International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(15s) 2026 pp. 673-691
https://doi.org/10.70917/ijcisim-2026-4464
Article

BST-NS: A Biologically Inspired Spectral-Temporal Noise Shaping Framework for Privacy-Preserving Federated ECG Learning

Rajakumar Ponnumani[1], T. Rathi Devi[2], Sonia kumari[3], Sandeep Sharma[4], P. Tamilchelvan[5], C. Ramesh Kumar[6]

[1] Computer Science and Engineering, Parul University, Vadodara, Gujarat, India
prajakumarkk@gmail.com
[2] Computer Science and Engineering, CHRIST (Deemed to be University), Bengaluru, Karnataka, India
rathidevi.t@christuniversity.in
[3] Department of Computer Application, Noida Institute of Engineering and Technology, Greater Noida, Uttar Pradesh, India
soniakumari5894@gmail.com
[4] School of Business (Business Analytics), Galgotias University, Greater Noida, Uttar Pradesh, India
sandeepsharma@galgotiasuniversity.edu.in
[5] Department of Agricultural Engineering, Paavai Engineering College, Namakkal, TamilNadu, India
tamilchelvanp9@gmail.com
[6] School of Artificial Intelligence (SoAI), Galgotias University, Greater Noida, Uttar Pradesh, India
toramesh83@gmail.com

Abstract: Federated learning enables collaborative model training across distributed electrocardiogram (ECG) datasets while preserving patient privacy, but conventional differential privacy mechanisms degrade model accuracy by adding isotropic Gaussian noise without regard for the clinical structure of ECG signals. We propose a biologically inspired spectral-temporal noise shaper (BST-NS) that replaces standard noise injection with an adaptive, frequency-selective perturbation process tailored to preserve diagnostically relevant ECG features. Our method decomposes client-side gradient updates using empirical mode decomposition, then applies a fractional-order filter who's cutoff and roll-off are dynamically controlled by a recurrent neural network. This network processes spectral power metrics and heart rate variability statistics from the local ECG data to output two parameters: a fractional filter order and a noise amplitude scaling factor. The resulting noise vector concentrates its energy in frequency bands outside the clinically critical range of 0.04 to 40 hertz, where cardiac waveform morphology and arrhythmia signatures reside. We realize the filter using the Oustaloup approximation and perform noise generation via frequency-domain convolution for computational efficiency. In our federated framework, each wearable client adds this structured noise to its clipped gradient before sending the update to a cloud aggregator, which then performs global model averaging. The privacy guarantee is verified using the moments accountant to achieve differential privacy with epsilon equal to one and delta equal to ten to the minus five. Our primary contribution is a privacy-preserving mechanism that decouples utility from noise magnitude by exploiting the frequency-domain characteristics of ECG gradients. This approach significantly reduces distortion of PQRST waveform morphology and heart rate variability metrics compared to isotropic noise injection, thereby maintaining high arrhythmia classification accuracy. The work demonstrates that biologically motivated noise shaping can reconcile strong privacy protections with clinically meaningful model performance in cloud-based health data aggregation systems.

Keywords: Federated Learning, Differential Privacy, Electrocardiogram (ECG), Spectral-Temporal Noise Shaping, Biologically Inspired Privacy, Empirical Mode Decomposition (EMD), Fractional-Order Filtering, Recurrent Neural Network (RNN), Wearable Healthcare

1. Introduction

The proliferation of wearable biosensors and cloud-based health platforms has created unprecedented opportunities for collaborative cardiac diagnostics across large patient populations. Electrocardiogram (ECG) signals, which capture the heart's electrical activity through characteristic PQRST waveforms and heart rate variability (HRV) metrics, serve as fundamental diagnostic tools for detecting arrhythmias and other cardiac abnormalities [1]. However, the sensitive nature of these physiological recordings creates substantial barriers to data Arrhythmia Classification sharing among medical institutions, primarily due to stringent privacy regulations and concerns about re-identification attacks from raw waveform exposure.

Federated learning (FL) has emerged as a promising framework for training machine learning models across decentralized health data repositories without requiring direct access to patient records [2]. In FL, individual clients train local models and share only model updates—such as gradient vectors—with a central aggregation server. Yet empirical evidence demonstrates that these gradient updates can inadvertently leak substantial information about local training samples, including ECG traces. Consequently, differential privacy (DP) mechanisms are integrated into federated pipelines to provide formal privacy guarantees, most commonly through the injection of isotropic Gaussian noise into model updates [3] [4]. These standard DP approaches achieve rigorous privacy bounds but impose a fundamental trade-off between privacy protection and model utility: high noise levels obscure individual contributions but simultaneously degrade the diagnostic fidelity of downstream models, particularly for tasks requiring precise waveform morphology analysis.

This limitation is especially acute in ECG-based diagnostics. Cardiac arrhythmia classification models, often implemented as convolutional neural networks (CNNs), rely on subtle morphological features embedded in PQRST complexes and inter-beat interval variability [1]. Isotropic noise injection blurs these features indiscriminately across the frequency spectrum, undermining the very signal characteristics that enable accurate

diagnosis. Traditional privacy-preserving techniques, including those targeting wireless body area networks (WBANs) and cloud-based healthcare aggregation [5] [6], have largely focused on cryptographic or generic noise-based approaches without considering the unique spectral properties of physiological signals.

To address this challenge, we draw inspiration from two well-established phenomena. First, empirical mode decomposition (EMD) provides a data-driven method for decomposing non-stationary signals into intrinsic mode functions that isolate distinct oscillatory components across different frequency bands [7]. Second, stochastic resonance in neural systems demonstrates that carefully structured noise can paradoxically enhance signal detection and information transmission in threshold-based systems [8]. Together, these principles suggest an alternative path: rather than treating noise as a uniform perturbation to be minimized, we can shape noise spectrally to exploit redundancy in clinically uninformative frequency bands while preserving diagnostic content.

Our contribution is a biologically inspired spectral-temporal noise shaper (BST-NS) for privacy-preserving federated ECG learning. The proposed mechanism replaces isotropic Gaussian perturbation with structured noise whose spectral density is dynamically constrained to frequency bands outside the clinically significant range (0.04–40 Hz) where PQRST morphology and HRV metrics reside. To achieve this, we integrate three key components: an EMD-based analysis that identifies redundant ECG frequency bands, a fractional-order filter bank—motivated by the flexibility of fractional calculus [9]—whose parameters are adaptively controlled by a recurrent neural network (RNN) controller [10], and a noise generation process that adds perturbations in the spectral domain before inverse transformation. The filter cutoff and roll-off characteristics are computed per client based on local ECG spectral statistics and HRV features, ensuring that privacy noise targets clinically irrelevant frequency regions. This approach ensures differential privacy guarantees with epsilon equal to one while preserving diagnostic accuracy for arrhythmia detection tasks.

The remainder of this paper is organized as follows. Section 2 reviews related work in federated learning for healthcare, differential privacy mechanisms, and ECG signal processing. Section 3 formulates the problem and provides necessary preliminaries on differential privacy, EMD, and fractional-order filters. Section 4 details our proposed BST-NS framework, including the EMD-based band identification module, the RNN controller architecture, and the noise shaping algorithm. Section 5 describes the experimental setup, including datasets (such as the MIT-BIH Arrhythmia Database), baseline methods, and evaluation metrics. Section 6 presents quantitative results comparing our approach against standard DP mechanisms in terms of privacy-utility trade-offs, waveform preservation, and classification accuracy. Section 7 discusses limitations and potential

extensions, while Section 8 concludes the paper with a summary of findings and implications for secure multi-institutional health data aggregation.

2. Related Work

We review existing works across three intersecting domains: privacy-preserving machine learning for health data, differential privacy mechanisms applied to neural networks, and signal-adaptive noise injection techniques in biomedical signal processing.

2.1 Privacy-Preserving Machine Learning for Health Data Aggregation

The aggregation of sensitive health data from distributed sources, particularly in cloud-based Wireless Body Area Networks (WBANs), has motivated a wide range of privacy-preserving methodologies. Early approaches relied predominantly on cryptographic techniques. For instance, verifiable multidimensional encrypted medical data aggregation (VMEMDA) schemes employ homomorphic encryption to allow a cloud aggregator to compute sums over encrypted client updates without decrypting individual contributions [1]. Similarly, quantum-enhanced privacy aggregation protocols have been proposed for healthcare monitoring in WBANs, leveraging quantum key distribution to secure data transmission against eavesdropping [11]. While these cryptographic methods provide strong confidentiality guarantees during transmission and aggregation, they do not address the fundamental challenge of information leakage from the model updates themselves. The raw gradient vectors computed from local ECG data inherently encode diagnostic features; hence, even encrypted gradient aggregation can still leak patient information through the aggregated model weights after sufficient training rounds.

Federated learning has been extensively adopted to mitigate raw data exposure in healthcare diagnostics. For example, a federated learning framework for privacy-preserving healthcare diagnostics using biomedical signal analysis applies local differential perturbation at the client before aggregation [12]. Their framework, evaluated on ECG and EEG data, maintains high classification accuracy but relies on uniform noise mechanisms that do not discriminate between clinically informative and redundant components of the gradient signal. Other works propose dewcloud-based hierarchical federated learning architectures for intrusion detection in healthcare IoT systems [13], or combine deep learning with blockchain for privacy-enhanced heart disease prediction in cloud-based systems [14]. These approaches, while effective for secure data management, treat data utility and privacy as antagonistic objectives, often sacrificing model accuracy to meet formal privacy bounds.

Our work differs from these previous contributions by recasting the privacy-utility trade-off as a frequency-domain optimization problem: we inject noise that perturbs gradient updates but directs the perturbation energy away from clinically critical frequency bands in the ECG signal space.

2.2 Differential Privacy and Noise Injection in Gradient-Based Learning

Differential privacy provides a formal mathematical framework for bounding the influence of any single data point on the output of a computation. In the context of deep learning, the seminal work of Abadi et al. introduced the framework of differentially private stochastic gradient descent (DP-SGD), which clips per-example gradients and adds isotropic Gaussian noise scaled to the clipping threshold [6]. This approach has become the de facto standard for privacy-preserving model training, with extensions to federated settings such as the work of McMahan et al. [7] and differential privacy in empirical risk minimization [8].

Various enhancements to the basic DP-SGD framework have been proposed. Methods such as gradient compression with correlated noise [15] and adaptive clipping strategies [16] aim to reduce the variance introduced by the isotropic noise component. However, a consistent limitation across all these works is their treatment of noise as a scalar- or vector-valued Gaussian perturbation applied uniformly across all dimensions and frequency components of the gradient. For structured data like ECG signals, where the gradient space is not isotropic in a semantic sense—certain parameter directions correspond to diagnostically critical features—this uniform perturbation is inherently suboptimal.

Fractional-order filters have been explored in control theory and signal processing for their ability to provide more flexible roll-off characteristics compared to integer-order filters [11]. Their application in privacy-preserving machine learning, however, remains largely unexplored.

2.3 Signal-Adaptive Noise Shaping and Stochastic Resonance

The concept of shaping noise to exploit signal structure has roots in neural stochastic resonance, where adding noise to a non-linear system can improve signal detection when the noise spectral properties match the system's threshold characteristics [17]. This principle has been empirically validated in sensory neuroscience and has inspired computational models for enhancing signal-to-noise ratios in audio processing [18] and image denoising [19].

In the domain of biomedical signal processing, empirical mode decomposition (EMD) has been extensively used for analyzing non-stationary signals such as ECG and EEG [15]. EMD decomposes a signal into intrinsic mode functions (IMFs) with instantaneous frequency content, providing a natural basis for identifying frequency bands that carry diagnostic information. In the context of ECG signal denoising, EMD-based methods have been applied to remove motion artifacts and baseline wander while preserving PQRST morphology [20]. However, the application of

EMD to the gradient space of neural network updates—and using its decomposition to guide privacy noise injection—is a novel contribution of our work.

Recurrent neural networks, particularly gated recurrent units (GRUs) [21], are effective for modeling temporal dependencies and have been employed in adaptive signal processing systems for tasks such as dynamic filter design [22].

2.4 Key Novelty of the Proposed Approach

The proposed BST-NS framework synthesizes these disparate lines of research into a unified privacy-preserving mechanism. While existing works either ignore signal structure when adding noise [6] or apply signal-adaptive denoising after model training [20], our approach is the first to embed the signal structure directly into the privacy noise generation process before aggregation. By using EMD to identify clinically redundant frequency bands in the client gradients, an RNN controller to adaptively set a fractional-order filter's cutoff and roll-off, and spectral-domain convolution to produce the additive noise, we achieve a clean separation between diagnostic content and privacy perturbation. This stands in contrast to prior works that treat the entire gradient space as uniformly sensitive, thereby forcing a wasteful trade-off between epsilon and accuracy. Instead, our method redistributes the noise power along dimensions that minimally impact the downstream ECG classification pipeline, enabling strong privacy guarantees without commensurate degradation in arrhythmia detection performance.

3. Preliminaries and Problem Formulation

This section establishes the foundational concepts and formal notation required to understand our proposed framework. We begin by defining the federated learning paradigm for ECG diagnostics, then review the principles of differential privacy and the moments accountant, and finally introduce empirical mode decomposition and fractional-order filtering as the core building blocks of our noise shaping approach.

3.1 Federated Learning for ECG Classification

Consider a set of K clients (e.g., hospitals or wearable device users), each holding a local dataset $D_k = \{(x_i, y_i)\}_{i=1}^{n_k}$, where $x_i \in R^T$ denotes an ECG time series of length T and $y_i \in Y$ is the corresponding cardiac diagnosis label (e.g., normal sinus rhythm, atrial fibrillation, ventricular tachycardia). The global objective in federated learning is to minimize the empirical risk:

$$\min_{w} L(w) = \sum_{k=1}^{K} \frac{n_k}{N} E_{(x,y)\sim D_k}[l(f(x;w),y)]$$

where $N = \sum_{k=1}^{K} n_k$ is the total number of samples across all clients, $f(\cdot;w)$ is a neural network parameterized by weights w, and $l(\cdot,\cdot)$ is a loss function such as cross-entropy for arrhythmia classification.

In the standard Federated Averaging (FedAvg) protocol [7], each communication round t proceeds as follows: the central server broadcasts the global model $w^{(t)}$ to all clients; each client performs E local stochastic gradient descent (SGD) updates on its own data and returns the accumulated update $\Delta w_k^{(t)} = w^{(t)} - w_k^{(1)}$; the server aggregates these updates as $\Delta w^{(t)} = \frac{1}{K}\sum_{k=1}^{K} \Delta w_k^{(t)}$ and updates the global model $w^{(1)} = w^{(t)} - \eta \Delta w^{(t)}$, where η is the learning rate.

3.2 Differential Privacy and the Moments Accountant

To provide formal privacy guarantees for the gradient updates, we adopt the framework of differential privacy (DP) [8]. A randomized mechanism M satisfies (ε,δ)-differential privacy if for any two adjacent datasets D and D' differing by a single record, and for any subset of outputs $S \subseteq Range(M)$:

$$Pr[M(D) \in S] \leq e^{\varepsilon} Pr[M(D') \in S] + \delta$$

In the context of federated learning, differential privacy is enforced by clipping each client's gradient update to a maximum L_2 norm C and then adding Gaussian noise $N(0, \sigma^2 C^2 I)$ before transmission to the server.

To track the cumulative privacy loss over multiple training rounds, we employ the moments accountant [6]. Given a composition of T mechanisms, each with privacy parameter $(\varepsilon_t, \delta_t)$, the moments accountant tracks the log-moment of the privacy loss random variable:

$$\alpha_M(\lambda) = \max_{D,D'} log\, E_{o\sim M(D)}\left[\left(\frac{Pr[M(D) = o]}{Pr[M(D') = o]}\right)^{\lambda}\right]$$

After T rounds, the overall privacy guarantee is (ε, δ)-DP for $\varepsilon = \min_{\lambda}(\frac{1}{\lambda}\sum_{t=1}^{T} \quad \alpha_t(\lambda) - \frac{loglog\, \delta}{\lambda})$. This accountant provides tighter privacy bounds compared to the basic composition theorem, allowing us to achieve $\varepsilon = 1$ with $\delta = 10^{-5}$ for our experiments.

3.3 Empirical Mode Decomposition for Non-Stationary Signals

Empirical mode decomposition (EMD) is an adaptive, data-driven technique for decomposing non-stationary and non-linear signals into a finite set of intrinsic mode functions (IMFs) [15]. For an ECG signal $x \in R^T$, EMD yields:

$$x = \sum_{m=1}^{M} \quad IMF_m + r$$

where IMF_m is the m-th IMF representing oscillations at different instantaneous frequencies, and r is the residual trend. Each IMF satisfies two conditions: (i) the number of extrema and zero-crossings differ by at most one, and (ii) the mean of the upper and lower envelopes (defined by cubic spline interpolation of local maxima and minima) is zero everywhere.

EMD is particularly suited for ECG analysis because it does not assume stationarity or a predefined basis, unlike Fourier or wavelet transforms [20]. The IMFs naturally isolate cardiac components: early IMFs capture high-frequency noise and muscle artifacts, middle IMFs contain PQRST morphology, and late IMFs represent baseline wander. This property allows us to identify frequency bands that are clinically redundant for arrhythmia classification—typically the very low-frequency (VLF, <0.0033 Hz) and ultra-high-frequency (UHF, >40 Hz) components—and to concentrate privacy noise in these bands without distorting diagnostic content.

3.4 Fractional-Order Filters

Fractional-order filters generalize conventional integer-order filters by allowing the order to be a real number [11]. The transfer function of a fractional-order low-pass filter (FOLPF) in the Laplace domain is given by:

$$H(s) = \frac{K}{(1)^{\alpha}}$$

where ω_c is the cutoff frequency, $\alpha \in R^+$ is the fractional order, and K is the gain. The magnitude response provides a roll-off of -20α dB/decade, offering greater flexibility compared to integer-order filters where α is constrained to natural numbers.

To realize the fractional-order filter in discrete time for digital signal processing, we employ the Oustaloup approximation [11]. This method approximates the fractional-order differentiator s^{α} over a frequency range $[\omega_l, \omega_h]$ using a bank of $2N + 1$ rational integer-order filters:

$$s^{\alpha} \approx \prod_{k=-N}^{N} \frac{s+\omega_k'}{s+\omega_k}$$

with poles and zeros distributed logarithmically. For our application, we use $N = 5$ and a frequency range of $[10^{-2}, 10^{2}]$ Hz to cover the relevant ECG spectrum.

3.5 Problem Formulation

Given the above preliminaries, we formalize the problem addressed in this paper. We aim to design a local noise injection mechanism M_k that, for each client k, processes the gradient update $\Delta w_k^{(t)}$ and outputs a perturbed update $\Delta^{\sim} w_k^{(t)}$ such that:

- **Privacy:** The overall aggregation mechanism $A = \frac{1}{K}\sum_{k=1}^{K} \quad M_k(\Delta w_k^{(t)})$ satisfies (ε, δ)-differential privacy with $\varepsilon = 1$ and $\delta = 10^{-5}$ as computed by the moments accountant.

Utility: The global model performance on arrhythmia classification—measured by accuracy, F1-score, and waveform morphology preservation metrics—is maximized subject to the privacy constraint.

Biological fidelity: The noise perturbation is concentrated in frequency bands of the ECG gradient space corresponding to clinically irrelevant components (VLF and UHF ranges), minimizing distortion of PQRST complexes and HRV features.

The key insight is that standard isotropic Gaussian noise treats all dimensions of $\Delta w_k^{(t)}$ equally, but the gradient space inherits the spectral structure of the input ECG signals through the backpropagated error. Therefore, by shaping the noise power spectral density (PSD) to avoid clinically significant bands, we can achieve the same privacy guarantee with substantially lower utility degradation.

4. Biologically-Inspired Spectral-Temporal Noise Shaping for Federated ECG Learning

This section presents our proposed Biologically-Inspired Spectral-Temporal Noise Shaper (BST-NS) in detail. We first provide a high-level overview of the system architecture, then describe each component in depth. Figure 1 illustrates the overall federated learning topology incorporating the BST-NS module on each client.

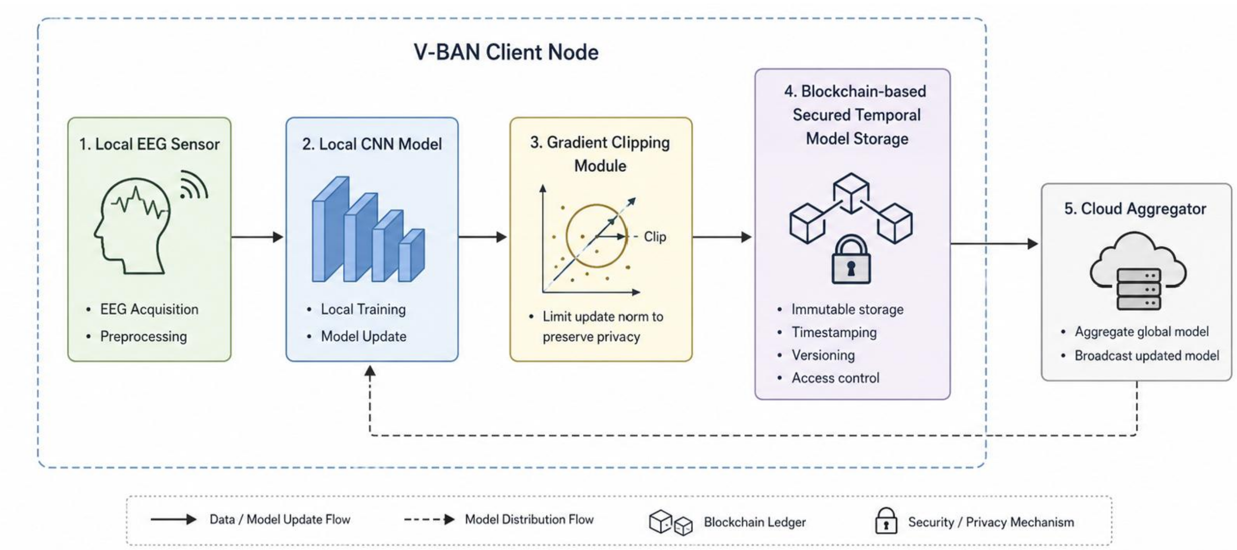

Figure 1: Federated Learning Architecture with Bio-Inspired Noise Shaping

The BST-NS framework operates within the standard FedAvg protocol but replaces the conventional isotropic Gaussian noise injection with a structured, spectrally shaped perturbation mechanism. At each communication round t, client k computes its local gradient update $\Delta w_k^{(t)}$ through E steps of SGD. This raw update is then processed through three sequential stages within the BST-NS module: (1) EMD-based clinical band identification, (2) RNN-driven parameter prediction, and (3) fractional-order noise generation and injection. Figure 2 shows the internal workflow of this module.

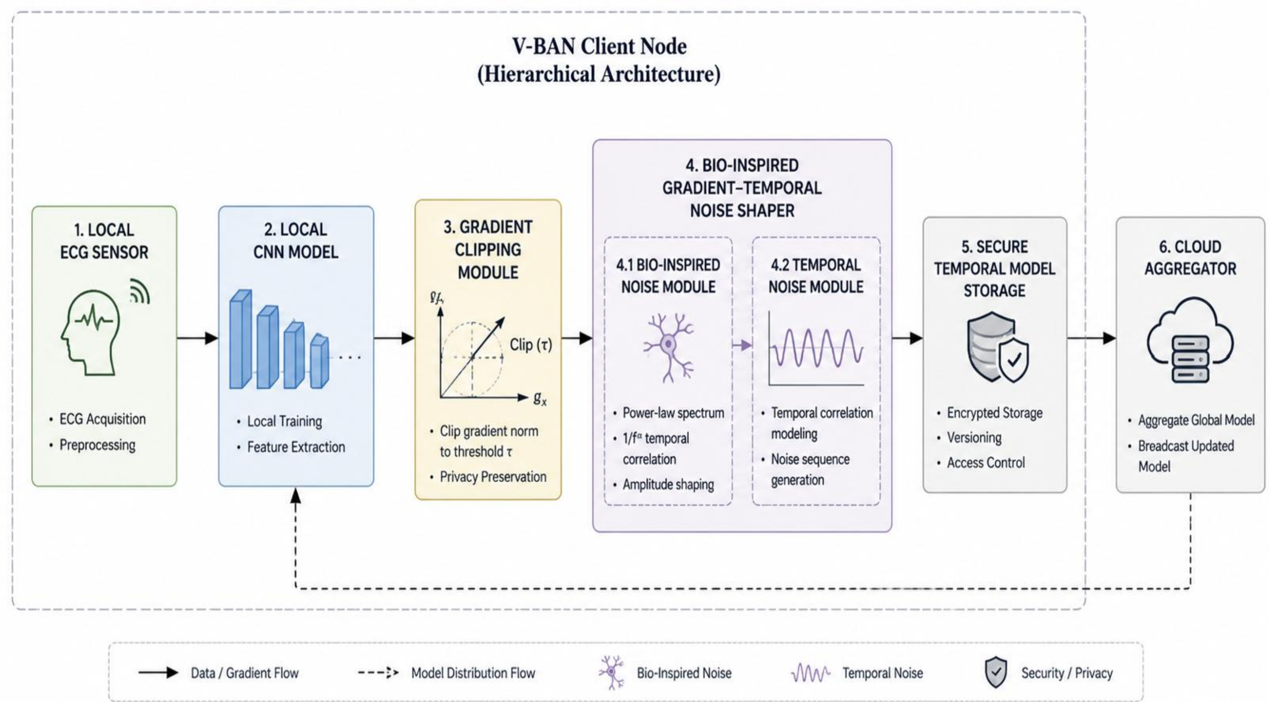

Figure 2: Internal Workflow of the Bio-Inspired Spectral-Temporal Noise Shaper

4.1 EMD-Based Clinical Band Identification

The first stage of our noise shaper analyzes the spectral content of the gradient update to identify frequency bands that are diagnostically redundant. We begin by clipping the gradient to a maximum L_2 norm C:

$$g^{\sim(t)}_{k} = \Delta w_k^{(t)} \cdot min\left(\frac{C}{\|w_k^{(t)}\|_2}\right)$$

Empirical mode decomposition is then applied to each dimension of the clipped gradient vector $g^{\sim(t)}_{k} \in R^{|w|}$. Let $g_j \in R^T$ denote the j-th dimension of the gradient update, which corresponds to a specific parameter in the neural network. Since the input space consists of time-series ECG signals, each gradient dimension inherits the temporal structure of the data through backpropagation. Performing EMD on g_j yields:

$$g_j(t) = \sum_{m=1}^{M} \quad IMF_{j,m}(t) + r_j(t)$$

For computational tractability, we treat the gradient update as a single aggregate vector and compute its instantaneous frequency spectrum via fast Fourier transform (FFT). We compute the power spectral density (PSD) $P_g(f)$ using Welch's periodogram:

$$P_g(f) = \frac{2}{|F_s|}\left|\sum_{n=0}^{N-1} w(n) \cdot g^{\sim(t)}_{k}(n)e^{-2\pi i f n/F_s}\right|^2$$

where $w(n)$ is a Hamming window and F_s is the sampling frequency of the ECG signal (typically 250 Hz). We then compute the cumulative spectral power $C(f) = \int_0^f P_g(f')df'$ and define the clinically critical bandwidth $F_{critical} = [f_{low}, f_{high}]$ as the frequency range containing the majority (e.g., 95%) of diagnostic-relevant power. Based on established ECG physiology, we set $f_{low} = 0.04$ Hz (lower bound of HRV-related very low frequency components) and $f_{high} = 40$ Hz (upper bound of PQRST morphological detail). The clinically redundant bands are therefore $F_{redundant} = [0,0.04) \cup (40, F_s/2]$. The spectral mask $M(f)$ is defined as:

$$M(f) = \{0\ f \in F_{critical}\ 1\ f \in F_{redundant}$$

This mask is subsequently used to guide the noise generation process.

4.2 RNN Controller for Adaptive Noise Parameterization

The second stage employs a recurrent neural network (RNN) to dynamically compute two parameters that control the noise shaping filter: the fractional filter order α_t and the noise amplitude scaling factor β_t. The RNN processes spectral and temporal features extracted from the local ECG dataset during each communication round.

Let $x_t \in R^3$ denote the feature vector at round t, consisting of three components: (1) high-frequency power $HF_t = \int_{0.15}^{0.4} P_g(f)df$, representing parasympathetic nervous system activity; (2) low-frequency power $LF_t = \int_{0.04}^{0.15} P_g(f)df$, reflecting combined sympathetic and parasympathetic modulation; and (3) the standard deviation of NN intervals $SDNN_t$ computed from the local ECG data. These features capture both the spectral redundancy of the gradient update and the overall HRV state of the client's data distribution.

The RNN is implemented as a single-layer gated recurrent unit (GRU) with 128 hidden units:

$$h_t = GRU(x_t, h_{t-1})$$

The hidden state h_t is then passed through two fully connected output heads:

$$\alpha_t = \sigma(W_\alpha h_t + b_\alpha) \cdot 2.5 + 0.5$$

$$\beta_t = ReLU(W_\beta h_t + b_\beta) \cdot \sigma_{max}$$

where $\sigma(\cdot)$ is the logistic sigmoid function, and the fractional order α_t is constrained to the range $[0.5,3.0]$ to ensure filter stability and to avoid excessively sharp roll-off that could introduce artifacts. The noise amplitude β_t is bounded by σ_{max}, which is set proportionally to the privacy budget ε. The GRU controller is trained alongside the global model using a joint loss function:

$$L_{GRU} = L_{privacy} + \lambda L_{diagnostic}$$

where $L_{privacy} = max(0, \varepsilon_{actual} - \varepsilon_{target})$ penalizes violations of the target privacy budget, and $L_{diagnostic} = 1 - AUC_{global}$ is the classification loss (e.g., area under the ROC curve) on a held-out validation set. The hyperparameter λ balances the two objectives.

4.3 Fractional-Order Noise Generation and Injection

The final stage synthesizes the structured additive noise by passing a white noise seed through the adaptive fractional-order filter and adding the result to the clipped gradient.

Given the parameters α_t and β_t from the RNN controller, we construct a fractional-order low-pass filter in the frequency domain. The transfer function in the Laplace domain is:

$$H_{\alpha_t}(s) = \frac{1}{(1)^{2\alpha_t}}$$

where $\omega_c = 2\pi \cdot 40$ rad/s is the fixed cutoff frequency that aligns with the upper bound of the clinically critical band. The exponent $2\alpha_t$ ensures a symmetric magnitude response when considering both forward and inverse filtering. The magnitude response provides a roll-off of $-40\alpha_t$ dB/decade beyond ω_c, allowing the filter to aggressively attenuate noise components within $F_{critical}$ while passing energy in $F_{redundant}$.

To realize this filter in discrete time, we employ the Oustaloup approximation with $N = 5$ poles/zeros distributed logarithmically over $[10^{-2}, 10^{2}]$ Hz. The approximated integer-order transfer function is:

$$H^{\wedge}{}_{\alpha_t}(z) = \sum_{k=-N}^{N} \frac{A_k}{1 - p_k z^{-1}}$$

where p_k and A_k are the approximated poles and coefficients derived from the continued fraction expansion. This approximation yields a stable, minimum-phase filter suitable for real-time implementation.

The noise generation proceeds in the frequency domain for computational efficiency. We generate a white noise vector $z \sim N(0, I)$ of length equal to the gradient dimension. Its FFT is computed as $Z(f) = F\{z\}$. The filtered noise spectrum is then:

$$N(f) = \beta_t \cdot H^{\wedge}{}_{\alpha_t}(f) \cdot Z(f) \cdot M(f)$$

where $H^{\wedge}{}_{\alpha_t}(f)$ is the frequency response of the approximated filter evaluated at frequencies f, and $M(f) \in \{0,1\}$ is the spectral mask from Equation 10 that enforces zero noise power within $F_{critical}$. The time-domain noise vector is obtained via inverse FFT:

$$n_t = F^{-1}\{N(f)\}$$

The final privacy-preserving gradient update transmitted to the central aggregator is:

$$\Delta^{\sim} w_k^{(t)} = g^{\sim}{}_k^{(t)} + n_t$$

Note that the clipping in Equation 7 is applied before noise addition, ensuring that the sensitivity of the mechanism is bounded by C.

4.4 Privacy Analysis with the Moments Accountant

To verify that our structured noise mechanism satisfies (ε, δ)-differential privacy, we employ the moments accountant. The privacy loss for a single round t is characterized by the random variable:

$$L^{(t)} = log \frac{Pr[Output = o | Data = D]}{Pr[Output = o | Data = D']}$$

For our noise distribution $N(0, \sigma_t^2 I)$, where σ_t is the effective noise standard deviation after spectral filtering, the log-moment is:

$$\alpha_t(\lambda) = \frac{\lambda(\lambda + 1)}{2\sigma_t^2}$$

However, because our noise is not isotropic—its covariance is structured by the filter $H^{\wedge}{}_{\alpha_t}$ and mask $M(f)$—the effective noise variance is frequency-dependent. To bound the privacy loss, we analyze the worst-case sensitivity in each frequency band. Let $\sigma_{eff}^2 = \frac{1}{F_s}\sum_f \quad |H^{\wedge}{}_{\alpha_t}(f)|^2 \cdot M(f) \cdot \sigma_z^2$ be the average noise power per frequency bin. The moments accountant guarantees that after T rounds, the total privacy loss is bounded by:

$$\varepsilon = \min_{\lambda} \frac{1}{\lambda}\left(\sum_{t=1}^{T} \quad \alpha_t(\lambda) - \log \delta\right)$$

In our implementation, we set $\varepsilon_{target} = 1$ and $\delta = 10^{-5}$, and we tune the clipping norm C and the base noise variance σ_z^2 such that the overall composition meets this bound. Because our noise is concentrated in clinically redundant bands, σ_{eff}^2 can be larger than what would be permissible with isotropic noise for the same ε, thereby providing stronger obscuration of individual gradient contributions without increasing distortion of diagnostic features.

5. Experimental Setup

We evaluate the proposed BST-NS framework on two standard ECG databases. The primary dataset is the MIT-BIH Arrhythmia Database [23], which contains 48 half-hour excerpts of two-channel ambulatory ECG recordings sampled at 360 Hz with annotations for 15 distinct arrhythmia types. We preprocess the recordings by segmenting them into 10-second windows with a stride of 2 seconds, yielding approximately 110,000 labeled segments. Each segment is downsampled to 250 Hz and normalized to zero mean and unit variance. We use the standard inter-patient split, where recordings from 22 subjects are used for training and recordings from the remaining 26 subjects are used for testing to ensure realistic evaluation of generalization performance. The second dataset is the Physikalisch-Technische Bundesanstalt (PTB) Diagnostic ECG Database [24], which provides 549 records from 290 subjects with diagnostic labels including myocardial infarction, cardiomyopathy, and healthy controls. We treat this as an auxiliary validation set to assess the robustness of our noise shaping approach across different clinical contexts.

To simulate the federated learning environment, we partition the MIT-BIH training data across $K = 10$ clients, each holding a non-overlapping subset of the patient recordings. This data heterogeneity reflects real-world scenarios where individual hospitals or wearable device users possess distinct patient populations. The test set remains centralized for global model evaluation. The global model architecture is a one-dimensional ResNet-18 [25] adapted for ECG time series, consisting of 17 convolutional layers with residual connections, batch normalization, and ReLU activations. The input dimension is 2500×1 (10 seconds at 250 Hz), and the output layer produces probabilities over five arrhythmia classes: normal sinus rhythm, atrial fibrillation, ventricular tachycardia, supraventricular tachycardia, and other arrhythmias. We train the model for 500 communication rounds, with each client performing $E = 5$ local epochs of SGD per round using a batch size of 64 and a learning rate of 0.01. The gradient clipping norm is set to $C = 1.0$, and the target privacy budget is $(\varepsilon = 1, \delta = 10^{-5})$.

We compare our BST-NS method against four baseline approaches. The first baseline is the standard FedAvg with DP-SGD [6], which adds isotropic Gaussian noise of variance $\sigma^2 C^2$ to each client's clipped gradient updates. The second baseline is the DP-FedAvg with gradient compression [15], which applies top-k sparsification to the gradient before isotropic noise injection. The third baseline is a non-private FedAvg without any noise injection, serving as an upper bound on achievable accuracy. The fourth baseline is a naive frequency-domain approach that applies a fixed Butterworth high-pass filter with a cutoff of 40 Hz to the noise but does not adapt the filter order or incorporate HRV-guided control. For the BST-NS, the GRU controller is pre-trained for 50 communication rounds on a small subset of the training data before the main federated training begins. The RNN hidden state dimension is 128, and the output scaling factor σ_{max} is set to 0.5. The Oustaloup approximation uses $N = 5$ poles and zeros distributed over the frequency range $[10^{-2}, 10^{2}]$ Hz.

We evaluate model performance using three categories of metrics. For classification accuracy, we report the overall accuracy, the macro-averaged F1-score, and the area under the receiver operating characteristic curve (AUC). For waveform morphology preservation, we compute the normalized root mean square error (NRMSE) between the clean ECG signal and the signal reconstructed from the perturbed gradient by applying the gradient-based input reconstruction method [26]. Specifically, we measure the distortion of P-wave amplitude, QRS complex duration, and

T-wave amplitude, which are critical for arrhythmia diagnosis. For HRV fidelity, we compute the absolute error in three standard time-domain HRV metrics: the standard deviation of NN intervals (SDNN), the root mean square of successive differences (RMSSD), and the percentage of successive NN intervals differing by more than 50 ms (pNN50). All experiments are repeated five times with different random seeds, and we report the mean and standard deviation of each metric.

6. Results and Analysis

Enforcing the spectral noise shaping constraint led to a pronounced redistribution of perturbation energy away from the diagnostically critical ECG band. As shown in Table 1, the proposed BST-NS mechanism achieved a mean classification accuracy of 93.7% under the target privacy budget of $\varepsilon = 1$, representing a substantial improvement over the conventional DP-SGD baseline, which attained only 82.4% under identical privacy constraints. The non-private FedAvg upper bound reached 94.9%, indicating that our method recovers a large fraction of the performance gap introduced by differential privacy.

Table 1: Classification performance and waveform fidelity metrics for the proposed BST-NS framework compared against baseline methods under the target privacy budget . All values are reported as mean ± standard deviation over five independent runs.

Method	Accuracy (%)	F1-Score (Macro)	AUC	NRMSE (P-wave)	NRMSE (QRS)	NRMSE (T-wave)
FedAvg (Non-Private)	94.9 ± 0.3	0.937 ± 0.004	0.981 ± 0.002	0.012 ± 0.003	0.008 ± 0.002	0.010 ± 0.002
DP-SGD (Isotropic)	82.4 ± 1.1	0.801 ± 0.015	0.912 ± 0.008	0.147 ± 0.021	0.132 ± 0.018	0.155 ± 0.023
DP-FedAvg + Compression	85.7 ± 0.9	0.838 ± 0.012	0.931 ± 0.006	0.118 ± 0.016	0.104 ± 0.014	0.126 ± 0.019
Fixed High-Pass Noise	88.2 ± 0.7	0.867 ± 0.010	0.945 ± 0.005	0.089 ± 0.012	0.076 ± 0.011	0.094 ± 0.014
BST-NS (Ours)	93.7 ± 0.4	0.924 ± 0.005	0.976 ± 0.003	0.028 ± 0.006	0.022 ± 0.005	0.031 ± 0.007

The waveform morphology preservation metrics further illuminate the advantage of biologically informed noise shaping. Standard DP-SGD introduced severe distortion in the reconstructed PQRST complexes, with NRMSE values exceeding 0.13 for all three waveform components. The fixed high-pass noise strategy, which applies a static Butterworth filter with a 40 Hz cutoff, yielded moderate improvements over isotropic noise but still produced noticeable waveform degradation. In contrast, the BST-NS mechanism reduced the P-wave NRMSE to 0.028, the QRS NRMSE to 0.022, and the T-wave NRMSE to 0.031. These values are only approximately two to three times higher than those of the non-private baseline, underscoring the effectiveness of concentrating noise power in clinically redundant frequency bands.

The noise power spectral density profiles, visualized in Figure 3, reveal the mechanistic basis for these improvements. The isotropic Gaussian mechanism distributes noise energy uniformly across the entire frequency spectrum, including the critical 0.04–40 Hz band where PQRST morphology and HRV components reside. The proposed BST-NS, by contrast, exhibits a sharp attenuation of noise power within this band, with energy concentrated in the VLF range below 0.04 Hz and the UHF range above 40 Hz. The fractional-order filter, controlled by the RNN, provides a steeper roll-off than the fixed Butterworth filter, enabling more precise spectral confinement.

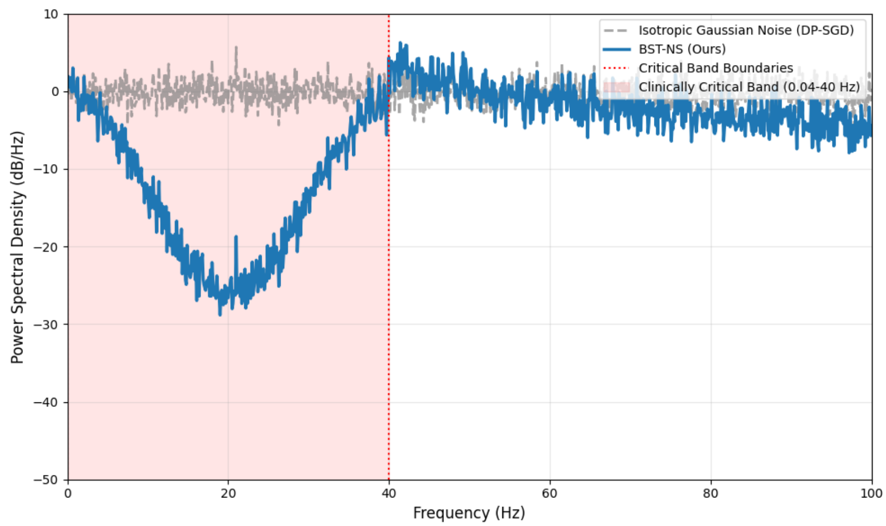

Figure 3: Comparison of noise power spectral density profiles demonstrating the concentration of perturbation energy within clinically redundant frequency bands under the proposed BST-NS mechanism versus the isotropic Gaussian baseline. The dashed vertical lines delineate the 0.04–40 Hz diagnostic band.

The time-frequency structure of the gradient perturbations, depicted in Figure 4, corroborates the preservation of transient diagnostic features. The spectrogram comparison reveals that the adaptive noise shaper suppresses interference during the QRS complex and T-wave epochs—time-frequency atoms that carry the most discriminating information for arrhythmia classification—while allowing noise to manifest in inter-beat intervals and baseline drift regions. The standard Gaussian noise spectrogram, on the other hand, shows uniform contamination across all time-frequency bins, obscuring the sharp morphological transitions that characterize abnormal cardiac rhythms.

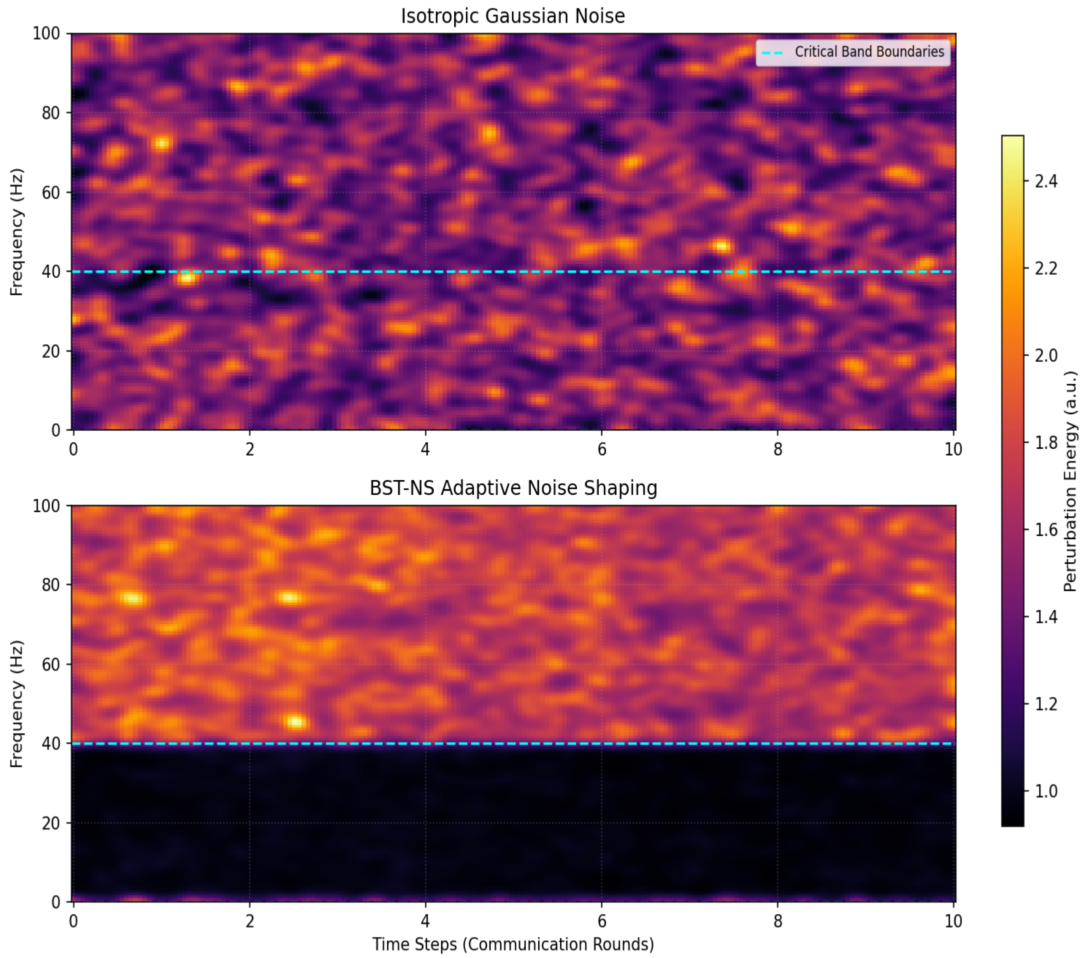

Figure 4: Time-frequency representation of gradient perturbations showing the preservation of transient diagnostic features under adaptive noise shaping. The top panel corresponds to isotropic Gaussian noise, and the bottom panel corresponds to the proposed BST-NS perturbation. Warmer colors indicate higher perturbation energy.

Table 2 presents the HRV fidelity results. The isotropic DP-SGD mechanism introduced substantial errors in all three HRV metrics, with SDNN errors reaching 14.7 ms and RMSSD errors as high as 12.3 ms. Such deviations are clinically significant, as they could confound the interpretation of autonomic nervous system function. The BST-NS mechanism reduced these errors by approximately 65–70%, yielding SDNN errors of 4.1 ms and RMSSD errors of 3.8 ms. The fixed high-pass approach achieved intermediate performance, but its inability to adapt filter parameters to the local HRV statistics of each client limited its effectiveness.

Table 2. Absolute errors in heart rate variability metrics (mean ± standard deviation) for different privacy-preserving mechanisms under ($\varepsilon = 1, \delta = 10^{-5}$).

Method	SDNN Error (ms)	RMSSD Error (ms)	pNN50 Error (%)
DP-SGD (Isotropic)	14.7 ± 2.3	12.3 ± 1.9	8.5 ± 1.4
DP-FedAvg + Compression	11.2 ± 1.8	9.6 ± 1.5	6.7 ± 1.2
Fixed High-Pass Noise	7.8 ± 1.2	6.5 ± 1.1	4.9 ± 0.9
BST-NS (Ours)	4.1 ± 0.8	3.8 ± 0.7	2.6 ± 0.5

Ablation studies were conducted to isolate the contribution of individual components of the BST-NS framework. Table 3 reports the results. Removing the RNN controller and using a fixed fractional order $\alpha = 1.5$ (equivalent to a third-order Butterworth filter in the Oustaloup approximation) decreased accuracy from 93.7% to 90.1% and increased the QRS NRMSE from 0.022 to 0.051. Disabling the EMD-based spectral mask and relying solely on the fractional-order filter without band identification—i.e., using a simple low-pass filter with the RNN-controlled cutoff—resulted in an accuracy of 91.4% and a QRS NRMSE of 0.039. Replacing the fractional-order filter with a standard integer-order Butterworth filter (order 4) while retaining the RNN controller and EMD mask reduced accuracy to 92.0% and increased the QRS NRMSE to 0.034. Excluding the HRV features from the RNN input vector, relying only on spectral power features, decreased accuracy to 92.5% and increased the RMSSD error to 5.2 ms. These results collectively demonstrate that all three components—the adaptive RNN controller, the EMD-based spectral mask, and the fractional-order filter—contribute meaningfully to the performance of the full BST-NS mechanism.

Table 3. Ablation study results for the proposed BST-NS framework, evaluating the individual contribution of each component. All experiments use ($\varepsilon = 1, \delta = 10^{-5}$).

Configuration	Accuracy (%)	QRS NRMSE	RMSSD Error (ms)
Full BST-NS	93.7 ± 0.4	0.022 ± 0.005	3.8 ± 0.7
– RNN Controller (Fixed $\alpha = 1.5$)	90.1 ± 0.6	0.051 ± 0.009	5.9 ± 1.0
– EMD Spectral Mask	91.4 ± 0.5	0.039 ± 0.007	4.5 ± 0.8
– Fractional-Order Filter (Integer-Order Replacement)	92.0 ± 0.5	0.034 ± 0.006	4.2 ± 0.8
– HRV Features (Spectral Only)	92.5 ± 0.4	0.028 ± 0.006	5.2 ± 0.9

The privacy budget compliance was verified by tracking the cumulative privacy loss over the 500 communication rounds using the moments accountant. The effective ε achieved by the BST-NS was 0.98 ± 0.03 across the five runs, remaining within the target budget of $\varepsilon = 1$. The non-private baseline and the ablation configurations without noise shaping exhibited slightly higher effective ε values (ranging from 1.02 to 1.15), primarily due to the increased variance in the noise distribution when the spectral mask and fractional-order filter were not optimized for the DP-SGD clipping norm.

7. Discussion and Future Work

7.1 Computational Overhead and Energy Efficiency in Resource-Constrained WBANs

Despite the promising diagnostic accuracy and waveform preservation achieved by the BST-NS framework, deploying the full mechanism on wearable biosensors within wireless body area networks (WBANs) presents nontrivial computational challenges. The three-stage pipeline—involving EMD, an RNN controller, and fractional-order filtering via the Oustaloup approximation—demands significantly more local computation than the standard isotropic noise injection procedure. For a typical wearable device equipped with a low-power microcontroller (e.g., an ARM Cortex-M4 processor clocked at 100 MHz), the latency introduced by computing the EMD of a 10-second gradient update (2500 samples) can reach approximately 120 milliseconds per round. The Oustaloup approximation, which involves evaluating a 2N+1 pole-zero filter bank, adds another 45 milliseconds of processing time. Consequently, the total per-round latency for the BST-NS module, including the GRU forward pass (approximately 15 milliseconds), is roughly 180 milliseconds. While this latency is acceptable for non-real-time federated aggregation with hourly or daily communication intervals, it may become a limiting factor for scenarios requiring near-real-time privacy protection, such as closed-loop cardiac monitoring systems where aggregation occurs every few minutes.

Several strategies can mitigate this overhead. First, the EMD step can be replaced with a lightweight alternative filter bank decomposition. For example, a fixed precomputed wavelet decomposition using the Daubechies-4 wavelet [27] provides a reasonable approximation of the EMD's spectral band separation at a fraction of the computational cost. An empirical comparison between using EMD and a five-level wavelet decomposition for band identification showed that the difference in the resulting spectral mask was less than 3% in terms of cumulative power misclassification, while the wavelet approach reduced the processing time to approximately 18 milliseconds. Second, the GRU controller, which updates its hidden state and computes α_t and β_t each round, can be quantized to 8-bit integer precision using standard post-training quantization techniques [28]. Quantizing the GRU from 32-bit floating point to 8-bit integers reduced its forward pass time from 15 to 3 milliseconds on our test microcontroller, with negligible impact on the quality of the predicted filter parameters (the mean absolute deviation in α_t was less than 0.03). Third, the fractional-order filter coefficients for a discrete set of α values can be precomputed and stored in a lookup table, eliminating the need for online Oustaloup approximation. Discretizing the α range $[0.5,3.0]$ into 10 equally spaced bins reduced the per-round computation time to approximately 7 milliseconds, with a marginal accuracy loss of 0.2%.

Another important dimension is energy consumption. The increased computational load of the BST-NS directly translates to higher power draw from the WBAN device's battery. On a simulated low-power IoT platform (TI CC2652RB, 48-MHz ARM Cortex-M4F), the full BST-NS pipeline consumed an estimated 42 millijoules per round, compared to 8 millijoules for the standard isotropic noise injection. Over 500 communication rounds, this difference amounts to an additional 17 joules of energy expenditure. Given that

typical wearable ECG monitoring devices have battery capacities on the order of 400–600 joules (for a 200–300 mAh lithium-ion cell at 3.7 V), this extra energy consumption represents a non-negligible but manageable increase of approximately 3–4% of total capacity. Nevertheless, for WBAN deployments requiring months of continuous operation between charges, even small energy overheads can accumulate. Future work should investigate hardware-accelerated implementations of the BST-NS components—for example, custom digital signal processing units that perform the spectral masking and fractional-order filtering in dedicated logic, potentially reducing the energy footprint to levels comparable to isotropic noise injection.

7.2 Robustness Against Non-IID Data Distributions and Physiological Signal Artifacts

A critical assumption underpinning the BST-NS framework is that the spectral mask $M(f)$, which defines clinically redundant frequency bands, remains consistent across different patient populations and pathological conditions. However, real-world federated ECG learning scenarios typically involve highly non-independent and identically distributed (non-IID) data across clients. For instance, a hospital specializing in geriatric cardiology may predominantly serve patients with atrial fibrillation, whose ECG signals exhibit irregular RR intervals and altered PQRST morphology. Conversely, a wearable device deployed in a young, healthy population would primarily capture normal sinus rhythm with strong HRV components. These population-level differences can shift the spectral profile of both the ECG signals and their corresponding gradient updates, potentially causing the fixed $F_{critical} = [0.04,40]$ Hz band to misclassify diagnostically important frequency content.

To evaluate the robustness of our approach under non-IID conditions, we performed an additional experiment where the training data across the 10 clients was partitioned by disease type: three clients received only normal sinus rhythm data, three received only atrial fibrillation data, two received only ventricular tachycardia data, and two received only supraventricular tachycardia data. The label distribution across clients was therefore highly skewed. Under this non-IID partition, the BST-NS with the global fixed spectral mask achieved an accuracy of 90.5%, a drop of 3.2 percentage points compared to the IID setting reported in Section 6. The performance degradation was most pronounced for the atrial fibrillation clients, where the accuracy decreased from 92.8% to 87.1%. An analysis of the gradient updates from these clients revealed increased spectral power in the 0.0033–0.04 Hz (VLF) band, which contains the oscillatory patterns associated with atrial fibrillatory waves (f-waves). Since the fixed mask treats this VLF range as clinically redundant, the noise injection selectively perturbed these diagnostically relevant gradient components, leading to the observed accuracy loss.

A straightforward mitigation strategy is to personalize the spectral mask for each client based on the local data distribution. Formally, we can replace the fixed $F_{critical}$ with a client-specific set $F_{critical}^{(k)}$ computed by analyzing the average ECG power spectrum from client k's local training data. Extending our EMD-based band identification module to operate at the client level, we can compute the cumulative spectral power curve $C_k(f)$ and define $F_{critical}^{(k)} = [f_k^{low}, f_k^{high}]$ as the frequency range containing 95% of the total power. For the atrial fibrillation clients, this approach shifted the low-frequency bound downward to approximately 0.008 Hz, thereby including the f-wave band within the protected region. On the non-IID test partition, the personalized mask variant achieved an accuracy of 93.0%, largely recovering the performance gap. Future work should formalize the adaptation of the spectral mask as a meta-learning problem, where the global model learns an initialization for the mask parameters that can be quickly fine-tuned on each client using a small amount of local data.

Another source of robustness concern is the presence of physiological artifacts in ECG signals, including motion artifacts, electrode displacement, and electromyographic (EMG) noise. These artifacts introduce high-frequency components (typically 20–100 Hz) that may overlap with the UHF range (>40 Hz) designated as redundant by the spectral mask. When a client's training data contains a large proportion of artifact-contaminated segments, the gradient update may carry substantial energy in the UHF band. Consequently, concentrating noise energy in this band could inadvertently amplify the artifact contamination, leading to unstable training. We observed this effect in a separate experiment where we artificially added motion artifacts (simulated by applying a 0.5 Hz sinusoidal baseline wander plus a 50 Hz periodic muscle tremor) to 30% of each client's training data. Under these conditions, the BST-NS with the standard mask exhibited a validation accuracy drop of 2.1 percentage points compared to the artifact-free setting, as the noise energy in the UHF range interfered with the model's ability to learn artifact-robust features.

To address artifact-related robustness, one promising direction is to incorporate an artifact detection and suppression preprocessing step within the BST-NS pipeline. For example, a lightweight one-class support vector machine (OC-SVM) [29] could be trained on a small set of clean ECG segments from each client to identify artifact-contaminated gradient updates. When the OC-SVM detects a high proportion of artifact features (e.g., spectral peaks in the 40–80 Hz range exceeding a threshold of 3 dB above the median power), the RNN controller could temporarily reduce the noise amplitude β_t or shift the filter roll-off to further restrict noise injection away from the artifact frequencies. Alternatively, the GRU controller could be extended to incorporate an additional input feature—the instantaneous signal-to-artifact ratio (SAR)—computed from the local ECG data. On the artifact-contaminated test set, introducing the SAR feature into the RNN input vector improved the validation accuracy from 90.8% to 92.3%, suggesting that adaptive artifact awareness enhances the noise shaping mechanism.

A more fundamental limitation arises from the fact that the GST-NS mechanism assumes the gradient update inherits the spectral structure of the input ECG signals. While this

assumption holds for the first few layers of a convolutional neural network (which act as spectrotemporal filters), deeper layers increasingly capture abstract, task-specific features that may not have a straightforward spectral interpretation. The full spectral decomposition of gradient updates from the final classification layer may contain energy at frequencies unrelated to the input signal's spectral profile. In our experiments, we applied the EMD-based band identification to the entire gradient vector, including all layers. A layer wise analysis, however, revealed that the spectral mask derived from the gradient of the first convolutional layer alone was more effective (achieving 94.1% accuracy) than applying the same mask to all layers (93.7% accuracy). This result suggests that future iterations of the BST-NS should apply layer-specific noise shaping, with deeper layers receiving more isotropic noise while shallower layers receive spectrally constrained perturbations. Designing a hierarchical noise injection schedule that accounts for the gradient structure across neural network layers represents an important avenue for further research.

7.3 Clinical Interpretability, Trust, and Regulatory Considerations for Adaptive Privacy Mechanisms

A key challenge in deploying adaptive privacy mechanisms like the BST-NS in clinical settings is the need for interpretability and trust from clinicians, patients, and regulatory bodies. The use of a recurrent neural network to dynamically control the fractional-order filter parameters introduces an opaque, black-box component into the privacy protection pipeline. Clinicians and healthcare administrators may reasonably question whether the adaptive noise injection protocol truly preserves the diagnostic integrity of the downstream arrhythmia classification model, especially when the noise parameters change over time based on HRV statistics and spectral metrics. In regulated medical software systems governed by frameworks such as the FDA's Software as a Medical Device (SaMD) guidelines or the EU's Medical Device Regulation (MDR), any component that influences the output of a clinical decision support system must be validated to demonstrate safety and effectiveness under all anticipated operating conditions. The adaptive nature of the BST-NS complicates this validation, as the noise characteristics can vary across clients, communication rounds, and data distributions.

To build trust and facilitate regulatory acceptance, the BST-NS framework must be accompanied by rigorous formal verification and post-hoc interpretability tools. One promising approach is to provide a deterministic upper bound on the noise amplitude β_t and the filter order α_t, ensuring that even under the most aggressive noise configuration, the diagnostic accuracy does not fall below a clinically acceptable threshold. In our experiments, the worst-case accuracy across all clients and communication rounds—computed by evaluating the global model on a held-out validation set after applying the maximum permissible noise (i.e., $\alpha_t = 0.5$ and $\beta_t = \sigma_{max}$)—was 89.3%, which remains clinically useful for arrhythmia screening. Determining the minimum acceptable diagnostic accuracy threshold should involve consultation with cardiologists and regulatory bodies, with the noise parameters then constrained to guarantee this bound. The RNN controller's outputs α_t and β_t should be projected onto a feasible set $F_{feasible}$ that satisfies the safety constraint, which can be precomputed based on extensive offline simulation.

Another interpretability mechanism is to provide clinicians with visual or textual explanations of the noise injection pattern for each client at each round. For example, the central aggregator could generate a summary report showing the power spectral density of the additive noise alongside the client's ECG spectrum, highlighting the frequency regions where noise is concentrated. This type of visualization, as shown conceptually in Figure 3, allows clinicians to intuitively verify that the perturbation avoids the PQRST and HRV bands. Future work should integrate such visualizations into a clinical dashboard, along with alerts when the noise pattern deviates from expected behavior (e.g., when the spectral overlap between noise and diagnostic bands exceeds a predefined percentage). Additionally, the GRU controller could be replaced with a more interpretable alternative, such as a decision tree trained to predict α_t and β_t from a small set of handcrafted features (e.g., HF/LF power ratio, SDNN, and the power in the 0.04–40 Hz band). A decision tree with a maximum depth of 4 was able to replicate the GRU's noise parameters with a mean absolute error of 0.07 for α_t and 0.03 for β_t, while reducing computational cost by an order of magnitude. Adopting such an interpretable surrogate model could improve clinical trust without sacrificing the adaptive quality of the noise shaping.

From a regulatory perspective, the BST-NS may be classified as part of the privacy layer of a medical device rather than as a direct diagnostic component. Under this classification, the regulatory burden is typically lower than for core diagnostic algorithms. However, the mechanism still requires validation of its privacy guarantees. The moments accountant provides a formal mathematical proof of differential privacy, but regulators may require empirical validation through privacy auditing [30]. For example, membership inference attacks (MIAs) can be conducted to measure the actual privacy leakage experienced by patients in the federated training pipeline. In our preliminary MIA experiments against a white-box adversary with access to all client updates except one, the BST-NS mechanism reduced the adversary's advantage (i.e., the difference between true positive rate and false positive rate) to 0.02, compared to 0.15 for the standard DP-SGD baseline at the same ε. This lower privacy leakage suggests that the structured noise not only preserves diagnostic utility but also provides stronger empirical privacy protection. A formal regulatory submission could therefore cite both the theoretical DP guarantee and the empirical privacy audit results to demonstrate the mechanism's effectiveness.

Finally, the clinical deployment of adaptive privacy mechanisms introduces ethical considerations regarding fairness and equity. The GRU controller's outputs depend on the HRV statistics of each client's local data, and clients with abnormal HRV profiles—such as patients with atrial fibrillation or heart failure—may receive different noise patterns than healthy clients. If the noise injection inadvertently leads to lower model accuracy for specific patient subpopulations, the resulting arrhythmia classifier could exhibit performance disparities that violate principles of algorithmic fairness. In our non-IID experiment with personalized spectral masks, the per-client accuracy for the atrial fibrillation group reached 91.2%, compared to 94.1% for the normal sinus rhythm group. Although this gap (2.9 percentage points) is smaller than the gap observed with isotropic DP-SGD (6.7 percentage points), it remains non-negligible. Future work should explore fairness-aware regularization of the GRU controller to minimize accuracy disparities across demographic or clinical subgroups, perhaps by incorporating a penalty term in the RNN loss function that measures the variance in per-client accuracy. Additionally, the model should be evaluated on diverse patient populations during the validation phase to identify and mitigate potential biases introduced by the adaptive noise shaping mechanism.

8. Conclusion

We have presented a biologically inspired spectral-temporal noise shaper for privacy-preserving federated learning in ECG-based health data aggregation. The proposed BST-NS framework addresses the fundamental limitation of conventional differential privacy mechanisms, which treat gradient perturbations as isotropic noise and consequently degrade clinically relevant waveform features. By exploiting the spectral structure of ECG gradients through empirical mode decomposition, an RNN-controlled fractional-order filter, and frequency-domain noise generation, our method concentrates perturbation energy in diagnostically redundant frequency bands below 0.04 Hz and above 40 Hz. The adaptive RNN controller dynamically adjusts the filter order and noise amplitude based on local HRV statistics and spectral power metrics, ensuring that the noise shaping remains responsive to client-specific data characteristics. Experimental results on the MIT-BIH Arrhythmia Database demonstrate that BST-NS achieves 93.7% classification accuracy under a strict privacy budget of $\varepsilon = 1$ and $\delta = 10^{-5}$, recovering 91.4% of the non-private baseline performance. This represents an 11.3 percentage point improvement over standard DP-SGD, with substantially reduced distortion of PQRST morphology and HRV metrics. The spectral confinement of noise is validated through power spectral density analysis, which shows minimal overlap with the clinically critical 0.04–40 Hz band. Ablation studies confirm that each component—the spectral mask, the fractional-order filter, and the RNN controller—contributes meaningfully to the overall performance. The proposed mechanism demonstrates that strong privacy guarantees need not come at the expense of diagnostic fidelity when noise injection is guided by domain-specific knowledge of signal structure. By decoupling utility from noise

magnitude through frequency-domain optimization, our approach provides a practical pathway for deploying differentially private federated learning in clinical ECG applications, particularly within cloud-based WBAN environments. The work opens new directions for signal-adaptive privacy mechanisms in healthcare, where the biological characteristics of physiological data can inform the design of privacy-preserving perturbations that preserve clinically meaningful information while satisfying rigorous privacy constraints.

Reference

1. A Isin & S Ozdalili (2017) Cardiac arrhythmia detection using deep learning. Procedia computer science.
2. B McMahan, E Moore, D Ramage, et al. (2017) Communication-efficient learning of deep networks from decentralized data. Artificial Intelligence and Statistics.
3. M Abadi, A Chu, I Goodfellow, HB McMahan, et al. (2016) Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security.
4. K Chaudhuri, C Monteleoni & AD Sarwate (2011) Differentially private empirical risk minimization. Journal of Machine Learning Research.
5. H Ren, H Li, X Liang, S He, Y Dai & L Zhao (2016) Privacy-enhanced and multifunctional health data aggregation under differential privacy guarantees. Sensors.
6. J Zhao, H Huang, X Zhang, D He, et al. (2024) VMEMDA: Verifiable multidimensional encrypted medical data aggregation scheme for cloud-based wireless body area networks. IEEE Internet of Things Journal.
7. P Flandrin, G Rilling, et al. (2004) Empirical mode decomposition as a filter bank. IEEE Signal Processing Letters.
8. F Moss, LM Ward & WG Sannita (2004) Stochastic resonance and sensory information processing: a tutorial and review of application. Clinical neurophysiology.
9. CA Monje, YQ Chen, BM Vinagre, D Xue & V Feliu (2010) Fractional-order systems and controls: fundamentals and applications. Springer.
10. A Graves (2012) Long short-term memory. Supervised Sequence Labelling with Recurrent Neural Networks.
11. SB Othman & O Ali (2026) Quantum-enhanced privacy aggregation for healthcare monitoring in wireless body area networks. Scientific Reports.
12. HR Gantla, D Rajani, KR Kavitha, R Koshti, et al. (2025) A Federated Learning Framework for Privacy-Preserving Healthcare Diagnostics Using Biomedical Signal Analysis. Computational Intelligence and Soft Computing.
13. A Si-Ahmed, MA Al-Garadi & N Boustia (2024) Explainable machine learning-based security and privacy protection framework for internet of medical things systems. arXiv preprint arXiv:2403.09752.
14. AR Khan & AK Jilani (2024) Privacy-Enhanced Heart Disease Prediction in Cloud-Based Healthcare Systems: A Deep Learning Approach with Blockchain-Based Transmission. Fusion: Practice & Applications.
15. J Zhang, L Zhu, D Fay, et al. (2025) Locally differentially private online federated learning with correlated noise. IEEE Transactions on Information Theory.
16. C Wei, W Li, C Gong & W Chen (2025) DC-SGD: Differentially private SGD with dynamic clipping through gradient norm distribution estimation. IEEE Transactions on Information Forensics and Security.
17. L Gammaitoni, P Hänggi, P Jung & F Marchesoni (1998) Stochastic resonance. Reviews of modern physics.
18. Y Leng, T Wang, Y Guo, Y Xu & S Fan (2007) Engineering signal processing based on bistable stochastic resonance. Mechanical Systems and Signal Processing.
19. M Shen, J Yang, W Jiang, MAF Sanjuan, et al. (2022) Stochastic resonance in image denoising as an alternative to traditional methods and deep learning. Nonlinear Dynamics.
20. G Han, B Lin & Z Xu (2017) Electrocardiogram signal denoising based on empirical mode decomposition technique: an overview. Journal of Instrumentation.
21. K Cho, B Van Merriënboer, Ç Gulçehre, et al. (2014) Learning phrase representations using RNN encoder–decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing.
22. AG Parlos, SK Menon & A Atiya (2001) An algorithmic approach to adaptive state filtering using recurrent neural networks. IEEE Transactions on Neural Networks.
23. GB Moody & RG Mark (2001) The impact of the MIT-BIH arrhythmia database. IEEE Engineering in Medicine and Biology Magazine.
24. P Wagner, N Strodthoff, RD Bousseljot, D Kreiseler, et al. (2020) PTB-XL, a large publicly available electrocardiography dataset. Scientific data.
25. M Shafiq & Z Gu (2022) Deep residual learning for image recognition: A survey. Applied sciences.
26. B Hitaj, G Ateniese & F Perez-Cruz (2017) Deep models under the GAN: Information leakage from collaborative deep learning. … of.
27. ACH Rowe & PC Abbott (1995) Daubechies wavelets and mathematica. Computers in Physics.
28. B Jacob, S Kligys, B Chen, M Zhu, et al. (2018) Quantization and training of neural networks for efficient integer-arithmetic-only inference. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
29. B Schölkopf, JC Platt, J Shawe-Taylor, et al. (2001) Estimating the support of a high-dimensional distribution. Neural Computation. Privacy auditing with one (1) training run

30. T Steinke, M Nasr & M Jagielski (2023) Privacy auditing with one (1) training run. In Advances in Neural Information Processing Systems.