intricacies and paves the way for optimizing neural network

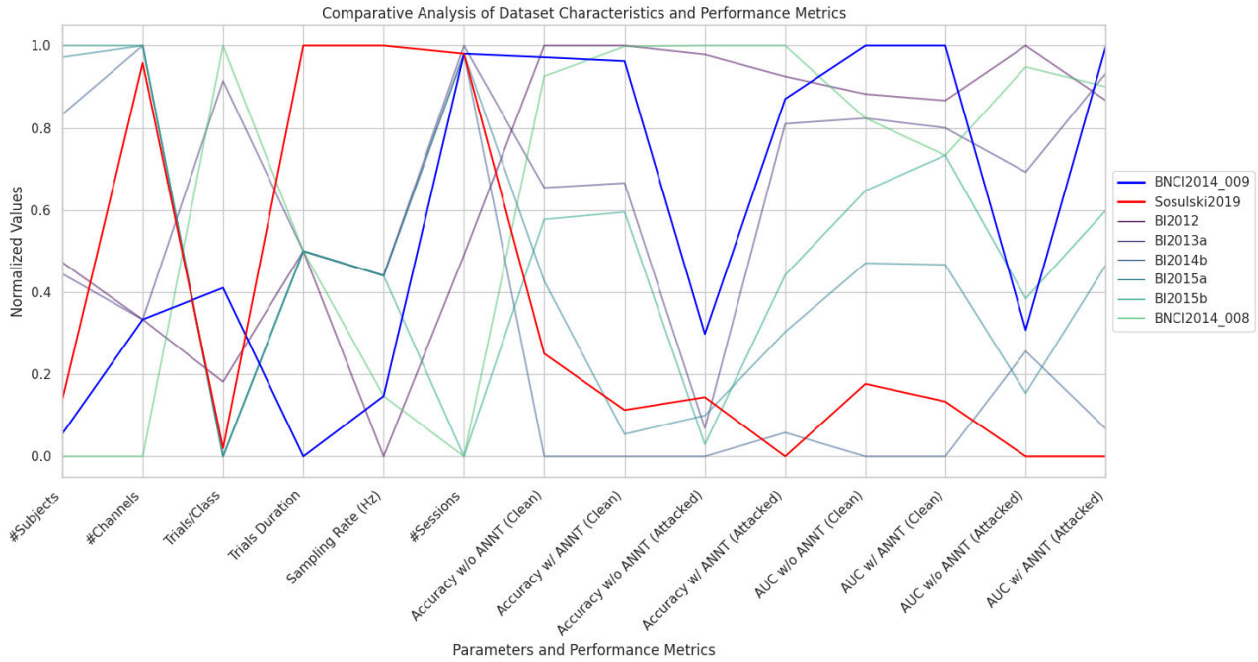

FIGURE 6. Normalised Association Between Performance Metrics and Dataset Characteristics. The values in this figure are normalized by subtracting the minimum value of each parameter and dividing by the range (maximum-minimum value), scaling all values to a range of 0-1 for uniform comparison across datasets. The bold blue line highlights the BNCI2014_009 dataset, demonstrating the most significant positive impact of ANNT due to its short Trial Duration and relatively lower Sampling Rate. The bold red line represents the Sosulski2019 dataset, indicating a negligible or non-positive impact of ANNT on improving performance in the presence of adversarial attacks, attributed to its long Trial Duration and high Sampling Rate.

TABLE 3. ANOVA results showing the influence of conditions and other factors on model performance metrics (Accuracy and AUC). This table compares performance under different conditions-with and without ANNT in clean and attacked scenarios-and examines how 'Trials Duration' and 'Sampling Rate' further affect these outcomes.

ACCURACY:

Source	Sum of Squares	df	F	PR(>F)
C(Condition)	3533.036	3.0	25.991	0.000002
Q("Trials Duration (s)")	188.295	1.0	4.156	0.058378
C(Condition):Q("Trials Duration (s)")	391.336	3.0	2.879	0.068506
Q("Sampling Rate (Hz)")	1525.151	1.0	33.659	0.000027
C(Condition):Q("Sampling Rate (Hz)")	381.198	3.0	2.804	0.073222
Q("Trials Duration (s)"):Q("Sampling Rate (Hz)")	420.182	1.0	9.273	0.007716
C(Condition):Q("Trials Duration (s)"):Q("Sampling Rate (Hz)")	77.325	3.0	0.569	0.643520
Residual	724.987	16.0	NaN	NaN

AUC:

Source	Sum of Squares	df	F	PR(>F)
C(Condition)	0.281	3.0	17.953	0.000023
Q("Trials Duration (s)")	0.003	1.0	0.495	0.491680
C(Condition):Q("Trials Duration (s)")	0.038	3.0	2.408	0.105162
Q("Sampling Rate (Hz)")	0.081	1.0	15.584	0.001152
C(Condition):Q("Sampling Rate (Hz)")	0.052	3.0	3.350	0.045450
Q("Trials Duration (s)"):Q("Sampling Rate (Hz)")	0.002	1.0	0.365	0.554274
C(Condition):Q("Trials Duration (s)"):Q("Sampling Rate (Hz)")	0.001	3.0	0.076	0.971969
Residual	0.083	16.0	NaN	NaN

models by considering computational resource allocation and the adoption of strategies like model pruning and quantization to mitigate computational burdens.

IV. DISCUSSION

A. EVALUATING METRIC RELEVANCE IN CLASS-IMBALANCED ERP EEG DATASETS FOR B2B-C SYSTEMS

In our study employing ERP EEG datasets, we recognize the inherent class imbalance resulting from the experimental design favoring less frequent target stimuli for P300 responses. This imbalance significantly impacts our metric choices.

Accuracy, while providing a quick performance overview, can be misleading due to the class distribution. A model might seem accurate by correctly predicting the more numerous non-target classes, but this does not accurately reflect its performance on the crucial target class.

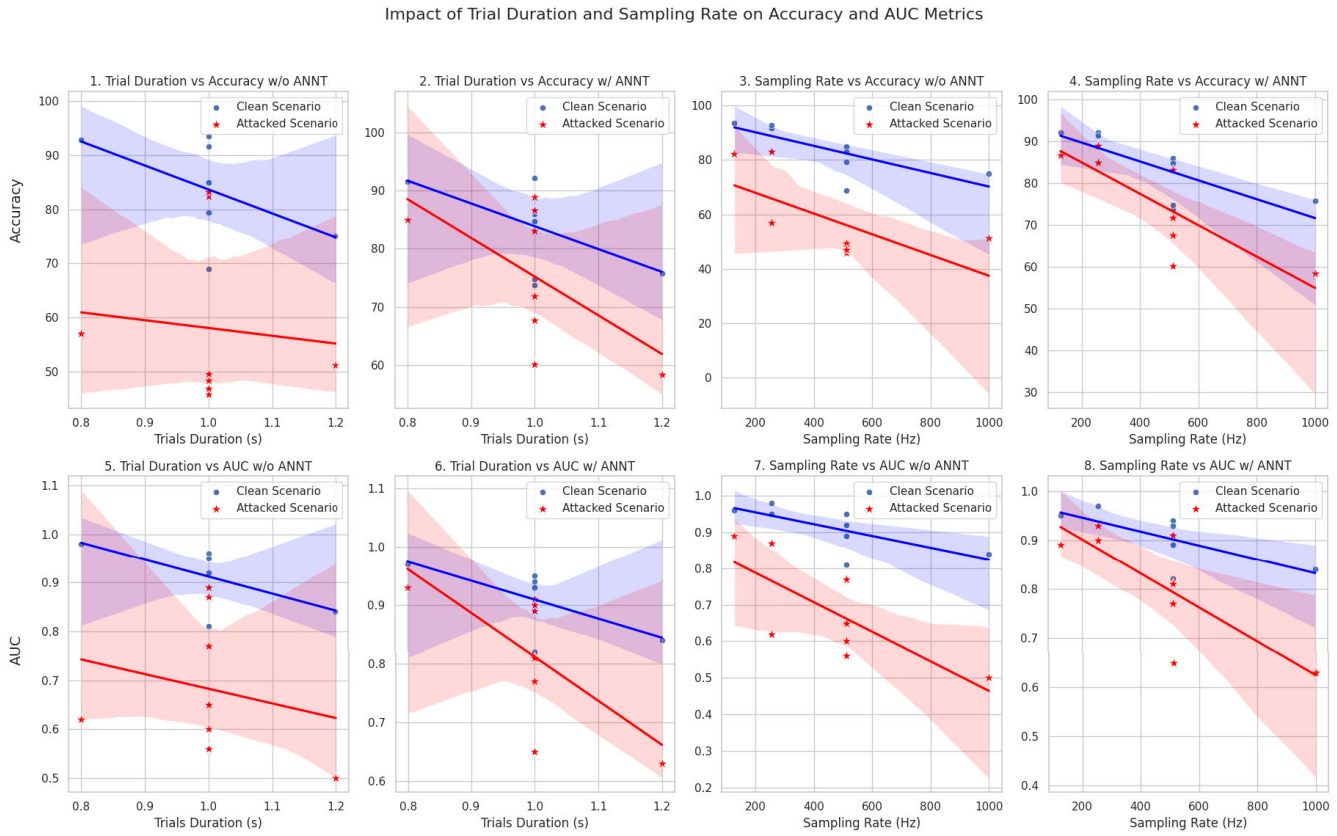

FIGURE 7. Impact of Trial Duration and Sampling Rate on the accuracy and AUC metrics in different scenarios, providing a nuanced understanding of how these dataset characteristics influence the performance of ANNT. The first two subplots (1 and 2) compare the accuracy against trial duration for clean and attacked data, showcasing the model's performance without and with ANNT. Subplots 3 and 4 continue this comparison for accuracy against the Sampling Rate, again for clean and attacked data conditions without and with ANNT application. The second row of plots shifts focus to AUC, with subplots 5 and 6 examining the relationship between AUC and Trial Duration and subplots 7 and 8 exploring AUC against Sampling Rate. Across all subplots, the differential effects of ANNT under varying levels of data integrity and attack simulation are visualized, allowing for a detailed assessment of model robustness.

TABLE 4. Breusch-Pagan test results for assessing homoscedasticity in model performance metrics (Accuracy and AUC).

Metric & Condition	Test Statistic	Test Statistic p-value	F-value	F-test p-value
AUC w/o ANNT (Clean)	0.246	0.884	0.079	0.925
AUC w/ ANNT (Clean)	0.237	0.888	0.076	0.928
AUC w/o ANNT (Attacked)	0.071	0.965	0.022	0.978
AUC w/ ANNT (Attacked)	0.203	0.904	0.065	0.938
Accuracy w/o ANNT (Clean)	0.335	0.846	0.109	0.899
Accuracy w/ ANNT (Clean)	0.732	0.693	0.252	0.787
Accuracy w/o ANNT (Attacked)	1.314	0.518	0.491	0.639
Accuracy w/ ANNT (Attacked)	0.181	0.913	0.058	0.944

We thus use AUC alongside accuracy, as it offers a comprehensive view of the model's ability to distinguish between target and non-target classes under varied thresholds. This is particularly pertinent for datasets with notable class imbalances. While accuracy gives a preliminary performance indication, AUC delves deeper, critically assessing the model's proficiency in identifying significant target events in B2B-C systems.

Linking to our study's aims, AUC's role becomes even more pivotal. In enhancing B2B-C systems' robustness against adversarial attacks, where precision in signal interpretation is key, relying solely on accuracy is insufficient due to the class distribution. AUC's broader evaluative scope, analyzing model robustness in differentiating stimuli under adversarial interference, aligns with our ANNT. By improving accuracy and adversarial robustness, ANNT benefits from AUC's insights into how well the model maintains communication integrity under threat.

After establishing the importance of selecting appropriate metrics for our class-imbalanced datasets, we now

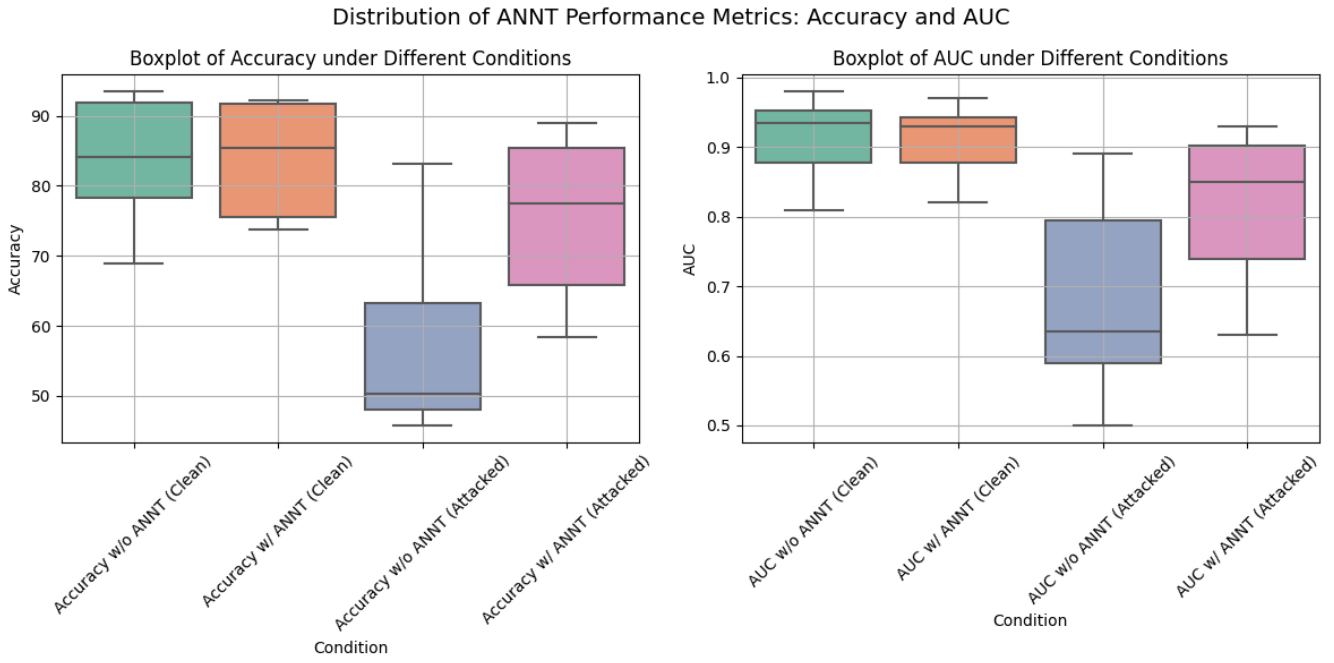

FIGURE 8. Comparative Distribution of Performance Metrics for ANNT. The left panel displays the range and distribution of accuracy metrics across different conditions, while the right panel focuses on the AUC metrics. Both panels highlight the variance and central tendencies of the model's performance, with and without ANNT, under clean and attacked scenarios.

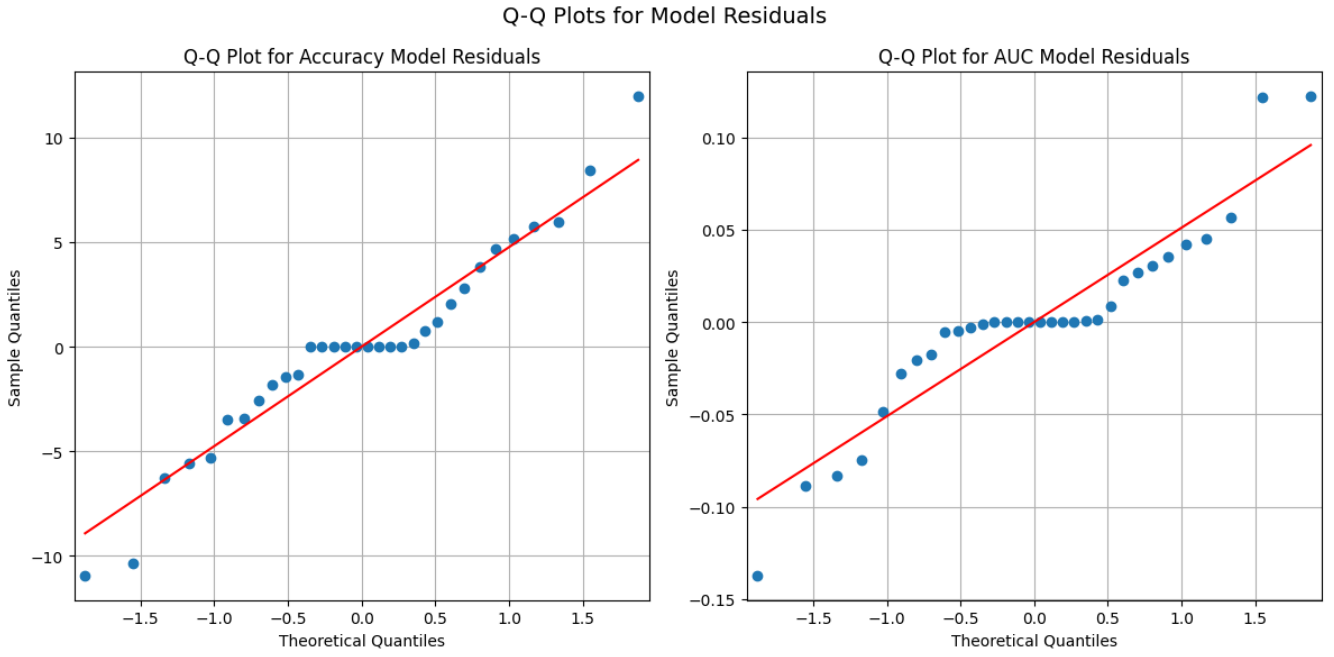

FIGURE 9. Q-Q Plot of Model Residuals. This quantile-quantile plot compares the distribution of residuals from the ANOVA model against a theoretical normal distribution. The close alignment of the data points along the red line indicates that the residuals approximate normality, satisfying one of the key assumptions required for the validity of ANOVA tests.

turn to the rationale behind our strategic dataset selection, which was instrumental in uncovering the nuanced impacts of various EEG features on the effectiveness of ANNT.

B. RATIONALE FOR DATASET SELECTION AND FEATURE IMPACT

In selecting our datasets, we meticulously assessed the influence of various EEG characteristics on model performance.

TABLE 5. Results of other attacks on BI2013a and BI2015b with and without ANNT.

Dataset	Attack Type	Condition	Accuracy	Precision	F1 Score	AUC	kappa
BI2013a	DeepFool	Without ANNT	0.4914	0.5395	0.4559	0.6663	0.3271
		With ANNT	0.7845	0.7860	0.7823	0.8535	0.7113
	PGD	Without ANNT	0.5690	0.6174	0.5434	0.6994	0.4122
		With ANNT	0.8190	0.8191	0.8186	0.8772	0.7572
	JSMA	Without ANNT	0.6207	0.6124	0.6157	0.7472	0.4933
		With ANNT	0.8276	0.8274	0.8267	0.8799	0.7684
BI2015b	DeepFool	Without ANNT	0.3707	0.5574	0.3356	0.5854	0.1728
		With ANNT	0.6305	0.6355	0.6298	0.7528	0.5070
	PGD	Without ANNT	0.5000	0.5473	0.5004	0.6710	0.3377
		With ANNT	0.6810	0.6832	0.6801	0.7850	0.5734
	JSMA	Without ANNT	0.5259	0.5206	0.5195	0.6847	0.3684
		With ANNT	0.6983	0.6975	0.6950	0.7974	0.5971

TABLE 6. Model complexity for each dataset.

Name	#Channel	Trials	Trials Duration	Sampling Rate	Model Complexity
BI2012	16	768	1 s	128 Hz	25,769,803,776
BNCI2014_008	8	4200	1 s	256 Hz	140,928,614,400
BNCI2014_009	16	1728	0.8 s	256 Hz	92,771,287,040
BI2013a	16	3840	1 s	512 Hz	515,396,075,520
BI2014b	32	6	1 s	512 Hz	1,610,612,736
BI2015a	32	6	1 s	512 Hz	1,610,612,736
BI2015b	32	6	1 s	512 Hz	1,610,612,736
Sosulski2019	31	90	1.2 s	1000 Hz	54,853,632,000

We commenced with BI2012, distinguished by its low Sampling Rate. To explore the influence of Sampling Rate, we then selected BNCI2014_008, which, while similar to BI2012 in Trial Duration, has a doubled Sampling Rate. Next, our focus shifted to the Trial Duration's impact with BNCI2014_009, which shared BNCI2014_008's Sampling Rate but featured shorter trials. We included BI2013a to investigate the effect of increased trials per class. Despite sharing channel count with BNCI2014_009, its doubled Sampling Rate, subject numbers, and trial per class differed, allowing us to evaluate the impact of more extensive data per class. Our analysis was further enriched by BI2014b, which offered a doubled channel count while retaining the Sampling Rate from BI2013a, shedding light on the influence of channel density on model performance. The inclusion of BI2015a and BI2015b, akin in channel number, trials per class, and Trial Duration, allowed us to probe deeper into the effects of these variables. With one matching the session number of BI2014b and the other having fewer, they provided a comparative perspective on the role of session frequency. Lastly, Sosulski2019 was integrated to examine the effects at the upper extremes of our variables, with a significantly higher Sampling Rate, equivalent channel and subject count, the same session numbers, and a moderate trial count, but notably, the longest Trial Duration.

With a careful selection of datasets designed to scrutinize the varied influences on model performance, we have laid the groundwork for a robust evaluation of ANNT's effectiveness. The subsequent analysis shifts from these methodological considerations to the tangible outcomes observed. We now turn to the critical impact and broader implications of ANNT, as evidenced by our empirical findings, to understand how this technique fortifies B2B-C systems against adversarial threats.

C. ADVERSARIAL NEURAL NETWORK TRAINING: IMPACT AND IMPLICATIONS

In assessing the robustness afforded by ANNT, we have refined our analysis to focus on the performance of individual datasets as depicted in Figure 4. This approach highlights the variability and dataset-specific effects of ANNT, providing a detailed view of the model's performance on each dataset, underscoring the nuanced nature of ANNT's benefits: improvements are evident, yet their magnitude and significance vary based on the characteristics of each dataset.

- **Improved Robustness to Adversarial Attacks.** Figure 4 shows a clear and significant improvement in accuracy and AUC for all datasets when ANNT is applied to attacked data, confirming the efficacy of ANNT in fortifying the model against adversarial

perturbations. The training with adversarial examples allows the model to retain its classification capabilities under adversarial conditions, indicative of a robust defense mechanism.

- **Minimal Impact on Clean Data.** The model performance on clean data, as inferred from individual dataset analysis, indicates minimal improvements post-ANNT application. This is attributed to the models' already high performance on clean data. Since adversarial training is designed to address perturbations absent in clean data, the primary advantage of ANNT is its maintenance of performance under compromised conditions rather than enhancing it in their absence.
- **Technical Explanation for the Observed Effects.** ANNT simulates potential attacks by incorporating adversarial examples during training, broadening the model's exposure to various perturbations that may otherwise lead to misclassification. This exposure cultivates a generalized data comprehension, bolstering the model's predictive robustness. Under adversarial attacks, the model leverages ANNT to discern and rectify disruptions, thus demonstrating its true value.
- **Implications for Model Deployment.** Our focus on individual dataset performance rather than averaged results emphasizes ANNT's value in practical scenarios where data integrity is at risk. The necessity of including ANNT in the training regimen becomes clear for ensuring model reliability and security in unpredictable environments. This analysis emphasizes ANNT's utility as a protective measure in real-world applications where data may be intentionally manipulated to provoke errors.

Thus far, our discussion has centered on the general efficacy of ANNT, as evidenced by improvements in model performance across various datasets. Figure 5 provides a targeted view, spotlighting the top and bottom datasets regarding ANNT's impact, as seen in the ROC curves. We conduct a normalized correlation analysis to deepen our examination and provide a more granular understanding of how dataset behaviors relate to specific characteristics. This allows us to dissect the extent to which individual dataset features contribute to the performance metrics observed. The forthcoming subsection delves into these findings, discussing the interplay between dataset attributes and the model's performance under the influence of ANNT.

D. ANALYSIS OF DATASET CHARACTERISTICS IN MODEL PERFORMANCE

In the realm of B2B-C systems, the diversity of EEG dataset characteristics can play a pivotal role in the performance of the models. To unravel the complex dynamics between these characteristics and the efficacy of ANNT, we conducted a normalized association analysis, the results of which are shown in Figure 6. This analysis informs our understanding of the datasets and sheds light on the implications of designing and applying robust B2B-C systems.

Figure 6 illustrates a marked variance in responsiveness to ANNT, with Trial Duration and Sampling Rate emerging as the most impactful factors. Notably, datasets with shorter trial durations, such as BNCI2014_009, seem to benefit more from ANNT, suggesting that the condensation of critical information within a smaller temporal window enhances the model's capability to resist adversarial interference. In contrast, extended trial durations, like those seen in Sosulski2019, may dilute the model's discriminatory power by providing a broader attack surface for adversarial noise.

Similarly, a higher Sampling Rate could be a double-edged sword. While it may increase the temporal resolution of the data, it also potentially introduces more opportunities for adversarial perturbations to integrate with the genuine EEG signals. This complexity is visually and quantitatively captured in Figure 6, where the interplay between Trial Duration and Sampling Rate significantly influences ANNT's effectiveness.

We have conducted a detailed statistical investigation to elucidate these two critical characteristics' individual and combined effects on model performance. We present this as a scatter plot analysis, referred to as the 'Performance Impact Scatter Plot,' shown in Figure 7. This analysis will allow us to discern how each characteristic affects the model's accuracy and AUC under clean and adversarial conditions.

We have undertaken a rigorous statistical investigation to elucidate further the individual and combined effects of Trial Duration and Sampling Rate on model performance in the following subsection.

E. STATISTICAL ANALYSIS

Figure 7 provides a visual confirmation of the significant roles that Trial Duration and Sampling Rate play in the performance of ANNT. Each panel contrasts the effects of these key characteristics on model accuracy and AUC under both clean and adversarial conditions.

Subplots 1, 2, 5, and 6 correlating Trial Duration with model performance metrics reveal that shorter trials generally correspond with higher accuracy and AUC, particularly when under adversarial attack. This suggests that shorter trial durations allow ANNT to effectively enhance the model's resistance to such attacks, likely due to a more condensed and distinct representation of the target stimuli.

Conversely, when examining the impact of the Sampling Rate as illustrated in subplots 3, 4, 7, and 8, it becomes evident that while higher rates provide more detail by capturing finer temporal resolutions, they may also introduce additional complexity that can be exploited by adversarial noise. The plots indicate a nuanced balance to be struck; a higher Sampling Rate does not unilaterally lead to better performance and may, in some cases, be detrimental when coupled with longer Trial Durations.

Figure 7 acts as a visual supplement to our normalized association analysis from Figure 6, supporting the assertion that Trial Duration and Sampling Rate are predominant factors affecting ANNT's performance. To quantify the

magnitude of their impact, we extended our analysis to include an ANOVA test. The results of this statistical analysis are presented in Table 3. The ANOVA test allows us to discern the individual contributions of each variable to the performance metrics.

Our statistical examination via ANOVA, presented in Table 3, underscores the significant influence of the Sampling Rate on the efficacy of ANNT for accuracy metrics. The F-statistic for Sampling Rate is substantial (F=33.659, p=0.000027), confirming its crucial impact on model performance under adversarial conditions. However, Trial Duration shows a less pronounced effect on accuracy (F=4.156, p=0.058378), suggesting a marginal trend that does not reach conventional levels of statistical significance.

For AUC metrics, the Sampling Rate again proves to be a significant factor (F=15.584, p=0.001152), reinforcing its pivotal role in determining model efficacy. In contrast, Trial Duration appears to have a minimal impact on AUC metrics (F=0.495, p=0.491680), suggesting its influence is not statistically significant in this context.

In the ANOVA table, the interaction effects, such as C(Condition): Q(Trials Duration (s)) and C(Condition): Q(Sampling Rate (Hz)), with p-values exceeding the 0.05 threshold, indicate an inconsistent influence of Trial Duration and Sampling Rate across different conditions.

Upon extending our analysis to examine homoscedasticity through the Breusch-Pagan test, as shown in Table 4, we confirmed the uniform variance across different levels of our independent variables. Our model diagnostics, including the boxplots in Figure 8, confirm the homogeneity of variances and provide a visual representation of performance metric distributions across conditions. The Q-Q plot in Figure 9 confirms the normality of residuals.

Including the Breusch-Pagan test results complements our diagnostic checks, the boxplots, and the Q-Q plot, reinforcing the robustness of our statistical conclusions. These checks confirm the assumptions necessary for the validity of ANOVA, providing a comprehensive understanding of our data and underlining the reliability of our statistical inferences.

F. QUANTITATIVE EVALUATION OF MODEL ACCURACY AND AUC

Our empirical investigation presents a quantitative enhancement in model performance attributable to the application of ANNT. Table 7 encapsulates this enhancement, delineating an average accuracy increment of 17% in the face of adversarial attacks and an average AUC increment of 0.12 points. These increments are not merely numerical improvements; they substantially enhance the model's ability to maintain its integrity under adversarial duress, an increasingly relevant scenario in deploying B2B-C systems in security-critical applications.

The application of ANNT, as substantiated by the comparative performance metrics, delineates a clear trajectory towards robustness—a facet of performance that is paramount in real-world settings. The bolstered accuracy and AUC under adversarial conditions indicate ANNT's efficacy in reinforcing the model's defense mechanisms. This is particularly noteworthy in the context of B2B-C systems, where the accuracy and precision of communication are critical. The augmented adversarial accuracy suggests a fortified system that can reliably interpret and transmit neural signals even under sophisticated electronic attack strategies, thus mitigating the risk of erroneous interpretations or malicious signal manipulations.

TABLE 7. Quantitative evaluation of model accuracy with ANNT application under clean and adversarial conditions.

Dataset	Accuracy (%)		AUC	
	Clean	Attacked	Clean	Attacked
BI2012	93.49	82.41	0.96	0.89
BI2013a	85.00	48.36	0.95	0.77
BI2014b	68.91	45.78	0.81	0.60
BI2015a	79.41	49.51	0.89	0.56
BI2015b	83.10	46.88	0.92	0.65
BNCI2014_008	91.67	83.21	0.95	0.87
BNCI2014_009	92.80	56.94	0.98	0.62
Sosulski2019	75.06	51.17	0.84	0.50
Average Increase	**17%**		**0.12**	

G. ASSESSING MODEL RESILIENCE: INSIGHTS FROM DIVERSE ADVERSARIAL ATTACKS

In exploring our model's robustness against adversarial attacks, we subjected it to a series of sophisticated threats, including DeepFool, PGD, and JSMA, each chosen for their relevance to the nuances of B2B-C signal analysis. The comparative analysis of these attacks provided a comprehensive understanding of our model's resilience, revealing notable differences in their impact on model performance, as presented in Table 5.

DeepFool emerged as the most effective attack in revealing the model's vulnerabilities, significantly reducing all metrics across both BI2013a and BI2015b datasets. Its efficacy can be attributed to its ability to craft minimal yet impactful perturbations, which closely mimic potential real-world signal disruptions. This precision underscores the importance of designing countermeasures that address subtle adversarial manipulations, particularly relevant in ERP EEG signals where minor perturbations can lead to misinterpretations.

The PGD attack, with its iterative approach to applying perturbations, also posed a substantial threat but was slightly less impactful than DeepFool. This suggests that while PGD effectively simulates attack scenarios with incremental complexity, DeepFool's minimal perturbation approach is more aligned with the specific challenges of securing B2B-C models.

Though computationally intensive, JSMA offered valuable insights into the model's sensitivity to feature-specific alterations, aligning its effectiveness closely with that of FGSM. This highlights the necessity of understanding and

mitigating targeted attacks that exploit specific vulnerabilities within the data or model architecture.

Crucially, the incorporation of ANNT showcases a significant advancement in bolstering the model's defense mechanisms against the adversarial attacks discussed. The comparative results with and without ANNT, as detailed in Table 5, illuminate its profound impact on recovering model performance metrics to levels near those observed under clean data conditions. Despite the varying degrees of threat posed by DeepFool, PGD, and JSMA, ANNT has consistently mitigated their adverse effects, restoring accuracy, precision, F1 score, AUC, and Kappa to substantially higher values. This resilience is particularly noteworthy in the face of DeepFool's sophisticated perturbations, which present the most considerable challenge to model robustness. The efficacy of ANNT against such diverse and potent attacks not only underlines its value as a critical component of our defense strategy but also reinforces the model's capability to maintain high performance in real-world scenarios where adversarial threats are unpredictable. By effectively countering these attacks, ANNT contributes to a more generalized, robust, and secure framework for B2B-C analysis, ensuring the integrity and reliability of the communication channel even in the presence of sophisticated adversarial interventions.

H. COMPUTATIONAL COMPLEXITY ANALYSIS

The analysis of model complexity, as shown in Table 6, reveals a deep connection between computational demands and dataset characteristics, setting the stage for optimizing neural network models more effectively. This analysis highlights how crucial factors such as sampling rate and trial count significantly impact computational complexity, emphasizing their importance in allocating computational resources. The increase in data processing requirements, driven by higher sampling rates and larger trial volumes, necessitates adopting advanced computational strategies to handle this growth efficiently.

Furthermore, the added complexity from the number of channels and the variation in trial duration complicates the computational landscape even further. These elements collectively influence the model's data processing efficiency, underlining the need for adaptive model architectures to meet the diverse requirements of different datasets.

This insight underpins the necessity to consider computational limits and optimization opportunities in the model's practical application. Employing strategies like model pruning, quantization, and efficient data processing algorithms is crucial for reducing computational load. These strategies are instrumental in boosting model performance and ensuring the sustainability of computational resources.

Additionally, the importance of balancing model complexity with performance, particularly in the context of adversarial threats, cannot be overstated. This balance requires fine-tuning model parameters to achieve optimal efficiency while preserving the integrity of adversarial defense mechanisms.

I. METHODOLOGICAL CONSIDERATIONS, BROADER IMPLICATIONS, AND FUTURE DIRECTIONS

Our research on ANNT for B2B-C systems underscores the potential and limitations of current methodologies and sets the stage for future explorations with wide-ranging implications. While we have demonstrated significant advancements in securing B2B-C systems against adversarial attacks, the challenges encountered beckon for comprehensive future research efforts.

Firstly, we propose expanding the scope of adversarial strategies explored in future studies to account for the unique perturbations each method introduces, aiming for a more nuanced understanding of ANNT's robustness and versatility. The critical challenge of transmitting EEG data over wireless channels without compromising its integrity also calls for advanced encryption and transmission techniques tailored to neural data.

Moreover, our study's focus on ERP EEG datasets opens up avenues for future research to include a broader spectrum of neural tasks and signals, thereby extending the applicability of our findings. Addressing practical challenges in real-world scenarios, such as jamming signals and hardware impairments, remains pivotal for evaluating and enhancing ANNT's deployment in operational environments.

Beyond these methodological considerations, our study hints at broader interdisciplinary research and development implications. The insights gained from enhancing B2B-C system security can inform advancements in smart healthcare, neurotechnology, and secure communication platforms. This suggests a future where such technologies are more integrated into daily life and healthcare practices.

In summary, our envisioned future direction involves methodological advancements, tackling practical challenges, and exploring the transformative potential of ANNT in B2B-C systems across various domains. This comprehensive approach aims to advance the efficacy and reliability of B2B-C systems and contribute to the broader field of secure and efficient communication technologies.

V. CONCLUSION

Our study underscores the pivotal role of ANNT in bolstering the resilience of B2B-C systems against adversarial interferences. A key finding is the significant influence of Sampling Rate over Trial Duration on the performance enhancement through ANNT, pointing towards an optimum balance that maximizes system robustness. These insights pave the way for developing more secure and efficient B2B-C frameworks and hold profound implications for the biomedical domain. Specifically, the enhanced robustness against adversarial threats ensures the integrity and reliability of B2B-C systems, which are crucial for applications such as remote healthcare monitoring, neurorehabilitation, and BCIs for assistive technologies. By ensuring the secure and

effective transmission of EEG signals, our research advances telemedicine and personalized healthcare, where accurate and reliable brain signal interpretation can significantly impact patient outcomes. Future research should expand on these findings to explore a broader spectrum of adversarial threats and their countermeasures, further solidifying the foundation for secure, reliable, and efficient B2B-C in biomedical applications.

REFERENCES

[1] C. Grau, R. Ginhoux, A. Riera, T. L. Nguyen, H. Chauvat, M. Berg, J. L. Amengual, A. Pascual-Leone, and G. Ruffini, "Conscious brain-to-brain communication in humans using non-invasive technologies," *PLoS ONE*, vol. 9, no. 8, Aug. 2014, Art. no. e105225, doi: 10.1371/journal.pone.0105225.

[2] S. L. Bernal, A. H. Celdrán, G. M. Pérez, M. T. Barros, and S. Balasubramaniam, "Security in brain–computer interfaces: State-of-the-art, opportunities, and future challenges," *ACM Comput. Surveys*, vol. 54, no. 1, pp. 1–35, Jan. 2022, doi: 10.1145/3427376.

[3] K. M. Hossain, M. A. Islam, S. Hossain, A. Nijholt, and M. A. R. Ahad, "Status of deep learning for EEG-based brain–computer interface applications," *Frontiers Comput. Neurosci.*, vol. 16, Jan. 2023, Art. no. 1006763, doi: 10.3389/fncom.2022.1006763.

[4] D. Adesina, C.-C. Hsieh, Y. E. Sagduyu, and L. Qian, "Adversarial machine learning in wireless communications using RF data: A review," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 77–100, 1st Quart., 2023, doi: 10.1109/COMST.2022.3205184.

[5] L. Meng, X. Jiang, and D. Wu, "Adversarial robustness benchmark for EEG-based brain–computer interfaces," *Future Gener. Comput. Syst.*, vol. 143, pp. 231–247, Jun. 2023, doi: 10.1016/j.future.2023.01.028.

[6] D. Wu, J. Xu, W. Fang, Y. Zhang, L. Yang, X. Xu, H. Luo, and X. Yu, "Adversarial attacks and defenses in physiological computing: A systematic review," *Nat. Sci. Open*, vol. 2, no. 1, Jan. 2023, Art. no. 20220023, doi: 10.1360/nso/20220023.

[7] J. K. Samriya, C. Chakraborty, A. Sharma, and M. Kumar, "Adversarial ML-based secured cloud architecture for consumer Internet of Things of smart healthcare," *IEEE Trans. Consum. Electron.*, early access, Dec. 12, 2023, doi: 10.1109/TCE.2023.3341696.

[8] J. Jung, H. Moon, G. Yu, and H. Hwang, "Generative perturbation network for universal adversarial attacks on brain–computer interfaces," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 11, pp. 5622–5633, Nov. 2023, doi: 10.1109/JBHI.2023.3303494.

[9] R. C. Moioli, P. H. J. Nardelli, M. T. Barros, W. Saad, A. Hekmatmanesh, P. E. G. Silva, A. S. de Sena, M. Dzaferagic, H. Siljak, W. Van Leekwijck, D. C. Melgarejo, and S. Latré, "Neurosciences and wireless networks: The potential of brain-type communications and their applications," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1599–1621, 3rd Quart., 2021, doi: 10.1109/COMST.2021.3090778.

[10] R. P. N. Rao, A. Stocco, M. Bryan, D. Sarma, T. M. Youngquist, J. Wu, and C. S. Prat, "A direct brain-to-brain interface in humans," *PLoS ONE*, vol. 9, no. 11, Nov. 2014, Art. no. e111332, doi: 10.1371/journal.pone.0111332.

[11] S. Rajesh, V. Paul, V. G. Menon, S. Jacob, and P. Vinod, "Secure brain-to-brain communication with edge computing for assisting post-stroke paralyzed patients," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 2531–2538, Apr. 2020, doi: 10.1109/JIOT.2019.2951405.

[12] R. Ajmeria, M. Mondal, R. Banerjee, T. Halder, P. K. Deb, D. Mishra, P. Nayak, S. Misra, S. K. Pal, and D. Chakravarty, "A critical survey of EEG-based BCI systems for applications in industrial Internet of Things," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 184–212, 1st Quart., 2023, doi: 10.1109/COMST.2022.3232576.

[13] K.-J. Wang, C. Y. Zheng, M. Shidujaman, M. Wairagkar, and M. von Mohr, "Jean Joseph v2.0 (REmotion): Make remote emotion touchable, seeable and thinkable by direct brain-to-brain telepathy neurohaptic interface empowered by generative adversarial network," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Toronto, ON, Canada, Oct. 2020, pp. 3488–3493, doi: 10.1109/SMC42975.2020.9283049.

[14] J. Polich, "Updating p300: An integrative theory of P3a and P3b," *Clin. Neurophysiol.*, vol. 118, no. 10, pp. 2128–2148, Oct. 2007, doi: 10.1016/j.clinph.2007.04.019.

[15] M. Kutas and K. D. Federmeier, "Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP)," *Annu. Rev. Psychol.*, vol. 62, no. 1, pp. 621–647, Jan. 2011, doi: 10.1146/annurev.psych.093008.131123.

[16] J. Polich and A. Kok, "Cognitive and biological determinants of p300: An integrative review," *Biol. Psychol.*, vol. 41, no. 2, pp. 103–146, Oct. 1995, doi: 10.1016/0301-0511(95)05130-9.

[17] E. Donchin and M. G. H. Coles, "Is the P300 component a manifestation of context updating?" *Behav. Brain Sci.*, vol. 11, no. 03, p. 357, Sep. 1988, doi: 10.1017/s0140525x00058027.

[18] H. Cecotti and A. Graser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 433–445, Mar. 2011, doi: 10.1109/TPAMI.2010.125. https://doi.org/10.1109/TPAMI.2010.125

[19] S. Bai, J. Zico Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*.

[20] R. K M and M. Siddappa, "Electronically linked brain to brain communication in humans using non-invasive technologies," in *Proc. Int. Conf. Emerg. Res. Electron., Comput. Sci. Technol. (ICERECT)*, Mandya, India, Dec. 2015, pp. 235–239, doi: 10.1109/ERECT.2015.7499019.

[21] Y. Yu, C. Qian, Z. Wu, and G. Pan, "Mind-controlled ratbot: A brain-to-brain system," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PERCOM WORKSHOPS)*, Budapest, Hungary, Mar. 2014, pp. 228–231, doi: 10.1109/PERCOMW.2014.6815207.

[22] J. Toppi, A. Ciaramidaro, P. Vogel, D. Mattia, F. Babiloni, M. Siniatchkin, and L. Astolfi, "Graph theory in brain-to-brain connectivity: A simulation study and an application to an EEG hyperscanning experiment," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Milan, Italy, Aug. 2015, pp. 2211–2214, doi: 10.1109/EMBC.2015.7318830.

[23] K. Hameed, I. S. Ateeq, S. H. Khan, and S. Tabassum, "An EEG analysis approach towards brain-to-brain synchronization," in *Proc. IEEE Asia–Pacific Conf. Comput. Sci. Data Eng. (CSDE)*, Gold Coast, QLD, Australia, Dec. 2020, pp. 1–6, doi: 10.1109/CSDE50874.2020.9411590.

[24] Q. Ma, W. Gao, Q. Xiao, L. Ding, T. Gao, Y. Zhou, X. Gao, T. Yan, C. Liu, Z. Gu, X. Kong, Q. H. Abbasi, L. Li, C.-W. Qiu, Y. Li, and T. J. Cui, "Directly wireless communication of human minds via non-invasive brain-computer-metasurface platform," *eLight*, vol. 2, no. 1, pp. 1–11, Jun. 2022, doi: 10.1186/s43593-022-00019-x.

[25] M. Pais-Vieira, M. Lebedev, C. Kunicki, J. Wang, and M. A. L. Nicolelis, "A brain-to-brain interface for real-time sharing of sensorimotor information," *Sci. Rep.*, vol. 3, no. 1, pp. 1–10, Feb. 2013, doi: 10.1038/srep01319.

[26] C. Micek, T. Wilaiprasitporn, and T. Yagi, "A study on SSVEP-based brain synchronization: Road to brain-to-brain communication," in *Proc. 9th Biomed. Eng. Int. Conf. (BMEiCON)*, Laung Prabang, Laos, Dec. 2016, pp. 1–5, doi: 10.1109/BMEiCON.2016.7859615.

[27] M. Dingemanse, "Brain-to-brain interfaces and the role of language in distributing agency," in *Distributed Agency, Foundations of Human Interaction*, N. J. Enfield and P. Kockelman, Eds. New York, NY, USA: Oxford Academic, Feb. 2017, doi: 10.1093/acprof:oso/9780190457204.003.0007.

[28] M. Pais-Vieira, G. Chiuffa, M. Lebedev, A. Yadav, and M. A. L. Nicolelis, "Building an organic computing device with multiple interconnected brains," *Sci. Rep.*, vol. 5, no. 1, pp. 1–15, Jul. 2015, doi: 10.1038/srep11869.

[29] S.-S. Yoo, H. Kim, E. Filandrianos, S. J. Taghados, and S. Park, "Non-invasive brain-to-brain interface (BBI): Establishing functional links between two brains," *PLoS ONE*, vol. 8, no. 4, Apr. 2013, Art. no. e60410, doi: 10.1371/journal.pone.0060410.

[30] F. R. Willett, D. T. Avansino, L. R. Hochberg, J. M. Henderson, and K. V. Shenoy, "High-performance brain-to-text communication via handwriting," *Nature*, vol. 593, no. 7858, pp. 249–254, May 2021, doi: 10.1038/s41586-021-03506-2.

[31] C. S. Nam, Z. Traylor, M. Chen, X. Jiang, W. Feng, and P. Y. Chhatbar, "Direct communication between brains: A systematic PRISMA review of brain-to-brain interface," *Frontiers Neurorobotics*, vol. 15, May 2021, Art. no. 656943, doi: 10.3389/fnbot.2021.656943.

[32] A. Hekmatmanesh, H. M. Azni, H. Wu, M. Afsharchi, M. Li, and H. Handroos, "Imaginary control of a mobile vehicle using deep learning algorithm: A brain computer interface study," *IEEE Access*, vol. 10, pp. 20043–20052, 2022, doi: 10.1109/ACCESS.2021.3128611.

[33] L. Lu, R. Wang, and M. Luo, "An optical brain-to-brain interface supports rapid information transmission for precise locomotion control," *Sci. China Life Sci.*, vol. 63, no. 6, pp. 875–885, Jun. 2020, doi: 10.1007/s11427-020-1675-x.

[34] L. Jiang, A. Stocco, D. M. Losey, J. A. Abernethy, C. S. Prat, and R. P. N. Rao, "BrainNet: A multi-person brain-to-brain interface for direct collaboration between brains," *Sci. Rep.*, vol. 9, no. 1, p. 6115, Apr. 2019, doi: 10.1038/s41598-019-41895-7.
[35] S. Ajrawi, R. Rao, and M. Sarkar, "Cybersecurity in brain–computer interfaces: RFID-based design-theoretical framework," *Informat. Med. Unlocked*, vol. 22, 2021, Art. no. 100489, doi: 10.1016/j.imu.2020.100489.
[36] Z. Tarkhani, L. Qendro, M. O'Connor Brown, O. Hill, C. Mascolo, and A. Madhavapeddy, "Enhancing the security & privacy of wearable brain–computer interfaces," 2022, *arXiv:2201.07711*.
[37] F. Brocal, "Brain–computer interfaces in safety and security fields: Risks and applications," *Saf. Sci.*, vol. 160, Apr. 2023, Art. no. 106051, doi: 10.1016/j.ssci.2022.106051.
[38] G. Van Veen, A. Barachant, A. Andreev, G. Cattan, P. C. Rodrigues, and M. Congedo, "Building brain invaders: EEG data of an experimental validation," 2019, *arXiv:1905.05182*.
[39] E. Vaineau, A. Barachant, A. Andreev, P. C. Rodrigues, G. Cattan, and M. Congedo, "Brain invaders adaptive versus non-adaptive P300 brain–computer interface dataset," 2019, *arXiv:1904.09111*.
[40] A. Barachant and M. Congedo, "A plug & play P300 BCI using information geometry," 2014, *arXiv:1409.0107*.
[41] M. Congedo, M. Goyat, N. Tarrin, G. Ionescu, B. Rivet, L. Varnet, R. Phlypo, N. Jrad, M. Acquadro, and C. Jutten, "'Brain invaders': A prototype of an open-source P300-based video game working with the OpenViBE platform," in *Proc. IBCI Conf.*, Graz, Austria, 2011, pp. 280–283.
[42] L. Korczowski, E. Ostaschenko, A. Andreev, G. Cattan, P. L. C. Rodrigues, V. Gautheret, M. Congedo. (2019). *Brain Invaders Solo Versus Collaboration: Multi-User P300-Based Brain-Computer Interface Dataset (BI2014b)*. [Online]. Available: https://hal.archives-ouvertes.fr/hal-02173958
[43] L. Korczowski, M. Cederhout, A. Andreev, G. Cattan, P. L. C. Rodrigues, V. Gautheret, M. Congedo. (2019). *Brain Invaders Calibration-less P300-based BCI With Modulation of Flash Duration Dataset (BI2015a)*. [Online]. Available: https://hal.archives-ouvertes.fr/hal-02172347
[44] L. Korczowski, M. Cederhout, A. Andreev, G. Cattan, P. L. C. Rodrigues, V. Gautheret, M. Congedo. (2019). *Brain Invaders Cooperative Versus Competitive: Multi-User P300-based Brain-Computer Interface Dataset (BI2015b)*. [Online]. Available: https://hal.archives-ouvertes.fr/hal-02172347
[45] A. Riccio, L. Simione, F. Schettini, A. Pizzimenti, M. Inghilleri, M. O. Belardinelli, D. Mattia, and F. Cincotti, "Attention and P300-based BCI performance in people with amyotrophic lateral sclerosis," *Frontiers Human Neurosci.*, vol. 7, p. 732, 2013.
[46] L. A. Farwell and E. Donchin, "Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalogr. Clin. Neurophysiology*, vol. 70, no. 6, pp. 510–523, Dec. 1988, doi: 10.1016/0013-4694(88)90149-6.
[47] P. Aricò, F. Aloise, F. Schettini, S. Salinari, D. Mattia, and F. Cincotti, "Influence of P300 latency jitter on event related potential-based brain–computer interface performance," *J. Neural Eng.*, vol. 11, no. 3, Jun. 2014, Art. no. 035008.
[48] J. Sosulski and M. Tangermann, "Electroencephalogram signals recorded from 13 healthy subjects during an auditory oddball paradigm under different stimulus onset asynchrony conditions," Dataset, 2019, doi: 10.6094/UNIFR/154576.
[49] J. Sosulski and M. Tangermann, "Spatial filters for auditory evoked potentials transfer between different experimental conditions," in *Proc. Graz BCI Conf.*, 2019, pp. 1–6.
[50] J. Sosulski, D. Hübner, A. Klein, and M. Tangermann, "Online optimization of stimulation speed in an auditory brain–computer interface under time constraints," 2021, *arXiv:2109.06011*.

HOSSEIN AHMADI received the B.S. degree in electronics engineering from Kurdistan University, Sanandaj, Iran, in 2010, and the M.S. degree in telecommunications engineering from the Amirkabir University of Technology, Tehran, Iran, in 2016. He is currently pursuing the Ph.D. degree with Politecnico di Torino, Italy, specializing in electrical, electronics, and communications engineering. His research interests include 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 currently an Associate Professor with 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.

LUCA MESIN received the bachelor's degree in electronics engineering, in 1999, and the Ph.D. degree in applied mathematics from Politecnico di Torino, Italy, in 2003. He is currently an Associate Professor of biomedical engineering and the 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.

Open Access funding provided by 'Politecnico di Torino' within the CRUI CARE Agreement