



Article

Q-DP-GAN: Improving EEG Data Privacy Through Quantum-Inspired Differential Privacy-Based GAN

Shouvik Paul and Garima Bajwa *

Department of Computer Science, Lakehead University, Thunder Bay, ON P7B 5E1, Canada; spaul4@lakeheadu.ca

* Correspondence: garima.bajwa@lakeheadu.ca

Abstract

Electroencephalography (EEG)-based brain–computer interface (BCI) systems pose significant privacy risks, as EEG data remain vulnerable to inference and reconstruction attacks. Conventional privacy-preserving techniques, including data anonymization, encryption, and perturbation, frequently compromise data utility or prove ineffective against advanced adversaries. To address these limitations and balance utility and privacy, we propose a quantum-inspired, differential privacy-based generative adversarial network (Q-DP-GAN). Unlike classical GANs, which lack adaptive privacy mechanisms during training, our method uses quantum-inspired stochasticity to dynamically calibrate noise and the privacy budget. The experimental results demonstrate that Q-DP-GAN is more robust to membership inference and reconstruction attacks than existing approaches. Evaluation on the widely used BCI Competition IV Datasets 2A and 2B indicates that our framework produces high-quality synthetic EEG data while maintaining utility and data confidentiality for BCI classification tasks.

Keywords: brain–computer interfaces; differential privacy; EEG; generative adversarial network (GAN); privacy; quantum-inspired; hybrid architecture

1. Introduction

Brain–computer interface (BCI) technology is rapidly evolving and has enabled the development of intricate neuroprosthetics and applications in healthcare, entertainment, and even personal security. Electroencephalography (EEG)-based BCIs record electrical activity in the brain and provide a bridge between human cognitive states and computerized procedures. EEG recordings were first demonstrated by Hans Berger in 1924 [1,2] and have since provided a new way of interacting with and controlling the human brain [3–6]. BCIs were developed in the 1970s to capture and analyze brain signals from a user and translate them into actions to control external devices [4]. Brain–machine interface (BMI) systems, in turn, use only embedded sensor signals for input, excluding external equipment [4].

1.1. Landscape: EEG Data Privacy

BCI technologies that use EEG data range from advanced brain-controlled prosthetics for people with severe motor disabilities to novel real-time mood-forecasting systems for personalized entertainment and mental health coaching. All of these developments illustrate the increased use of EEG data across domains and represent a major advancement in user interface techniques [7–9].

The increasing reliance on EEG data in emerging BCI applications poses significant privacy concerns. Recent EEG data breaches have exposed vulnerabilities and threats in



Received: 4 March 2026

Revised: 4 May 2026

Accepted: 5 May 2026

Published: 11 May 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

current data protection measures, leading to identity theft, privacy violations, and unwarranted surveillance. Many researchers have shown that wearable BCI devices pose a significant risk to user privacy and could be detrimental if brain data are not adequately encrypted [7,8,10,11]. For example, unauthorized access to EEG data can expose users' mental health or emotional states to targeted advertising, or more harmful ones, such as coercion or manipulation [9,12]. Beltran et al. [13] identified several cumulative threats indicating that noise-based attacks can occur in BCI systems that rely on P300 waves. To mitigate these threats, increasing the number of electrodes used to monitor the EEG output of such systems is advantageous. Bernal et al. [14] showed that these threats can affect normal neuronal activity, demonstrating neuronal scanning and flooding attacks in a neuronal simulator.

The security landscape is further complicated by the integration of BCIs with other technologies, such as cloud computing and IoT devices. Thus, it is necessary to protect privacy at the key stages of data collection, transmission, and storage [8,10,15].

1.2. Existing Approaches

Ongoing efforts to ensure the anonymity of EEG data to preserve its utility for BCI applications include techniques such as encryption, blockchain, anonymization, federated learning, and differential privacy [7,15–19]. Despite numerous efforts, existing methods face several challenges and limitations that hinder their effectiveness in protecting sensitive EEG data.

Homomorphic encryption is a popular technique that enables computation on encrypted data without decryption [20]. Popescu et al. [21] applied it to obtain a privacy-preserving EEG data classification. However, the computational inefficiency and high latency of their method limit its applicability to real-time EEG data processing in BCI applications.

Differential privacy (DP) provides a framework for adding noise to data or queries to prevent the disclosure of individual-specific information [22]. Debie et al. [23] proposed a method using Generative Adversarial Networks (GANs) to generate synthetic EEG data that resembles real EEG data. Their GAN framework, developed using a differential privacy approach, implemented a privacy budget to balance data utility and privacy. However, their model achieved slightly lower but similar performance to that of the model trained on the original EEG data. Although DP is effective at ensuring privacy, EEG data-specific mechanisms [15,21] struggle to preserve data utility, which significantly affects the accuracy of downstream machine learning models [24,25].

Federated learning (FL) enables training a decentralized model on local devices, ensuring that raw data remain on the user's device, thus improving privacy [26]. However, FL also faces challenges related to communication overhead, model convergence, and data heterogeneity across devices. Xia et al. [15] highlighted these limitations in their review of privacy-preserving brain-computer interfaces, noting that while FL reduces the risk of data breaches, it struggles with scalability and efficiency in processing large-scale EEG datasets.

Wang et al. [27] proposed a novel privacy-preserving domain adaptation approach (PPDA) for intracranial EEG classification that addresses individual differences and privacy concerns without requiring access to source data. This method showed promise in reducing domain shifts and preserving patient privacy but relies on pseudolabeling techniques, which introduce inaccuracies that accumulate and degrade the model's performance over time, particularly with complex and noisy EEG signals.

Agarwal et al. proposed cryptographic techniques to ensure the confidentiality of multiple users [28]. Their approach relies on secure multiparty computation (SMC) [29], which ensures that no party has access to an individual's brain/EEG signal. However, this

method involves substantial effort to make it feasible on less powerful machines, which presents a major disadvantage. The complexity of implementing SMC in EEG-based BCI has been pointed out in [21,28,30], which poses a significant barrier to its widespread adoption.

Focusing on AI-driven cybersecurity solutions, Schiliro et al. [31] introduced a novel cognitive privacy technique to protect EEG data against cyber threats. To ensure the security of brain data within the BCI architecture against unauthorized credentials, a normalized correlation analysis approach was used, as also reported in [2]. Pazouki et al. [32] developed a model to mitigate common cyber threats, such as flooding and jamming, associated with the control of smart-home devices using brain implants. Their ANN-based model showed effective performance against false-data injection and scanning attacks [14].

Bidgoly et al. [33] proposed a privacy-preserving technique for BCI authentication that stores confidential user data as an EEG fingerprint, similar to a cryptographic hash. Gui et al. [34] introduced a detection mechanism that uses residual noise characteristics to detect replay attacks and input modifications. Initially, it recognizes the user using a convolutional neural network (CNN) and then classifies the replay attacks. By mixing the EEG channels, abnormalities in the communication channel due to hacking can be detected. Mezzina et al. [35] identified vulnerabilities between the BCI device and the framework and addressed them using a versatile cyber-secure architecture that is resistant to noise-based attacks.

Maiorana et al. [36] discovered a hill climbing attack that iteratively modifies EEG data until success with a BCI-based biometric system. To protect EEG features against such attacks, Wang et al. [37] proposed a system based on polynomial transformation with cosine functions that modulate the graph features and devised a template-corrupting process to provide cancelable templates for improved EEG security.

1.3. Quantum Computing's Emerging Role

Limitations of Single-Paradigm Models: Although significant progress has been made in developing privacy-preserving techniques for EEG data, existing methods face several challenges that limit their effectiveness and practicality. The dynamic nature of EEG data, characterized by its high dimensionality and temporal dependencies, poses unique challenges to preserving privacy. Many existing techniques fail to adequately address these characteristics, leading to oversimplification of the data or inadequate protection measures, especially with small datasets [38].

Quantum algorithms provide a unique advantage for their potential to implement differential privacy in a more secure manner, employing quantum noise mechanisms that naturally align with the probabilistic and unpredictable nature of quantum mechanics [10,39].

The intersection of quantum computing and privacy is an emerging field, with ongoing research aimed at addressing the unique challenges posed by integrating quantum principles into existing privacy-preserving frameworks. The application of quantum computing to BCI has been explored in hybrid models that combine classical and quantum computing elements to enhance performance and security. For example, quantum neural networks (QNNs) are being developed to improve the accuracy and robustness of EEG data processing by leveraging quantum properties to handle complex patterns and noise more effectively than traditional neural networks [39,40].

Quantum-enhanced BCIs can potentially revolutionize the way we approach brain signal processing, enabling faster and more secure data handling, which is crucial for applications such as neuroprosthetics and brain-controlled interfaces [41–43]. As technology matures, quantum-enhanced privacy techniques are expected to become increasingly viable to protect sensitive data, particularly in high-security applications such as EEG-based BCIs [41,44,45].

Limited research on the application of differential privacy, GANs, and quantum principles specifically to EEG data underscores a gap in the literature that requires further exploration of methods that can effectively protect EEG data without compromising their integrity.

1.4. Contributions: Q-DP-GAN

To address the limitations of existing approaches, we present a novel integration of quantum principles with differential privacy to reconfigure the architecture of sequential GAN (Q-DP-GAN). Our approach aims to balance privacy and utility in EEG-based BCI applications by generating synthetic data that preserve the essential characteristics of the original signals while reducing privacy risks. The term quantum-inspired describes the probabilistic state-transition behavior of the noise mechanism, particularly the exponential decay resembling decoherence and the transition to a stabilized phase analogous to measurement collapse. We do not rely on quantum hardware or quantum algorithms. Instead, the approach is designed to maintain compatibility with classical deep learning systems, avoiding the computational and hardware constraints [46–48] of quantum computing while providing a practical and scalable privacy mechanism for EEG data.

The main contributions of this study are:

- Quantum-inspired noise dynamics: Our study is among the first to integrate quantum-inspired noise dynamics into GAN training by emulating quantum decoherence. Dynamically adjustable noise not only meets but also significantly enhances differential privacy standards. By transitioning from high to low uncertainty over training epochs, this approach effectively reduces the risk of information leakage.
- Hybrid privacy-preserving mechanism: We use a stochastic gradient descent framework in the Q-DP-GAN model that combines traditional differential privacy techniques with quantum-inspired noise dynamics. This hybrid approach enables dynamic adjustment of noise scale and gradient clipping based on the model's convergence state.
- Privacy budget management: To align with the quantum uncertainty principle, we calculate the privacy budget by incorporating a decay factor into the total budget accumulated during training. As the model stabilizes, each epoch contributes less to the overall privacy budget, resulting in a more predictable and manageable budget for long-term data security.
- Privacy attacks evaluation: We evaluated our system against membership inference in both White-Box and Black-Box settings, as well as reconstruction attacks. These tests confirm the model's effectiveness in concealing data sources and preventing replication of the original training set. Notably, the model demonstrates resistance to reconstruction attacks, which, to our knowledge, have not been previously assessed on EEG data. As a result, even if synthetic EEG data is obtained, the risk of reconstructing or inferring sensitive source data remains minimal.

2. Preliminaries

2.1. Generative Adversarial Network (GAN)

Generative adversarial networks (GANs), introduced by Goodfellow et al. in 2014 [49], represent a significant innovation in machine learning by facilitating the generation of highly realistic synthetic data. GANs comprise two competing neural networks: generator (G) and discriminator (D). The generator's objective is to generate data that appear indistinguishable from real data, while the discriminator's task is to distinguish real from synthesized data. This adversarial interaction encourages both networks to continuously improve, thus generating highly realistic data.

Sequential Generative Adversarial Networks (SGANs) [50] extend the GAN architecture to handle sequential data such as time series or text. SGANs are particularly useful for generating sequences that exhibit temporal dependencies, making them suitable for applications in natural language processing and bioinformatics.

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x_{1:T})] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z_{1:T})))] \quad (1)$$

where $x_{1:T}$ denotes a sequence of data points and $G(z_{1:T})$ represents the generated sequence [50].

The selection of SGANs over other GAN architectures for modeling EEG data was motivated by their ability to handle sequences and preserve temporal dependencies. SGANs offer improved training stability and are well suited to the complexities of sequence generation, including the avoidance of mode collapse, a common challenge in traditional GAN frameworks. Thus, it provides a targeted approach to generate high-fidelity synthetic EEG data.

2.2. Differential Privacy (DP)

Differential Privacy (DP) [51] quantifies and bounds the privacy loss incurred when releasing information derived from a dataset. By ensuring that adding or removing a single data point does not significantly alter the analysis outcomes, DP provides strong privacy guarantees, which are essential for sensitive data such as EEG records. Its integration into data analysis, particularly machine learning, is now central to privacy-preserving technologies, underscoring the need to balance data utility and privacy.

Here, we define the core concepts and theorems essential for our proposed methodology.

Definition 1 (Differential Privacy). A randomized algorithm $\mathcal{M}$ with domain $\mathcal{D}$ and range $\mathcal{R}$ satisfies (ϵ, δ) -differential privacy if, for any two adjacent datasets $d, d' \in \mathcal{D}$ (differing by one element), and for all subsets $S \subseteq \mathcal{R}$,

$$\Pr[\mathcal{M}(d) \in S] \leq e^\epsilon \Pr[\mathcal{M}(d') \in S] + \delta \quad (2)$$

where ϵ represents the privacy budget and δ represents the privacy guarantee [52].

Theorem 1 (Gaussian Mechanism). For any function $f : \mathcal{D} \rightarrow \mathbb{R}^k$ with sensitivity Δf , and parameters ϵ, δ where $0 < \epsilon < 1$ and $c^2 > 2 \ln(1.25/\delta)$, the Gaussian Mechanism adds noise with standard deviation $\sigma \geq \frac{c \Delta f}{\epsilon}$ to the output of f to ensure (ϵ, δ) -differential privacy [52].

Theorem 2 (Composition Theorem). For a series of independent randomized mechanisms $\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_k$, each providing (ϵ, δ) -differential privacy, the sequence of mechanisms provides $(k\epsilon, k\delta)$ -differential privacy [53].

Theorem 3 (Enhanced Composition Theorem). For any $\epsilon, \delta, \delta' > 0$, the class of ϵ -differentially private mechanisms satisfies $(\epsilon', k\delta + \delta')$ -differential privacy under k -fold adaptive composition, where

$$\epsilon' = \sqrt{2k \ln(1/\delta')} \epsilon + k\epsilon(e^\epsilon - 1) \quad (3)$$

This theorem provides a tighter bound on cumulative privacy loss when multiple differentially private mechanisms are composed [51].

2.3. Quantum Principles

Quantum computing uses the principles of quantum mechanics to perform computations in fundamentally different ways from classical computing. Fundamental principles

include quantum bits (qubits), superposition, entanglement, quantum interference, the uncertainty principle, quantum decoherence, and quantum tunneling [54–56]. Each of these principles has profound implications for computational efficiency and data security. We specifically incorporated quantum decoherence and the quantum uncertainty principle due to their direct applicability to enhancing privacy in data synthesis.

2.3.1. Quantum Uncertainty Principle

The Heisenberg Uncertainty Principle is a fundamental concept in quantum mechanics, stating that certain pairs of physical properties, such as position (x) and momentum (p), cannot be known precisely simultaneously. This principle is mathematically expressed as

$$\Delta x \Delta p \geq \frac{\hbar}{2} \quad (4)$$

where Δx and Δp are the uncertainties in position and momentum, respectively, and $\hbar$ is the reduced Planck constant. In quantum computing, the uncertainty principle implies that measuring quantum states can disturb them. This principle is exploited in quantum cryptographic systems to detect eavesdropping, since any measurement by an eavesdropper disturbs the quantum state and reveals its presence [57].

2.3.2. Quantum Decoherence

Quantum decoherence describes the loss of quantum coherence as a quantum system interacts with its environment, leading it to transition from a coherent superposition state to an incoherent classical mixture. This process can be described by the evolution of the density matrix ρ of the quantum state, given by the Lindblad master equation.

$$\frac{d\rho}{dt} = -\frac{i}{\hbar}[H, \rho] + \sum_n \left(\zeta_n L_n \rho L_n^\dagger - \frac{1}{2} \{L_n^\dagger L_n, \rho\} \right) \quad (5)$$

where H is the Hamiltonian of the system, L_n are the Lindblad operators representing interactions with the environment, and $\{\cdot, \cdot\}$ denotes the anticommutator. Decoherence is a major challenge in quantum computing, as it can destroy delicate quantum states required for computation. However, understanding and controlling decoherence are key to developing quantum error-correction techniques [58].

3. Methodology

3.1. Hybrid Architecture

The Q-DP-GAN model combines differential privacy, implemented through DP-SGD, with dynamically adjusted privacy parameters in a generative adversarial network (GAN) for EEG data synthesis. While the approach is conceptually quantum-inspired, its mathematical implementation is classical, relying on adaptive noise and clipping mechanisms. This integration is designed to achieve strong privacy guarantees while maintaining the utility of the generated EEG data. Our methodology is outlined in Algorithm 1 and depicted in Figure 1.

3.1.1. Generator and Discriminator Architecture

The generator (G) and the discriminator (D) form the backbone of our model, where the generator is tasked with generating synthetic EEG data, and the discriminator is responsible for validating its authenticity.

Generator Architecture: The generator is designed to convert a latent noise vector into synthetic EEG data that resemble the real EEG recordings in both structure and dynamics. To capture the temporal and sequential nature of the EEG signals, the generator

employs stacked GRU [59–61] layers that model dependencies across time steps. The latent vector is first transformed into a structured sequence representation, enabling recurrent processing. This design allows the model to generate temporally coherent multi-channel EEG signals while preserving inter-channel relationships. The generator function $G(\mathbf{z}; \theta_G)$ maps the latent space vector $\mathbf{z}$ to the data space, where θ_G denotes the generator parameters. The output is activated by a function $\tanh$, to ensure that the output matches the amplitude characteristics typical of EEG signals. The generator architecture is designed as shown in Table 1.

Algorithm 1 Training Q-DP-GAN with dynamic privacy

```

1: Input: EEG dataset, total number of epochs  $\text{Epoch}_{\text{total}}$ , initial and final privacy parameters  $(\sigma_{\text{initial}}, C_{\text{initial}}, \sigma_{\text{final}}, C_{\text{final}})$ , transition threshold  $(\theta)$ , number of stable epochs required  $(n)$ , learning rates for discriminator  $(\text{lr}_D)$  and generator  $(\text{lr}_G)$ 
2: Output: Trained GAN model with privacy guarantees
3: Initialize parameters: Generator parameters  $\theta_G$ , Discriminator parameters  $\theta_D$ 
4: Initialize  $\sigma_{\text{current}} = \sigma_{\text{initial}}, C_{\text{current}} = C_{\text{initial}}$ 
5: Initialize  $\lambda$  (decay rate for  $\sigma$ ),  $\beta$  (decay rate for  $C$ )
6: Initialize  $\text{loss\_stabilization\_counter} = 0, \text{previous\_loss} = \infty$ 
7: for each epoch  $t = 1$  to  $\text{Epoch}_{\text{total}}$  do
8:   Shuffle and batch the dataset
9:   for each batch do
10:     Sample noise vector  $\mathbf{z}$  from a normal distribution
11:     Generate synthetic data  $G(\mathbf{z}; \theta_G)$  using generator
12:     Train discriminator ( $D$ ) on both real and generated data:
13:       Compute discriminator loss  $L_D$  on real data and  $G(\mathbf{z}; \theta_G)$ 
14:       Clip gradients of  $L_D$  by norm  $C_{\text{current}}$ 
15:       Add Gaussian noise  $\sim N(0, \sigma_{\text{current}}^2)$  to the gradients
16:       Update  $\theta_D$  using gradient descent with learning rate  $\text{lr}_D$ 
17:     Train generator ( $G$ ) to fool discriminator:
18:       Compute generator loss  $L_G$  using  $D$ 's response to  $G(\mathbf{z}; \theta_G)$ 
19:       Update  $\theta_G$  using gradient descent with learning rate  $\text{lr}_G$ 
20:   end for
21:   Compute and record the discriminator loss for this epoch ( $\text{current\_loss}$ )
22:   Assess model stability:
23:   if  $|\text{previous\_loss} - \text{current\_loss}| < \theta$  then
24:      $\text{loss\_stabilization\_counter} += 1$ 
25:   else
26:      $\text{loss\_stabilization\_counter} = 0$ 
27:   end if
28:   if  $\text{loss\_stabilization\_counter} \geq n$  then
29:     Transition to static privacy parameters:
30:      $\sigma_{\text{current}} = \sigma_{\text{final}}$ 
31:      $C_{\text{current}} = C_{\text{final}}$ 
32:   end if
33:   Update  $\text{previous\_loss}$ :
34:    $\text{previous\_loss} = \text{current\_loss}$ 
35:   if not yet transitioned then
36:     Adjust privacy parameters dynamically:
37:      $\sigma_{\text{current}}^* = e^{-\lambda}$ 
38:      $C_{\text{current}}^* = e^{-\beta}$ 
39:   end if
40: end for
41: Return: trained GAN model  $(\theta_G, \theta_D)$ 

```

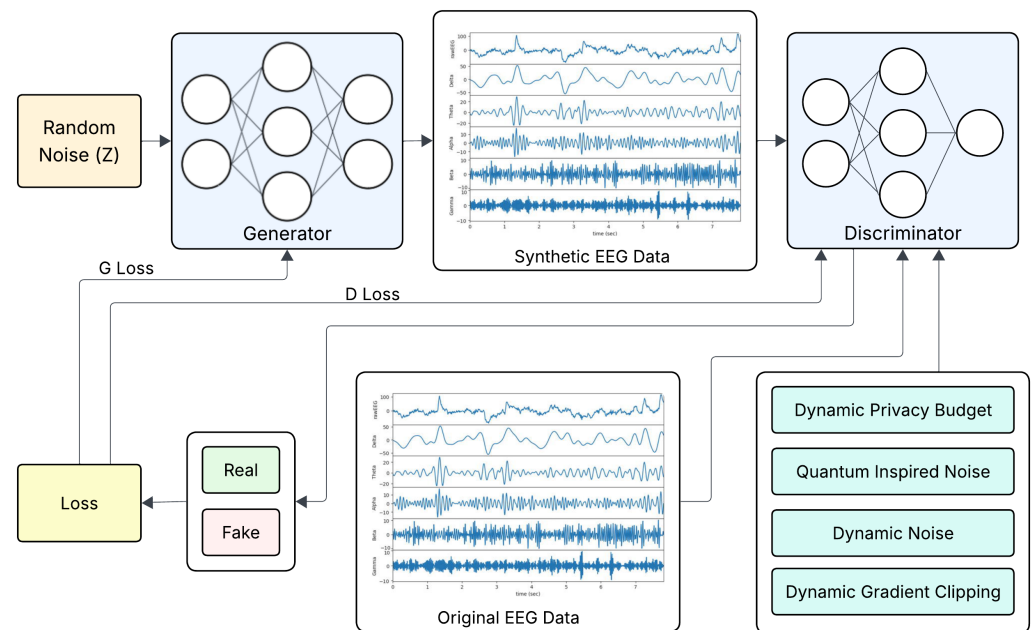


Figure 1. System architecture of the proposed Q-DP-GAN (quantum-inspired differential privacy) model.

Table 1. Generator architecture details of our proposed model.

Block	Layer Name	Units/Size	Activation	Output Shape
1	Input (latent vector)	-	-	(100, 1)
2	Dense	128	LeakyReLU	(128, 1)
3	Reshape	-	-	(64, 2, 1)
4	GRU Layer 1	256	Tanh	(64, 256)
5	GRU Layer 2	64	Tanh	(64, 64)
6	Flatten	-	-	(4096,)
7	Dense	3000	ReLU	(3000,)
8	Reshape/Output Layer	-	Tanh	(3, 1000)

Discriminator Architecture: The discriminator assesses the authenticity of EEG data, classifying them as either real or generated. It uses convolutional layers to extract features from the input, which are then used to compute the probability that the data are real. The function $D(\mathbf{x}; \theta_D)$ evaluates the input $\mathbf{x}$, with θ_D representing the discriminator parameters. A sigmoid function at the output layer provides a probabilistic estimate of authenticity. The discriminator architecture is designed as shown in Table 2.

Table 2. Discriminator architecture details of our proposed model.

Block	Layer Name	Units/Size	Kernel Size	Activation	Purpose	Additional Info	Output Shape
1	Input	-	-	-	Accepts EEG data matrices as input.	Input layer to receive data.	(3, 1000)
2	Reshape	-	-	-	Reshapes input to 3D for compatibility with convolutional layers.	Prepares data for convolutional processing.	(1, 3, 1000)
3	2D Convolution	10	(1, 10)	ReLU	Applies 2D convolution across time.	Initial feature extraction.	(10, 3, 991)
4	2D Convolution	10	(3, 1)	ReLU	Further extracts detailed features across channels.	Enhances channel integration.	(10, 1, 991)

Table 2. Cont.

Block	Layer Name	Units/Size	Kernel Size	Activation	Purpose	Additional Info	Output Shape
5	Dropout	-	-	-	Adds regularization to prevent overfitting.	Dropout rate: 0.25	(10, 1, 991)
6	Batch Normalization	-	-	-	Normalizes activations to improve convergence.	Stabilizes the learning process.	(10, 1, 991)
7	Max Pooling 2D	-	(1, 2)	-	Reduces temporal dimension moderately.	Downsampling to reduce feature dimension.	(10, 1, 495)
8	2D Convolution	16	(1,10)	ReLU	Deeper feature extraction across time.	Processes deeper temporal patterns.	(16, 1, 486)
9	Max Pooling 2D	-	(1, 3)	-	Further reduces temporal dimension more aggressively.	Additional downsampling.	(16, 1, 162)
10	2D Convolution	32	(1, 10)	ReLU	Increases the depth of feature extraction.	Captures complex temporal patterns.	(32, 1, 153)
11	Max Pooling 2D	-	(1, 2)	-	Reduces temporal dimension significantly.	Balances feature retention and reduction.	(32, 1, 76)
12	2D Convolution	64	(1, 10)	ReLU	Extracts higher-level features.	Captures highest-level temporal features.	(64, 1, 67)
13	Dropout	-	-	-	Adds regularization to further prevent overfitting.	Dropout rate: 0.25	(64, 1, 67)
14	Batch Normalization	-	-	-	Ensures stable feature normalization.	Ensures consistency in feature scaling.	(64, 1, 67)
15	Bidirectional LSTM	64	-	Tanh	Captures long-term dependencies from both directions.	Processes sequence data bidirectionally.	(64, 67)
16	Bidirectional LSTM	128	-	Tanh	Enhances the model's ability to detect complex patterns.	Provides depth to temporal data processing.	(128, 67)
17	Flatten	-	-	-	Flattens the output to prepare for dense layer.	Converts multidimensional data to 1D.	(8704,)
18	Dense	64	-	ReLU	Processes the flattened feature vector.	Final processing before output.	(64,)
19	Dropout	-	-	-	Helps reduce overfitting in the final stages.	Dropout rate: 0.5	(64,)
20	Dense (Output Layer)	1	-	Sigmoid	Outputs the probability that the input data are real.	Produces a probability output.	(1,)

3.1.2. Noise Dynamics

The model initially applies variable high-intensity noise that is adaptively reduced over time as training progresses. This behavior can be viewed as conceptually analogous to a reduction in uncertainty in quantum measurement, although the implementation itself is entirely classical and is based on standard adaptive decay. The mathematical formulation of the noise dynamics is given by

$$\sigma(t) = \sigma_{\text{initial}} e^{-\lambda t} \quad (6)$$

$$C(t) = C_{\text{initial}} e^{-\beta t} \quad (7)$$

where $\sigma(t)$ represents the standard deviation of the noise applied to the gradients at epoch t . This noise is dynamically adjusted, decreasing as t increases from 1 to $\text{epochs}_{\text{total}}$. This

decrease simulates the transition from a quantum state (characterized by high uncertainty) to a more classical state (characterized by reduced uncertainty). Similarly, $C(t)$ denotes the clipping threshold for the gradients at each epoch t . It is also dynamically adjusted, ensuring that the influence of any individual training sample remains bounded throughout the training, thereby mitigating the risk of revealing individual contributions in the synthesized data. Here, λ and β are the decay rates. We conducted several trials, varying λ and β to assess how these parameter settings affected the model's accuracy and privacy, and to identify the optimal values. In this study, λ and β were set to 0.05 and 0.03, respectively, as these values showed the right balance between privacy and accuracy (Section 4.3).

We follow the standard definition of (ϵ, δ) -differential privacy introduced in Section 2.2. Our approach is based on the Gaussian Mechanism and its application in differentially private stochastic gradient descent (DP-SGD) [53,62].

Let g_t denote the per-sample gradient of the discriminator at iteration t . To enforce privacy, gradients are clipped and perturbed as follows:

$$\tilde{g}_t = \text{Clip}(g_t, C(t)) + \mathcal{N}(0, \sigma(t)^2 I) \quad (8)$$

where $C(t)$ is the clipping norm and $\sigma(t)$ is the noise scale.

After clipping, the gradient satisfies

$$\|g_t\|_2 \leq C(t), \quad (9)$$

which bounds the ℓ_2 -sensitivity of the mechanism by $C(t)$.

To ensure sustained privacy protection as the model stabilizes, the proposed mechanism maintains a baseline level of uncertainty, thereby limiting the ability to precisely infer individual data contributions. Once the transition criterion is satisfied, the privacy parameters are kept constant and are defined as follows:

$$\sigma = \sigma_{\text{final}} \quad (10)$$

$$C = C_{\text{final}} \quad (11)$$

These parameters remain unchanged after the transition (described in the following section). By maintaining a steady noise level and minimizing the influence of individual data points, they reliably ensure data privacy.

Since $\sigma_{\text{final}} > 0$, noise continues to be applied to gradients. Therefore, the mechanism remains differentially private. Unlike learning rate scheduling, which affects optimization dynamics, this transition directly modifies noise scale and clipping bounds, which govern gradient sensitivity and perturbation. Hence, it operates within the differential privacy mechanism and does not violate the differential privacy guarantees.

3.2. Transition Algorithm and Privacy Budget Computation

3.2.1. Transition Algorithm

The model employs an algorithmic approach to determine when to switch from dynamic to static privacy parameters based on the stabilization of the model's performance.

- if $\Delta\text{Loss} < \theta$ for consecutive n epochs, then transition to the static phase.
This condition checks if the change in loss ΔLoss falls below a threshold θ over n epochs, indicating that the model has reached a sufficiently stable state and requires reduced dynamism in privacy controls.

To ensure effective execution of these parameters, ΔLoss is calculated as follows:

$$\Delta\text{Loss} = \frac{1}{n} \sum_{i=t-n+1}^t |\text{Loss}_i - \text{Loss}_{i-1}| \quad (12)$$

This formula computes the average absolute change in loss over n epochs, providing a smooth and representative measure of loss changes during this period. It effectively captures both the magnitude and the consistency of loss changes, providing a reliable trigger for the parameter transition.

3.2.2. Privacy Budget Computation

To incorporate quantum noise while ensuring consistency with the principles of differential privacy, the privacy budget [63–65] at a given epoch t is modified as follows:

$$\epsilon(t) = \sqrt{2 \log(1.25/\delta)} \cdot \sum_{i=1}^t \omega(i) \cdot \frac{q \cdot C(i)}{\sigma(i)}, \quad (13)$$

where $w(i)$ is a weighting function defined as $w(i) = e^{-\gamma i}$, with γ representing the decay rate. This rate governs how quickly the influence of each epoch reduces, highlighting the diminishing privacy risk as the model stabilizes. Here, $\gamma = 2\lambda\beta/(\lambda + \beta)$. $\sigma(i)$ and $C(i)$ are dynamically adjusted noise and clipping parameters, respectively, up to the stabilization point, after which they remain constant, and q represents the proportion of the dataset used in each training batch.

3.2.3. Validity of Dynamic Privacy Budget

Each iteration i satisfies the (ϵ_i, δ) -differential privacy through the Gaussian Mechanism. By the composition property of differential privacy [53], the total privacy loss is

$$\epsilon_{\text{total}} = \sum_{i=1}^t \epsilon_i. \quad (14)$$

The proposed formulation introduces a weighting function $\omega(i) \in (0, 1]$, leading to

$$\epsilon_{\text{weighted}} = \sum_{i=1}^t \omega(i) \epsilon_i. \quad (15)$$

Since $\omega(i) \leq 1$ for all i , we have

$$\epsilon_{\text{weighted}} \leq \epsilon_{\text{total}}. \quad (16)$$

Furthermore, both ϵ_i and $\omega(i)$ decay exponentially, ensuring that the series

$$\sum_{i=1}^t \omega(i) \epsilon_i \quad (17)$$

is bounded and convergent. Therefore, the proposed weighted privacy budget does not violate standard differential privacy guarantees.

3.2.4. Privacy-Utility Trade-Off via Dynamic Allocation

The proposed adaptive mechanism improves the privacy-utility trade-off by redistributing privacy across training iterations while maintaining the same overall (ϵ, δ) guarantee.

In standard DP-SGD, the privacy loss accumulates uniformly across iterations as follows:

$$\epsilon_{\text{static}} \propto \sum_{t=1}^T \frac{C}{\sigma}. \quad (18)$$

In the proposed method,

$$\epsilon_{\text{adaptive}} \propto \sum_{t=1}^T \frac{C(t)}{\sigma(t)}. \quad (19)$$

Substituting exponential decay,

$$\epsilon_{\text{adaptive}} \propto \sum_{t=1}^T e^{-(\beta-\lambda)t} \quad (20)$$

Although the rate of decay depends on the relative values of β and λ , the overall privacy guarantee remains governed by the composition property of differential privacy. Specifically, the cumulative privacy budget over T iterations is

$$\epsilon_{\text{total}} = \sum_{t=1}^T \epsilon_t \quad (21)$$

which remains finite for any finite number of training iterations T . Therefore, the proposed mechanism preserves standard differential privacy guarantees while adaptively redistributing the privacy budget across training.

3.2.5. Effect of Dynamic Noise and Clipping

In the proposed framework, the noise and clipping evolve as follows:

$$\sigma(t) = \sigma_{\text{initial}} e^{-\lambda t}, \quad C(t) = C_{\text{initial}} e^{-\beta t}. \quad (22)$$

Substituting into the privacy expression,

$$\frac{C(t)}{\sigma(t)} = \frac{C_{\text{initial}}}{\sigma_{\text{initial}}} \cdot e^{-(\beta-\lambda)t}. \quad (23)$$

Thus, the per-iteration privacy loss evolves as

$$\epsilon_t \propto e^{-(\beta-\lambda)t}. \quad (24)$$

This shows that privacy is not uniformly distributed across iterations. Instead, early iterations inject higher noise (lower signal-to-noise ratio), thereby reducing the risk of memorization, whereas later iterations allow finer updates to improve utility. Importantly, the proposed method does not modify the underlying Gaussian noise distribution but instead dynamically adjusts its magnitude over time. Therefore, the mechanism remains fully compatible with the standard differential privacy theory.

4. Experimental Setup

4.1. Dataset Description

To improve our model's reliability, robustness, and generalizability, we used two benchmark datasets: BCI Competition IV 2008 2A [66] and 2B [67]. These widely used seminal EEG-based BCI datasets [68,69] enable reproducible and comparable evaluation. They contain multi-subject motor-imagery EEG data with rich temporal and spectral features. Because EEG intrinsically encodes subject-specific patterns [70,71], it can reveal identity and sensitive neurological traits. Therefore, these datasets provide a realistic and

challenging setting to test whether our framework can generate synthetic EEG data that preserve task utility while reducing privacy risks in real-world BCI applications.

Dataset 2A contains EEG recordings of nine participants across four motor imagery tasks involving movements of the left hand, right hand, both feet, and tongue, spanning multiple sessions. These sessions are meticulously structured, offering insight into neural patterns associated with motor tasks. Dataset 2B also contains EEG data from 9 subjects performing left- and right-hand movements, with and without feedback mechanisms, unlike Dataset 2A. A detailed comparison of these two datasets is shown in Table 3.

Table 3. Comparison of BCI Competition IV datasets, 2A and 2B.

Attribute	Details for BCI-2A	Details for BCI-2B
Dataset Origin	Graz University of Technology, 2008	Graz University of Technology, 2008
Subjects	9 healthy subjects	9 right-handed subjects with normal or corrected vision
Motor Imagery Tasks	4 tasks: left hand, right hand, both feet, tongue	2 tasks: left hand, right hand
Number of Sessions	2 per subject	5 per subject
Runs per Session	6 runs	Screening: 6 runs, Feedback: 4 runs
Trials per Run	48 trials (12 per task)	Screening: 20 trials (10 per task), Feedback: 40 trials (20 per task)
Total Trials per Session	288 trials	Screening: 120 trials, Feedback: 160 trials
Total Trials per Subject	576 trials (2 sessions $\times$ 288 trials)	720 trials per subject (240 from screening, 480 from feedback)
EEG Electrodes	22 electrodes	3 bipolar electrodes (C3, Cz, C4)
EOG Channels	3 monopolar electrodes, for artifact processing	3 monopolar electrodes, for artifact processing
Sampling Rate	250 Hz	250 Hz
Filtering	Bandpass 0.5–100 Hz, 50 Hz notch filter	Bandpass 0.5–100 Hz, 50 Hz notch filter
Feedback Provided	No	Yes, in last 3 sessions
Artifact Handling	Expert visual inspection, trials with artifacts marked	Trials containing artifacts marked, required artifact removal
Data File Format	GDF (General Data Format)	GDF (General Data Format)
File Distribution	Training and evaluation sets separate	First 3 sessions for training, last 2 sessions for evaluation
Data Access Software	BioSig toolbox (Octave 8.4/FreeMat 4.2/MATLAB R2024a, C/C++)	BioSig toolbox (Octave 8.4/MATLAB R2024a, C/C++)

4.2. Data Preprocessing

For Dataset 2A, our focus was on three channels, C3, Cz, and C4, which efficiently capture brain patterns associated with the identification of imagined movement states [72]. Among the four initial motor imagery tasks, the movements of the feet and tongue were omitted, and the study focused solely on the movements of the left and right hands. The dataset contains artifacts, and Table 4 shows the accepted and rejected trials for Sessions I and II of all subjects. Session I was used as the train set, and Session II as the test set.

For Dataset 2B, we used all three EEG channels (C3, Cz, and C4) and tasks (left- and right-hand movements). Artifacts are also present in this dataset. Table 5 shows the accepted and rejected trials for the five sessions. Sessions I, II, and III were used as training sets, and the remaining sessions as test sets.

Table 4. Accepted and rejected/artifact trials by subject and task of Dataset 2A.

Dataset: 2A Session I (Train Set)										
Task	Status	A1	A2	A3	A4	A5	A6	A7	A8	A9
Left Hand	Accepted	69	67	69	62	63	56	67	66	53
	Rejected	3	5	3	10	9	16	5	6	19
Right Hand	Accepted	69	69	68	67	66	57	66	66	63
	Rejected	3	3	4	5	6	15	6	6	9
Dataset: 2A Session II (Test Set)										
Left Hand	Accepted	71	71	67	59	70	59	71	66	65
	Rejected	1	1	5	13	2	13	1	6	7
Right Hand	Accepted	70	71	70	57	65	55	69	68	65
	Rejected	2	1	2	15	7	17	3	4	7

Table 5. Accepted and rejected/artifact trials by subject and task of Dataset 2B.

Dataset: 2B Session I (Train Set)										
Task	Status	B1	B2	B3	B4	B5	B6	B7	B8	B9
Left Hand	Accepted	51	48	47	57	51	35	55	52	43
	Rejected	9	12	13	3	9	25	5	8	17
Right Hand	Accepted	51	50	43	60	52	41	52	40	48
	Rejected	9	10	17	0	8	19	8	20	12
Dataset: 2B Session II (Train Set)										
Left Hand	Accepted	48	51	43	56	58	42	55	49	45
	Rejected	12	9	17	4	2	18	5	11	15
Right Hand	Accepted	50	49	45	53	58	43	56	41	46
	Rejected	10	11	15	7	2	17	4	19	14
Dataset: 2B Session III (Train Set)										
Left Hand	Accepted	60	64	56	78	74	65	68	54	69
	Rejected	20	16	24	2	6	15	12	26	11
Right Hand	Accepted	64	67	62	75	69	69	70	52	66
	Rejected	16	13	18	5	11	11	10	28	14
Dataset: 2B Session IV (Test Set)										
Left Hand	Accepted	51	51	62	78	79	71	51	69	56
	Rejected	29	29	18	2	1	9	29	11	24
Right Hand	Accepted	61	51	55	75	77	66	56	57	59
	Rejected	19	29	25	5	3	14	24	23	21
Dataset: 2B Session V (Test Set)										
Left Hand	Accepted	58	71	60	76	55	52	63	50	65
	Rejected	22	9	20	4	25	28	17	30	15
Right Hand	Accepted	58	72	53	78	62	62	62	54	65
	Rejected	22	8	27	2	18	18	18	26	15

In the preprocessing stage of our study, an initial analysis was performed using 7.5 s trials to capture a broad spectrum of neural activity. To improve model training efficiency and optimize computational resource use, we systematically evaluated the efficacy of

shorter time windows. Our methodological approach involved segmenting the data into 6- and 4 s intervals. This segmentation was designed to identify the most critical time windows for motor-imagery classification. Specifically, the 6 s segments were obtained by excluding the first 1 s and the last 0.5 s of the trials, reducing the data from 1875 (7.5×250) data points to 1500 (6×250) data points. In addition, we tested 4 s segments in which the initial 2 s and the final 1.5 s were removed, yielding 1000 data points. These truncated segments were analyzed to determine if they contained sufficient information for accurate classification.

Our initial findings validated the 4 s window, as it yielded a classification accuracy comparable to that of longer durations. Previous work with the BCI Competition IV dataset also shows that temporal window selection affects performance. Studies often use windows after the cue onset, such as 0.5–2.5 s [69] or later windows, such as 2.5–4.5 s [73,74], supporting the exclusion of early segments. This indicates that the key neural signatures of the motor imagery tasks are within this interval. Using these segments, we achieved improved model training and real-time processing efficiency without compromising classification quality. Thus, the 4 s window was chosen for subsequent analysis, balancing computational efficiency and empirical effectiveness.

4.3. Model Training: Details, Hyperparameters, and Configurations

The selection of the appropriate parameter values is crucial to balance the effectiveness of the privacy mechanism and the learning capacity of the model. The following paragraphs describe the parameters of the hybrid quantum-inspired model, which were empirically determined through extensive experimental trials.

4.3.1. Initial Phase (Quantum Decoherence)

We chose the initial noise standard deviation (σ_{initial}) as 2.0. The chosen value provides robust privacy protection at the beginning of the training process, when the model is most susceptible to information leakage. A higher initial noise level effectively obscures the contributions of individual training samples, thereby reducing the risk of overfitting to specific data points, which could compromise privacy.

We have set the initial clipping threshold (C_{initial}) to 1.0, a value determined through experimentation. Beginning with a higher clipping threshold also ensures that the impact of any single data point remains constrained during the early stages of training. This configuration helps mitigate the undue influence of outliers on the model's learning process, thereby enhancing privacy protection.

4.3.2. Decay Rates (λ for Noise and β for Clipping)

Similarly, following our experimental testing, we set decay rates, λ to 0.05 and β to 0.03. The decay rate of 0.05 for noise ensures that the model gradually transitions to less noisy updates as it stabilizes, balancing privacy protection with learning efficiency. The slightly lower decay rate of 0.03 for clipping helps prevent the model from becoming overly sensitive to individual samples too quickly, thereby maintaining a conservative privacy approach for a longer duration during training.

4.3.3. Stabilization Phase (Quantum Uncertainty)

We have defined the final noise standard deviation (σ_{final}) as 0.5. It provides sufficient privacy protection while the model stabilizes without significantly impairing its fine-tuning capabilities. The reduced noise level prevents precise inferences about individual data points while allowing the GAN to generate high-quality synthetic EEG data.

We have set the final clipping threshold (C_{final}) to 0.1. This lower threshold refines the model's sensitivity to the training data during stabilization. It improves the quality of

generated outputs by limiting the influence of individual training samples without overly constraining model learning.

4.3.4. Privacy Guarantee (δ)

We set δ at $1/N$, where N is the total number of data points. For our 4 s data window, N is 1000 (4×250).

4.3.5. Privacy Budget (ϵ)

For our method, the final ϵ values for Dataset 2A were 4.291, 4.365, 4.246, 4.381, 4.309, 4.226, 4.317, 4.403, and 4.192 for Subjects A1 through A9, respectively. For Dataset 2B, the final values for Subjects B1 through B9 were 4.457, 4.324, 4.355, 4.410, 4.493, 4.445, 4.397, 4.382, and 4.363, respectively.

4.3.6. Total Epochs and Number of Consecutive Epochs (n)

We set $\text{Epoch}_{\text{total}}$ as 1000. For a total of 1000 epochs, n should be large enough to ensure that the detection of model stabilization is robust against normal fluctuations in training loss but not so large that it delays the transition unduly. A reasonable choice for n is about 1% to 2% of $\text{Epoch}_{\text{total}}$, which translates to 10 to 20 epochs. This range typically offers a good balance by smoothing out regular variations in loss while being responsive enough to changes, indicating true model stabilization. Setting $n = 20$ allowed the model to demonstrate consistent stabilization over a sufficient period to mitigate the effects of short-term fluctuations but is short enough to ensure timely adaptation to stabilized conditions.

4.3.7. Loss Stabilization Threshold (θ)

Considering a typical GAN loss scale, values such as $\theta = 0.0001$ to 0.1 are effective in ensuring model stabilization. After initial experiments, we chose $\theta = 0.005$. This value guarantees that reductions in privacy parameters are triggered only by stabilization during training. Thus, decrements are aligned with the model's steady phase.

5. Results and Discussion

5.1. Classification Performance: Varying Original (Real) and Synthetic Data

We evaluated how different ratios of original (real/train) and synthetic EEG data affect classification performance, which was measured only on real EEG samples from the test set. Dataset 2A used session I as original training data, and Dataset 2B used sessions I–III. We trained four neural networks—ATCNet [75], EEGNet [76], ShallowNet [61], and CapsNet [77]—to recognize hand movements using (1) only original EEG data and (2) mixed original and synthetic EEG data. As illustrated in Figure 2, the experiment examined how these data proportions affect model performance.

We tested five specific ratios of original to synthetic data for training the model as follows:

- 100% original data and 0% synthetic data. This model was trained exclusively on the original training dataset (Tr) for optimal comparison with augmented models.
- 50% original data and 50% synthetic data. This model used a balanced combination of 50% original training data and 50% synthetic data (S) to augment the training set while maintaining the total number of samples constant.
- 30% original data and 70% synthetic data.
- 20% original data and 80% synthetic data.
- 10% original data and 90% synthetic data.

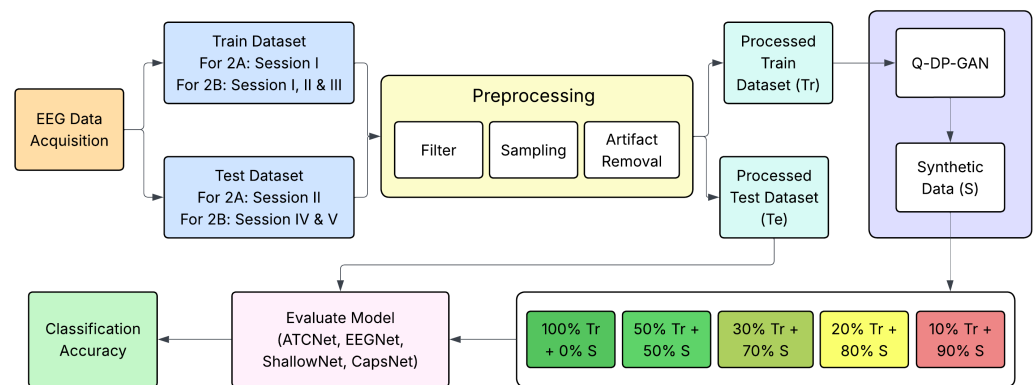


Figure 2. Overview of our experiments using original and synthetic data. The number of synthetic data samples (S) produced was equal to the number of original training samples (Tr). We then used varying proportions of Tr and S, merging them into a training set for further analysis.

We evaluated each model's classification performance over 20 runs to ensure reliability and robustness. Training used the datasets defined in the original publication (Dataset 2A: Session I; Dataset 2B: Sessions I–III), and testing used the remaining sessions. We applied our method to generate synthetic data, trained four neural networks (ATCNet, EEGNet, ShallowNet, and CapsNet) with different mixtures of original and synthetic training data, and then tested them on the original test set (real EEG samples). We selected $k\%$ of the original training data and $l\%$ of the synthetic data, where $k = 100, 50, 30, 20$, or 10 and $l = 0, 50, 70, 80$, or 90 . Because sample sizes differed across subjects and datasets, we rounded the resulting sample counts to the nearest integer using a threshold of 0.50.

The results of this experiment are illustrated in Tables 6 and 7, respectively, for Subject A1 of Dataset 2A and Subject B1 of Dataset 2B. The remaining tables for other subjects are provided in the Supplementary File. The changes in model accuracy were more noticeable when the percentage of synthetic data was progressively increased, as discussed below:

- Models trained on 100% original data performed similarly to those trained on 50% original data and 50% synthetic data, with the accuracy dropping by only 1–2.5% on the test data.
- The model's accuracy decreased by 2.1–4.1% when 30% of the original data were mixed with 70% synthetic data.
- A ratio of 20% of original data mixed with 80% synthetic data resulted in 3.2–5.6% accuracy drops.
- The greatest change was observed using 10% of the original data mixed with 90% synthetic data, resulting in 5.1–7.2% change in model accuracy.

Table 6. Classification performance of BCI Dataset IV 2A using various training scenarios for Subject A1. All models are evaluated on test data using 100% of real EEG data samples composed of session II.

Training Scenario (Trained on)	ATCNet		EEGNet		ShallowNet		CapsNet	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
100% original training data (Tr)	86.11	1.18	85.43	1.67	83.66	2.09	84.33	1.95
50% original training data (Tr) and 50% synthetic data (S)	85.12	1.93	83.91	2.17	81.55	2.78	83.03	2.61
30% original training data (Tr) and 70% synthetic data (S)	84.30	2.32	82.61	3.02	79.79	3.44	80.55	4.23
20% original training data (Tr) and 80% synthetic data (S)	83.84	3.82	83.97	2.76	78.11	2.93	79.80	3.45
10% original training data (Tr) and 90% synthetic data (S)	81.12	3.74	82.89	4.05	76.64	3.67	78.47	4.38

Table 7. Classification performance of BCI Dataset IV 2B using various training scenarios for Subject B1. All models are evaluated on test data using 100% of real EEG data samples composed of sessions IV and V.

Training Scenario (Trained on)	ATCNet		EEGNet		ShallowNet		CapsNet	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
100% original training data (Tr)	85.43	1.83	83.10	1.27	78.28	2.41	79.72	2.05
50% original training data (Tr) and 50% synthetic data (S)	83.92	2.13	81.64	1.92	75.94	3.14	78.31	2.47
30% original training data (Tr) and 70% synthetic data (S)	82.77	3.12	80.69	2.68	75.13	2.91	77.18	3.55
20% original training data (Tr) and 80% synthetic data (S)	82.04	2.95	79.82	3.74	74.46	3.61	76.39	2.95
10% original training data (Tr) and 90% synthetic data (S)	81.65	3.47	78.58	4.11	73.64	3.02	75.00	3.86

These findings suggest that a balanced mix of 50% original and 50% synthetic data effectively augments the EEG datasets while maintaining robust model performance. Other configurations, such as 30% or 20% original data with synthetic data, are also viable, albeit with slightly larger accuracy drops. The correct balance between the original and synthetic data depends on the error-rate tolerance of the BCI applications. Our findings show that varying proportions of synthetic EEG data are helpful to achieve comparable results in training processes that use real EEG data.

5.2. Privacy Analysis: Attack Scenarios

After evaluating classification performance, we simulated two attack scenarios to examine privacy preservation. We then compared our method with five state-of-the-art methods: WGAN [78], CGAN [79], DPGAN [80], ST-DP-GAN [81], and LDPGAN [82].

We ensured that each model was configured with the ϵ values derived from our method, so that each model maintained the same privacy budget for each subject, allowing a fair comparison across datasets. For each method, synthetic data were generated in quantities equivalent to the original training dataset and then exposed to the two attack scenarios.

5.2.1. Membership Inference Attacks

To assess the privacy-preserving capabilities of our Q-DP-GAN, we implemented membership inference attacks (MIAs) under Black-Box (BB) and White-Box (WB) conditions [38,83]. These attacks test whether an attacker can determine if specific data samples were in the model's training set. Three data subsets were used to evaluate the model's ability to protect EEG data against inference threats.

- *S*—Synthetic data generated by different models.
- *Tr*—Real training data used to train different models (Dataset 2A session I and Dataset 2B sessions I–III).
- *Te*—Real data not used in training, serving as test data (Dataset 2A session II and Dataset 2B sessions IV–V).

We augment the full training set (*Tr*) with an equal amount of synthetic data (*S*), doubling the training size. Model performance is then evaluated on the original test set (*Te*) to measure generalization to unseen data.

MIAs are conducted in two main modes based on the attacker's access level as follows:

- *Black-Box Membership Inference Attacks:* The attacker does not have access to Q-DP-GAN's internals. Black-Box (BB) scenarios simulate an external adversary using shadow models trained on various subsets of *S* and *Tr*, mimicking the target

model's behavior without accessing its actual parameters. The BB attack is detailed in Algorithm 2.

- *White-Box Membership Inference Attacks*: In this scenario, the adversary uses complete information about Tr and the structure of Q-DP-GAN to examine the influence of training data on the model's outputs for Te and S . It can reveal vulnerabilities in the model's processing and learning from Tr . The WB attack is outlined in Algorithm 3.

Algorithm 2 Black-Box Membership Inference Attack

Require: Real training data Tr , Synthetic data S , Real test data Te , Number of shadow models N_s

Ensure: Attack model to predict membership inference

- 1: Combine Tr and S to create $T_{combined}$
 - 2: Split $T_{combined}$ into N_s subsets: $\{T_{combined_1}, T_{combined_2}, \dots, T_{combined_{N_s}}\}$
 - 3: **for** each subset $T_{combined_i}$ ($i = 1$ to N_s) **do**
 - 4: Train shadow model S_i on $T_{combined_i}$
 - 5: **end for**
 - 6: Generate Predictions:
 - 7: **for** each shadow model S_i ($i = 1$ to N_s) **do**
 - 8: Predict probabilities on Te : $P_i^{test} = S_i.predict(Te)$
 - 9: Predict probabilities on S : $P_i^{syn} = S_i.predict(S)$
 - 10: **end for**
 - 11: Create and Label Attack Dataset:
 - 12: Combine all P_i^{test} to form X_{test} $\triangleright$ X_{test} contains the predictions for the real test data
 - 13: Combine all P_i^{syn} to form X_{syn} $\triangleright$ X_{syn} contains the predictions for the synthetic data
 - 14: Label X_{test} with 1s (real) and X_{syn} with 0s (synthetic)
 - 15: Form attack dataset $X_{attack} = \{X_{test}, X_{syn}\}$ and labels $Y_{attack} = \{1s, 0s\}$
 - 16: Train Attack Model:
 - 17: Split X_{attack} into training and testing sets
 - 18: Train binary classifier *AttackModel* on the training set
 - 19: Evaluate Attack Model:
 - 20: Test *AttackModel* on the testing set
 - 21: Output accuracy of *AttackModel*
-

The results for both types of MIA are essential for evaluating Q-DP-GAN's ability to obfuscate the origin of data points, particularly from Tr and S , and its efficacy in preventing adversaries from inferring the membership status of Te . The MIA results, shown in Tables 8–11, illustrate that Q-DP-GAN offers robust protection against both external and internal threats, thus ensuring the confidentiality of EEG data. The low predictive accuracy observed in the BB and WB scenarios indicates substantial data protection, highlighting the model's effectiveness in protecting privacy.

5.2.2. Reconstruction Attacks Analysis

The second phase of our privacy assessment used reconstruction attacks [84] to evaluate the resilience of Q-DP-GAN to attempts to recover original EEG data from its synthetic output. These attacks approximate the inverse of the Q-DP-GAN generator to test how closely synthetic data can be mapped back to training data and to expose weaknesses in anonymization. We trained deep learning models to emulate this reverse process by optimizing a latent representation to match the generator's output to the target synthetic data. The attack model used dense layers for feature extraction, dropout layers to mitigate overfitting, batch normalization for stable training, and regularization to improve generalization. For reconstruction attacks, a higher mean-squared error value is desirable. It signifies challenges and complexity in accurately reconstructing original data from syn-

thetic data, thereby increasing privacy [84]. Tables 12 and 13 show the superior performance of Q-DP-GAN against this attack compared with other state-of-the-art methods.

Algorithm 3 White-Box Membership Inference Attack

Require: Real training data T_r , Synthetic data S , Real test data T_e , Labels y_r (for T_r) and y_e (for T_e), Trained discriminator model discriminator, List of intermediate layer names intermediate_layer_names

Ensure: Attack model to predict membership inference

- 1: Load the trained discriminator model discriminator
 - 2: List the layers of discriminator to choose appropriate intermediate layers
 - 3: **for** each layer in discriminator **do**
 - 4: Print the layer index, name, and output shape
 - 5: **end for**
 - 6: Choose intermediate layers for feature extraction based on the output from step 2
 - 7: Split the real training data into training and testing sets:
 - 8: Split T_r and y_r into $T_{r_{train}}$, $T_{r_{test}}$, $y_{r_{train}}$, $y_{r_{test}}$ with a test size of 20%
 - 9: Prepare attack data:
 - 10: Extract intermediate outputs from the chosen layers of discriminator for $T_{r_{train}}$, S , and T_e
 - 11: Build an intermediate model with the chosen layers
 - 12: Predict the features using the intermediate model on $T_{r_{train}}$, S , and T_e
 - 13: Concatenate features from multiple intermediate layers
 - 14: Create and Label Attack Dataset:
 - 15: Combine features from T_e and label as 1 (real test data)
 - 16: Combine features from S and label as 0 (synthetic data)
 - 17: Form attack dataset $X_{attack} = \{X_t, X_s\}$ and labels $Y_{attack} = \{1s, 0s\}$
 - 18: Build the White-Box attack model:
 - 19: Define the input shape based on the shape of X_{attack}
 - 20: Build the attack model with several dense, batch normalization, and dropout layers
 - 21: Compile the attack model with Adam optimizer, binary cross-entropy loss, and accuracy metric
 - 22: Train the White-Box attack model:
 - 23: Split X_{attack} and Y_{attack} into attack training and testing sets
 - 24: Train the attack model on the attack training set with validation on the attack testing set
 - 25: Evaluate the attack model:
 - 26: Evaluate the attack model on the attack testing set
 - 27: Output the accuracy of the attack model
-

Table 8. Black Box attack simulated on Dataset 2A. Accuracy indicates whether the model can correctly identify the test data (Te). Reduced accuracy scores signify a more robust model against potential attacks. By adding an equivalent quantity of synthetic data (S) to 100% original data (Tr), the total amount of original training data is doubled.

Subject	Our Method	WGAN	CGAN	DPGAN	LDPGAN	ST-DP-GAN
A1	7.78	30.19	32.36	26.69	40.42	29.71
A2	8.26	37.25	40.12	33.21	47.21	30.06
A3	6.55	34.41	28.73	29.38	44.29	31.32
A4	9.47	41.32	42.89	36.52	46.31	38.67
A5	10.14	39.74	45.48	35.07	48.93	28.72
A6	8.82	42.78	39.23	31.98	42.72	32.28
A7	7.22	31.43	29.22	28.05	41.91	25.13
A8	11.58	41.12	48.27	38.53	49.64	34.41
A9	9.08	40.38	36.63	33.14	43.82	27.64

Table 9. Black Box attack simulated on Dataset 2B. Our model's lower accuracy scores indicate that it is more robust to potential attacks than other GAN models.

Subject	Our Method	WGAN	CGAN	DPGAN	LDPGAN	ST-DP-GAN
B1	10.54	44.89	39.48	34.29	45.67	35.14
B2	8.03	36.48	41.07	28.46	39.30	32.67
B3	8.62	39.89	39.67	32.50	42.61	25.39
B4	7.52	37.31	35.76	29.56	43.43	31.05
B5	6.57	27.17	30.17	24.96	38.56	24.68
B6	5.96	26.72	33.03	26.15	40.75	21.43
B7	8.26	34.07	38.76	29.49	40.31	26.82
B8	7.08	30.09	28.08	23.75	37.19	25.21
B9	8.03	36.48	41.07	28.46	39.30	24.73

Table 10. White Box attack simulated on Dataset 2A. Accuracy indicates whether the model can correctly identify the test data (Te). Reduced accuracy scores signify a more robust model against potential attacks. We use 100% synthetic data for each model. Adding an equivalent amount of synthetic data (S) to 100% of the original data (Tr) doubles the training dataset.

Subject	Our Method	WGAN	CGAN	DPGAN	LDPGAN	ST-DP-GAN
A1	17.15	56.38	62.91	48.14	67.12	50.32
A2	19.64	63.27	69.23	51.83	75.33	45.19
A3	15.24	61.56	59.45	45.48	71.21	53.41
A4	19.35	67.23	74.11	56.17	82.67	58.87
A5	18.43	65.16	78.43	54.52	78.72	56.46
A6	17.94	60.35	66.15	53.78	73.82	48.93
A7	16.83	59.23	60.41	49.34	65.25	49.45
A8	22.78	70.36	80.38	60.32	84.31	55.71
A9	20.31	65.88	75.72	53.51	76.23	47.28

Table 11. White Box attack simulated on Dataset 2B. Our model demonstrated performance comparable to the Black Box attack results (lower accuracy scores), indicating its robustness to potential attacks relative to other GAN models.

Subject	Our Method	WGAN	CGAN	DPGAN	LDPGAN	ST-DP-GAN
B1	19.59	73.95	64.03	54.29	76.72	57.84
B2	15.43	60.86	54.92	48.40	67.18	43.12
B3	16.64	68.18	61.18	50.55	75.23	53.57
B4	16.89	63.69	57.58	46.65	68.55	47.90
B5	14.47	55.58	53.90	44.41	59.38	39.28
B6	13.11	54.69	56.02	40.93	64.80	44.63
B7	17.47	69.66	59.95	48.16	69.37	42.03
B8	15.75	57.88	51.31	43.33	61.08	50.91
B9	15.43	60.86	54.92	48.40	67.18	46.45

Table 12. Reconstruction attack simulated on Dataset 2A to assess the model’s ability to reconstruct original input data x from its outputs $M(x)$. Synthetic data (S), equivalent in quantity to the original dataset (Tr), are used for this assessment.

Subject	Our Method	WGAN	CGAN	DPGAN	LDPGAN	ST-DP-GAN
A1	0.128	0.066	0.075	0.084	0.053	0.092
A2	0.123	0.054	0.068	0.073	0.046	0.085
A3	0.132	0.065	0.072	0.079	0.051	0.076
A4	0.116	0.058	0.062	0.068	0.041	0.072
A5	0.118	0.055	0.053	0.062	0.044	0.059
A6	0.127	0.057	0.065	0.068	0.047	0.075
A7	0.134	0.060	0.079	0.082	0.057	0.077
A8	0.110	0.049	0.058	0.060	0.036	0.064
A9	0.132	0.059	0.063	0.072	0.044	0.069

Table 13. Reconstruction attack simulated on Dataset 2B. A higher mean squared error value achieved by our model, similar to that of Dataset 2A, indicates resilience in reconstructing the original data from synthetic data.

Subject	Our Method	WGAN	CGAN	DPGAN	LDPGAN	ST-DP-GAN
B1	0.116	0.066	0.057	0.067	0.041	0.063
B2	0.137	0.074	0.067	0.075	0.054	0.082
B3	0.122	0.071	0.068	0.075	0.047	0.073
B4	0.129	0.078	0.063	0.081	0.053	0.074
B5	0.141	0.090	0.070	0.091	0.066	0.095
B6	0.142	0.082	0.076	0.088	0.059	0.086
B7	0.139	0.072	0.069	0.080	0.051	0.084
B8	0.136	0.086	0.077	0.093	0.061	0.102
B9	0.133	0.074	0.067	0.075	0.054	0.082

5.3. Overfitting and Data Fidelity

Dynamic Time Warping (DTW) [85,86] and Spectral Similarity (SS) [87] are used to evaluate the fidelity of the generated EEG data. DTW measures temporal similarity by aligning real and synthetic signals, with lower DTW values indicating greater similarity. Similarly, SS compares the power spectral densities (PSDs) of real and synthetic EEG signals, with higher values in the range [0, 1] indicating better preservation of frequency characteristics. Tables 14–17 show that the generated EEG data achieve favorable DTW and SS values, indicating high-fidelity synthetic signals that closely resemble real data [88].

In addition to fidelity, we analyze the risk of overfitting due to the relatively small size of the BCI datasets. To mitigate this, the model employs adversarial training and structured latent representations, encouraging the learning of the underlying data distribution rather than memorizing individual samples. From a privacy perspective, overfitting would lead to near-identical reconstructions of the training signals. However, the observed reconstruction differences and downstream classification performance indicate that the generated EEG data preserve task-relevant features without replicating the original samples, suggesting that the model captures generalizable patterns.

Table 14. The table shows the mean Dynamic Time Warping (DTW) values with standard deviations obtained for the generated EEG data of all subjects, A1 to A9 (Dataset 2A).

Subject	A1	A2	A3	A4	A5	A6	A7	A8	A9
Mean	9.85	11.7	9.32	12.89	13.22	14.46	12.62	11.13	12.02
Std	0.15	0.19	0.24	0.14	0.23	0.17	0.18	0.21	0.16

Table 15. The table shows the mean Dynamic Time Warping (DTW) values with standard deviations obtained for the generated EEG data of all subjects, B1 to B9 (Dataset 2B).

Subject	B1	B2	B3	B4	B5	B6	B7	B8	B9
Mean	12.46	8.44	14.12	9.53	11.19	10.35	13.77	10.21	9.63
Std	0.22	0.16	0.20	0.19	0.18	0.26	0.23	0.14	0.21

Table 16. The table shows the mean Spectral Similarity Score (SS) values with standard deviations obtained for the generated EEG data of all subjects, A1 to A9 (Dataset 2A).

Subject	A1	A2	A3	A4	A5	A6	A7	A8	A9
Mean	0.78	0.83	0.79	0.76	0.81	0.83	0.74	0.81	0.84
Std	0.07	0.09	0.05	0.08	0.06	0.08	0.09	0.06	0.05

Table 17. The table shows the mean Spectral Similarity Score (SS) values with standard deviations obtained for the generated EEG data of all subjects, B1 to B9 (Dataset 2B).

Subject	B1	B2	B3	B4	B5	B6	B7	B8	B9
Mean	0.82	0.77	0.83	0.85	0.78	0.75	0.84	0.76	0.80
Std	0.08	0.06	0.09	0.05	0.07	0.06	0.08	0.09	0.07

6. Conclusions

We developed a generative adversarial network (GAN) model that incorporates principles of quantum uncertainty and decoherence to enhance the privacy of EEG data in BCI applications. Quantum uncertainty introduces non-reversible randomness to strengthen security, while quantum decoherence enables adaptive noise during training, aligning privacy with model stability. Dynamic clipping helps balance privacy and learning by preventing early gradient leakage and gradually relaxing constraints to improve data representation. We use Dynamic Time Warping (DTW) to assess temporal alignment between real and synthetic EEG signals, and Spectral Similarity (SS) to evaluate preservation of frequency-domain characteristics, both of which are critical for EEG analysis. These metrics provide comprehensive perspectives on signal fidelity in the time and frequency domains. Our quantum-inspired differential privacy model (Q-DP-GAN) is also resistant to membership inference and reconstruction attacks, thus protecting the source of EEG data. Together, these results establish the significant value of Q-DP-GAN for researchers and practitioners in BCI, providing a solution to balance the utility, privacy, and security of EEG data in practical applications. The synthetic EEG data generated by our model remain valuable for BCI applications while preserving privacy.

7. Limitations and Future Work

Evaluating the model's scalability in real-world scenarios is essential, despite its development in this study with two seminal EEG-BCI datasets. A user-friendly framework or toolkit based on Q-DP-GAN can improve its adoption among researchers and practitioners who lack expertise in privacy-preserving technologies. Although our focus has been on BCI

applications based on EEG data, applying Q-DP-GAN to finance, healthcare, and social media is promising. These fields can benefit substantially from the privacy-preserving capabilities of the model.

While the proposed framework demonstrates strong empirical performance in balancing privacy and utility, several limitations remain. First, the current evaluation of reconstruction attacks relies primarily on MSE, which does not fully capture the temporal and spectral characteristics of EEG signals. Future work will extend this analysis by incorporating metrics such as DTW and frequency-domain similarity to better quantify privacy leakage. Second, although the proposed adaptive (quantum-inspired) noise mechanism is grounded in standard differential privacy theory, a rigorous theoretical proof establishing its formal advantage over existing DP approaches remains an open direction. Additionally, future work will incorporate more comprehensive evaluation metrics, such as precision, recall, and confusion matrices, to provide a deeper understanding of the class-wise performance of downstream tasks.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cryptography10030031/s1>, Tables S1–S5: Classification performance of BCI Dataset IV 2A using various training scenarios. All models are evaluated on the test set using 100% of the real EEG data samples from session II; Tables S6–S10: Classification performance of BCI Dataset IV 2B using various training scenarios. All models are evaluated on test data using 100% of real EEG data samples composed of sessions IV and V.

Author Contributions: Conceptualization, S.P. and G.B.; methodology, S.P. and G.B.; software, S.P.; validation, S.P. and G.B.; formal analysis, S.P.; investigation, S.P.; data curation, S.P.; writing—original draft preparation, S.P.; writing—review and editing, G.B.; supervision, G.B.; project administration, G.B. All authors have read and agreed to the published version of the manuscript.

Funding: Shouvik Paul was supported by the Vector Scholarship in Artificial Intelligence from the Vector Institute and by the Ontario Graduate Scholarship (OGS), funded by the Government of Ontario. Research expenses and APC were supported by Garima Bajwa’s startup funds provided by Lakehead University.

Data Availability Statement: The data used in this study are publicly available from the BCI Competition IV 2008 Graz Dataset 2A and Dataset 2B, provided by Graz University of Technology. The datasets can be accessed via the official BCI Competition repository (<http://www.bbci.de/competition/iv/>, accessed on 5 October 2023).

Acknowledgments: The authors acknowledge the use of large language models (LLMs), including ChatGPT-5, Grammarly, QuillBot, and GPTZero, to refine content and improve clarity, grammar, and sentence structure. All methodological designs, experiments, analyses, and conclusions were conceived and validated by the authors, who take full responsibility for the content of the published article. The authors thank Likhith Varma Muddulur for helping us to improve the quality of the figures.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Chaudhary, P.; Agrawal, R. Emerging threats to security and privacy in brain computer interface. *Int. J. Adv. Stud. Sci. Res.* **2018**, *3*, 340–344. Available online: <https://ssrn.com/abstract=3326692> (accessed on 13 May 2024).
2. Landau, O.; Puzis, R.; Nissim, N. Mind your mind: EEG-based brain-computer interfaces and their security in cyber space. *ACM Comput. Surv. CSUR* **2020**, *53*, 1–38. [CrossRef]
3. Pfurtscheller, G.; Flotzinger, D.; Kalcher, J. Brain-computer interface—A new communication device for handicapped persons. *J. Microcomput. Appl.* **1993**, *16*, 293–299. [CrossRef]

4. Wolpaw, J.R.; Birbaumer, N.; Heetderks, W.J.; McFarland, D.J.; Peckham, P.H.; Schalk, G.; Donchin, E.; Quatrano, L.A.; Robinson, C.J.; Vaughan, T.M. Brain-computer interface technology: A review of the first international meeting. *IEEE Trans. Rehabil. Eng.* **2000**, *8*, 164–173. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Mason, S.G.; Birch, G.E. A general framework for brain-computer interface design. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2003**, *11*, 70–85. [\[CrossRef\]](#)
6. Birbaumer, N.; Ghanayim, N.; Hinterberger, T.; Iversen, I.; Kotchoubey, B.; Kübler, A.; Perelmouter, J.; Taub, E.; Flor, H. A spelling device for the paralysed. *Nature* **1999**, *398*, 297–298. [\[CrossRef\]](#)
7. Janapati, R.; Dalal, V.; Sengupta, R. Advances in modern EEG-BCI signal processing: A review. *Mater. Today Proc.* **2021**, *80*, 2563–2566. [\[CrossRef\]](#)
8. Torres, E.P.; Torres, E.A.; Hernández-Álvarez, M.; Yoo, S.G. EEG-Based BCI Emotion Recognition: A Survey. *Sensors* **2020**, *20*, 5083. [\[CrossRef\]](#)
9. Douibi, K.; Le Bars, S.; Lemontey, A.; Nag, L.; Balp, R.; Breda, G. Toward EEG-Based BCI Applications for Industry 4.0: Challenges and Possible Applications. *Front. Hum. Neurosci.* **2021**, *15*, 705064. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Li, J.; Wu, H.; Ding, L.; Wu, D. Meta-Learning for Fast and Privacy-Preserving Source Knowledge Transfer of EEG-Based BCIs. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2023**, *31*, 123–134. [\[CrossRef\]](#)
11. Landau, O.; Cohen, A.; Gordon, S.; Nissim, N. Mind your privacy: Privacy leakage through BCI applications using machine learning methods. *Knowl.-Based Syst.* **2020**, *198*, 105932. [\[CrossRef\]](#)
12. Brocol, F. Brain-computer interfaces in safety and security fields: Risks and applications. *Saf. Sci.* **2023**, *160*, 106051. [\[CrossRef\]](#)
13. Martínez Beltrán, E.T.; Quiles Pérez, M.; López Bernal, S.; Huertas Celdran, A.; Martínez Pérez, G. Noise-based cyberattacks generating fake P300 waves in brain–computer interfaces. *Clust. Comput.* **2022**, *25*, 33–48. [\[CrossRef\]](#)
14. Bernal, S.L.; Celdran, A.H.; Maimo, L.F.; Barros, M.T.; Balasubramaniam, S.; Perez, G.M. Cyberattacks on miniature brain implants to disrupt spontaneous neural signaling. *IEEE Access* **2020**, *8*, 152204–152222. [\[CrossRef\]](#)
15. Xia, K.; Duch, W.; Sun, Y.; Xu, K.; Fang, W.; Luo, H.; Zhang, Y.; Sang, D.; Xu, X.; Wang, F.Y. Privacy-preserving brain–computer interfaces: A systematic review. *IEEE Trans. Comput. Soc. Syst.* **2022**, *10*, 2312–2324. [\[CrossRef\]](#)
16. Mokhiemar, M.O.; Mahmoud, A.; Eldagla, M.I.; Aribi, Y.; Anter, A.M. Blockchain-integrated multi-modal LSTM-CNN fusion for high-precision epileptic seizure detection from EEG signals. *Knowl.-Based Syst.* **2025**, *323*, 113703. [\[CrossRef\]](#)
17. Vărbu, K.; Muhammad, N.; Muhammad, Y. Past, Present, and Future of EEG-Based BCI Applications. *Sensors* **2022**, *22*, 3331. [\[CrossRef\]](#)
18. Singh, G.; Patel, P.; Asaduzzaman, M.; Bajwa, G. Selective eeg signal anonymization using multi-objective autoencoders. In Proceedings of the 2023 20th Annual International Conference on Privacy, Security and Trust (PST), Copenhagen, Denmark, 21–23 August 2023; pp. 1–7.
19. Paul, S.; Bajwa, G. Privacy-Preserving EEG Data Generation: A Federated Split Learning Approach Using Privacy-Adaptive Autoencoders and Secure Aggregation with GFlowNet. In Proceedings of the 22nd International Conference on Security and Cryptography, Bilbao, Spain, 11–13 June 2025; pp. 638–643.
20. Wood, A.; Najarian, K.; Kahrobaei, D. Homomorphic encryption for machine learning in medicine and bioinformatics. *ACM Comput. Surv. CSUR* **2020**, *53*, 1–35. [\[CrossRef\]](#)
21. Popescu, A.B.; Taca, I.A.; Nita, C.I.; Vizitiu, A.; Demeter, R.; Suciu, C.; Itu, L.M. Privacy Preserving Classification of EEG Data Using Machine Learning and Homomorphic Encryption. *Appl. Sci.* **2021**, *11*, 7360. [\[CrossRef\]](#)
22. Dwork, C. Differential privacy. In *Proceedings of the International Colloquium on Automata, Languages, and Programming*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 1–12.
23. Debie, E.; Moustafa, N.; Whitty, M.T. A privacy-preserving generative adversarial network method for securing eeg brain signals. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–8.
24. Blanco-Justicia, A.; Sánchez, D.; Domingo-Ferrer, J.; Muralidhar, K. A critical review on the use (and misuse) of differential privacy in machine learning. *ACM Comput. Surv.* **2022**, *55*, 160. [\[CrossRef\]](#)
25. Zhao, J.; Chen, Y.; Zhang, W. Differential privacy preservation in deep learning: Challenges, opportunities and solutions. *IEEE Access* **2019**, *7*, 48901–48911. [\[CrossRef\]](#)
26. Alshebli, S.; Alshehhi, M.; Yeun, C.Y. Investigating How Data Poising Attacks Can Impact An EEG-Based Federated Learning Model. In Proceedings of the 2024 2nd International Conference on Cyber Resilience (ICCR), Dubai, United Arab Emirates, 26–28 February 2024; pp. 1–6.
27. Wang, K.; Yang, M.; Li, C.; Liu, A.; Qian, R.; Chen, X. Privacy-Preserving Domain Adaptation for Intracranial EEG Classification via Information Maximization and Gaussian Mixture Model. *IEEE Sens. J.* **2023**, *23*, 26390–26400. [\[CrossRef\]](#)
28. Agarwal, A.; Dowsley, R.; McKinney, N.D.; Wu, D.; Lin, C.T.; De Cock, M.; Nascimento, A.C. Protecting privacy of users in brain-computer interface applications. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 1546–1555. [\[CrossRef\]](#) [\[PubMed\]](#)

29. Agarwal, A.; Dowsley, R.; McKinney, N.D.; Wu, D.; Lin, C.T.; De Cock, M.; Nascimento, A. Privacy-preserving linear regression for brain-computer interface applications. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 5277–5278.
30. Volgushev, N.; Schwarzkopf, M.; Getchell, B.; Varia, M.; Lapets, A.; Bestavros, A. Conclave: Secure multi-party computation on big data. In *Proceedings of the Fourteenth EuroSys Conference (EuroSys 2019)*; ACM: New York, NY, USA, 2019; pp. 1–18.
31. Schiliro, F.; Moustafa, N.; Beheshti, A. Cognitive privacy: AI-enabled privacy using EEG signals in the internet of things. In Proceedings of the 2020 IEEE 6th International Conference on Dependability in Sensor, Cloud and Big Data Systems and Application (DependSys), Nadi, Fiji, 14–16 December 2020; pp. 73–79.
32. Pazouki, S.; Golilarz, N.A.; Kazemi-Razi, S.M.; Aydeger, A. A self-healing cybersecurity mechanism for cyberattacks targeting artificial neural network-based human brain implants controlling smart homes. In Proceedings of the 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET), Prague, Czech Republic, 20–22 July 2022; pp. 1–6.
33. Bidgoly, A.J.; Bidgoly, H.J.; Arezoumand, Z. Towards a universal and privacy preserving EEG-based authentication system. *Sci. Rep.* **2022**, *12*, 2531. [[CrossRef](#)] [[PubMed](#)]
34. Gui, Q.; Yang, W.; Jin, Z.; Ruiz-Blondet, M.V.; Laszlo, S. A residual feature-based replay attack detection approach for brainprint biometric systems. In Proceedings of the 2016 IEEE International Workshop on Information Forensics and Security (WIFS), Abu Dhabi, United Arab Emirates, 4–7 December 2016; pp. 1–6.
35. Mezzina, G.; Annese, V.F.; De Venuto, D. A Cybersecure P300-Based Brain-to-Computer Interface against Noise-Based and Fake P300 Cyberattacks. *Sensors* **2021**, *21*, 8280. [[CrossRef](#)]
36. Maiorana, E.; Hine, G.E.; La Rocca, D.; Campisi, P. On the vulnerability of an EEG-based biometric system to hill-climbing attacks algorithms' comparison and possible countermeasures. In Proceedings of the 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), Arlington, VA, USA, 29 September–2 October 2013; pp. 1–6.
37. Wang, M.; Wang, S.; Hu, J. PolyCosGraph: A Privacy-Preserving Cancelable EEG Biometric System. *IEEE Trans. Dependable Secur. Comput.* **2022**, *20*, 4258–4272. [[CrossRef](#)]
38. Shateri, M.; Messina, F.; Labeau, F.; Piantanida, P. Preserving privacy in GANs against membership inference attack. *IEEE Trans. Inf. Forensics Secur.* **2023**, *19*, 1728–1743. [[CrossRef](#)]
39. Huang, D.; Wang, M.; Wang, J. A survey of quantum computing hybrid applications with brain-computer interface. *Cogn. Robot.* **2022**, *2*, 164–176. [[CrossRef](#)]
40. Ullah, U. and Garcia-Zapirain, B. Quantum Machine Learning Revolution in Healthcare: A Systematic Review of Emerging Perspectives and Applications. *IEEE Access* **2024**, *12*, 11423–11450. [[CrossRef](#)]
41. Singh, D.; Kanathe, Y.; Waykole, Y.; Mishra, R.K.; Walambe, R.; Aqeel, K.H.; Kotecha, K. Exploring the Usability of Quantum Machine Learning for EEG Signal Classification. In *Proceedings of the 13th International Conference on Advanced Computing (IACC 2023)*; Springer Nature Switzerland: Cham, Switzerland, 2024; pp. 427–438.
42. Tasci, G.; Gun, M.V.; Keles, T.; Tasci, B.; Barua, P.D.; Tasci, I.; Dogan, S.; Baygin, M.; Palmer, E.E.; Tuncer, T.; et al. QLBP: Dynamic patterns-based feature extraction functions for automatic detection of mental health and cognitive conditions using EEG signals. *Chaos Solitons Fractals* **2023**, *172*, 113472. [[CrossRef](#)]
43. Kim, J.H.; Cho, Y.; Suh, Y.A.; Yim, M.S. Development of an information security-enforced eeg-based nuclear operators' fitness for duty classification system. *IEEE Access* **2021**, *9*, 72535–72546. [[CrossRef](#)]
44. Liao, K.; Yang, Z.; Tao, D.; Zhao, L.; Pires, N.; Dorao, C.A.; Stokke, B.T.; Roseng, L.E.; Liu, W.; Jiang, Z. Exploring the intersection of brain-computer interfaces and quantum sensing: A review of research progress and future trends. *Adv. Quantum Technol.* **2024**, *7*, 2300185. [[CrossRef](#)]
45. Alawsi, W.A.; Mosleh, M.; Obayes, H.K.; Mosleh, M. TL-FL-QWOA: A quantum-optimized transfer-federated framework for privacy-preserving EEG-based emotion recognition in IoMT. *Int. J. Math. Model. Comput.* **2025**, *16*, 173–189.
46. Cerezo, M.; Verdon, G.; Huang, H.Y.; Cincio, L.; Coles, P.J. Challenges and opportunities in quantum machine learning. *Nat. Comput. Sci.* **2022**, *2*, 567–576. [[CrossRef](#)]
47. Tang, E. A quantum-inspired classical algorithm for recommendation systems. In Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, Phoenix, AZ, USA, 23–26 June 2019; pp. 217–228.
48. Biamonte, J.; Wittek, P.; Pancotti, N.; Rebentrost, P.; Wiebe, N.; Lloyd, S. Quantum machine learning. *Nature* **2017**, *549*, 195–202. [[CrossRef](#)]
49. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2014; Volume 27, pp. 2672–2680.
50. Yu, L.; Zhang, W.; Wang, J.; Yu, Y. SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient. In Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; Volume 31.
51. Dwork, C.; Rothblum, G.N.; Vadhan, S. Boosting and differential privacy. In Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science, Las Vegas, NV, USA, 23–26 October 2010; pp. 51–60.

52. Dwork, C.; McSherry, F.; Nissim, K.; Smith, A. Calibrating noise to sensitivity in private data analysis. *J. Priv. Confid.* **2006**, *7*, 17–34.
53. Dwork, C.; Roth, A. The Algorithmic Foundations of Differential Privacy. *Found. Trends Theor. Comput. Sci.* **2014**, *9*, 211–487. [[CrossRef](#)]
54. Amoroso, R.L. *Universal Quantum Computing: Supervening Decoherence-Surmounting Uncertainty*; World Scientific: Singapore, 2017.
55. Bi, L.; Liang, K.; Czap, G.; Wang, H.; Yang, K.; Li, S. Recent progress in probing atomic and molecular quantum coherence with scanning tunneling microscopy. *Prog. Surf. Sci.* **2023**, *98*, 100696. [[CrossRef](#)]
56. Abdikhakimov, I. The Uncertainty Principle: How Quantum Mechanics Is Transforming Jurisprudence. *Int. J. Cyber Law* **2023**, *1*. Available online: <https://irshadjournals.com/index.php/ijcl/article/view/122> (accessed on 10 June 2024).
57. Moore, M.; Narayanan, A. *Quantum-Inspired Computing*; Technical Report; Department of Computer Science, University of Exeter: Exeter, UK, 1995.
58. Jaeger, G. *Quantum Information*; Springer: Berlin/Heidelberg, Germany, 2007.
59. Lew, K.L.; Sim, K.S.; Ting, Z. Deep Learning Approach EEG Signal Classification. *JOIV Int. J. Inform. Vis.* **2024**, *8*, 1693–1702. [[CrossRef](#)]
60. Banville, H.; Chehab, O.; Hyvärinen, A.; Engemann, D.A.; Gramfort, A. Uncovering the structure of clinical EEG signals with self-supervised learning. *J. Neural Eng.* **2021**, *18*, 046020. [[CrossRef](#)] [[PubMed](#)]
61. Schirrmester, R.T.; Springenberg, J.T.; Fiederer, L.D.J.; Glasstetter, M.; Eggensperger, K.; Tangermann, M.; Hutter, F.; Burgard, W.; Ball, T. Deep learning with convolutional neural networks for EEG decoding and visualization. *Hum. Brain Mapp.* **2017**, *38*, 5391–5420. [[CrossRef](#)]
62. Abadi, M.; Chu, A.; Goodfellow, I.; McMahan, H.B.; Mironov, I.; Talwar, K.; Zhang, L. Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, 24–28 October 2016; pp. 308–318.
63. Saha, D.K.; Calhoun, V.D.; Du, Y.; Fu, Z.; Kwon, S.M.; Sarwate, A.D.; Panta, S.R.; Plis, S.M. Privacy-preserving quality control of neuroimaging datasets in federated environments. *Hum. Brain Mapp.* **2022**, *43*, 2289–2310. [[CrossRef](#)]
64. Wang, P.; Lei, Y.; Ying, Y.; Zhang, H. Differentially private SGD with non-smooth losses. *Appl. Comput. Harmon. Anal.* **2022**, *56*, 306–336. [[CrossRef](#)]
65. Wang, D.; Xu, J. Differentially private high dimensional sparse covariance matrix estimation. *Theor. Comput. Sci.* **2021**, *865*, 119–130. [[CrossRef](#)]
66. Brunner, C.; Leeb, R.; Müller-Putz, G.; Schlögl, A.; Pfurtscheller, G. *BCI Competition 2008–Graz Data Set A*; Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology: Graz, Austria, 2008; Volume 16, pp. 1–6.
67. Leeb, R.; Brunner, C.; Müller-Putz, G.; Schlögl, A.; Pfurtscheller, G. *BCI Competition 2008–Graz Data Set B*; Graz University of Technology: Graz, Austria, 2008; Volume 16, pp. 1–6.
68. Roy, Y.; Banville, H.; Albuquerque, I.; Gramfort, A.; Falk, T.H.; Faubert, J. Deep learning-based electroencephalography analysis: A systematic review. *J. Neural Eng.* **2019**, *16*, 051001. [[CrossRef](#)]
69. Ang, K.K.; Chin, Z.Y.; Wang, C.; Guan, C.; Zhang, H. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Front. Neurosci.* **2012**, *6*, 39. [[CrossRef](#)] [[PubMed](#)]
70. Campisi, P.; La Rocca, D. Brain waves for automatic biometric-based user recognition. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 782–800. [[CrossRef](#)]
71. Marcel, S.; Millán, J.d.R. Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 743–752. [[CrossRef](#)]
72. Hu, S.; Wang, H.; Zhang, J.; Kong, W.; Cao, Y. Causality from Cz to C3/C4 or between C3 and C4 revealed by Granger causality and new causality during motor imagery. In Proceedings of the 2014 International Joint Conference on Neural Networks (IJCNN), Beijing, China, 6–11 July 2014; pp. 3178–3185.
73. Gaur, P.; Pachori, R.B.; Wang, H.; Prasad, G. A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry. *Expert Syst. Appl.* **2018**, *95*, 201–211. [[CrossRef](#)]
74. Gaur, P.; Pachori, R.B.; Wang, H.; Prasad, G. An automatic subject specific intrinsic mode function selection for enhancing two-class EEG-based motor imagery-brain computer interface. *IEEE Sens. J.* **2019**, *19*, 6938–6947.
75. Altaheri, H.; Muhammad, G.; Alsulaiman, M. Physics-informed attention temporal convolutional network for EEG-based motor imagery classification. *IEEE Trans. Ind. Inform.* **2022**, *19*, 2249–2258. [[CrossRef](#)]
76. Lawhern, V.J.; Solon, A.J.; Waytowich, N.R.; Gordon, S.M.; Hung, C.P.; Lance, B.J. EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces. *J. Neural Eng.* **2018**, *15*, 056013. [[CrossRef](#)] [[PubMed](#)]
77. Ha, K.W.; Jeong, J.W. Motor imagery EEG classification using capsule networks. *Sensors* **2019**, *19*, 2854. [[CrossRef](#)]
78. Faisal, F.; Mohammed, N.; Leung, C.K.; Wang, Y. Generating privacy preserving synthetic medical data. In Proceedings of the 2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA), Shenzhen, China, 13–16 October 2022; pp. 1–10.

79. Torfi, A.; Fox, E.A.; Reddy, C.K. Differentially private synthetic medical data generation using convolutional GANs. *Inf. Sci.* **2022**, *586*, 485–500. [[CrossRef](#)]
80. Xie, L.; Lin, K.; Wang, S.; Wang, F.; Zhou, J. Differentially private generative adversarial network. *arXiv* **2018**, arXiv:1802.06739. [[CrossRef](#)]
81. Shao, W.; Zhu, R.; Yang, C.; Thapa, C.; Ahmed, M.E.; Camtepe, S.; Zhang, R.; Kim, D.Y.; Menouar, H.; Salim, F.D. ST-DPGAN: A Privacy-preserving Framework for Spatiotemporal Data Generation. *IEEE Internet Things J.* **2025**, *12*, 29333–29342. [[CrossRef](#)]
82. Gwon, Y. LDP-GAN: Generative adversarial networks with local differential privacy for patient medical records synthesis. *J. Med. Syst.* **2024**, *168*, 107738. [[CrossRef](#)]
83. Hayes, J.; Melis, L.; Danezis, G.; De Cristofaro, E. Logan: Membership inference attacks against generative models. *arXiv* **2017**, arXiv:1705.07663. [[CrossRef](#)]
84. Li, Z.; Yang, M.; Liu, Y.; Wang, J.; Hu, H.; Yi, W.; Xu, X. GAN You See Me? Enhanced Data Reconstruction Attacks against Split Inference. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 54554–54566.
85. Müller, M. Dynamic Time Warping. In *Information Retrieval for Music and Motion*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 69–84.
86. Li, H. Time works well: Dynamic time warping based on time weighting for time series data mining. *Inf. Sci.* **2021**, *547*, 592–608. [[CrossRef](#)]
87. Gupta, K.; Thomas, D.; Vidya, S.; Venkatesh, K.V.; Ramakumar, S. Detailed protein sequence alignment based on Spectral Similarity Score (SSS). *BMC Bioinform.* **2005**, *6*, 105. [[CrossRef](#)] [[PubMed](#)]
88. Geng, D.; Chen, Z.S. Auxiliary classifier generative adversarial network for interictal epileptiform discharge Modeling and EEG data augmentation. In Proceedings of the 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER), Virtual, 4–6 May 2021; pp. 1130–1133.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.