Differentially Private ECG Heartbeat Classification

João Volney G. Santos, Felipe T. Brito, Javam C. Machado

Laboratório de Sistemas e Bancos de Dados (LSBD)
Department of Computing / UFC – Fortaleza – CE – Brazil

Abstract. *Automated heartbeat classification models have become crucial for modern cardiac monitoring and the early detection of cardiovascular diseases. However, these models' tendency to memorize sensitive patient data poses a significant privacy risk. Differential Privacy (DP) has emerged as the mathematical foundation for providing formal, provable privacy guarantees during model training, effectively mitigating risks of data memorization. Despite its robustness, directly applying standard DP algorithms, such as DP-SGD, to neural networks often results in a significant drop in classification performance. We address this challenge by introducing DP-ECG-HC, a privatized architecture that leverages a Conformer-based, tokenized approach to extract morphological patterns from raw ECG signals. Experimental results on real-world ECG datasets demonstrate that DP-ECG-HC preserves robust model utility and effectively even under stringent privacy requirements, outperforming standard private deep learning baselines and achieving a superior privacy-utility trade-off.*

1. Introduction

The importance of electrocardiogram (ECG) analysis has become increasingly prominent in clinical practice. Cardiovascular diseases remain the leading cause of mortality among adults worldwide, and the ECG serves as a primary tool for anomaly screening and diagnosis [Zanchi et al. 2025]. Beyond monitoring electrical activity, ECG analysis enables the detection of arrhythmias, myocardial and pericardial disorders, and certain pulmonary conditions [Rafie et al. 2021].

A particularly notable characteristic of ECG signals is their role as a repository of personal data, reflecting the unique geometry of the heart, as well as demographic attributes and physiological and emotional states. As such, ECG data inherently carries highly sensitive information [Pinto and Cardoso 2020]. From this perspective, the classification of individual heartbeats has become essential for identifying potentially life-threatening arrhythmias, such as atrial fibrillation or myocardial infarction, which require early detection to prevent severe outcomes.

In this context, machine learning models make the classification process more efficient and robust. Traditional methods of classification that rely on manual feature extraction by specialists can be accelerated by these models, which can automatically learn hierarchical representations of different types of cardiac anomalies. Furthermore, machine learning–based classification techniques enable scalability for analyzing massive volumes of data at a population scale [Khalid et al. 2023]. Additionally, wearable devices with embedded models can already detect intermittent anomalies that might go unnoticed in conventional clinical exams [Rafie et al. 2021]. In this setting, classification models can serve as unique biometric identifiers, enabling individual bioidentification with only ECG data.

However, the high predictive accuracy that makes these models clinically valuable also introduces critical privacy vulnerabilities. For instance, recent studies have demonstrated that it is possible to accurately re-identify individuals using deep learning models trained on ECG data. Specifically, [Al-Jibreen et al. 2024] introduced a CNN that achieves 99.28% accuracy in heartbeat re-identification. This capability can enable linkage attacks, allowing an adversary to associate anonymous ECG records with their true owners' identities. The consequences of such attacks extend beyond the disclosure of cardiac comorbidities, as highly sensitive information such as daily routines, mood variations, training schedules, and even indicators of pulmonary diseases may also be inferred [Rafie et al. 2021]. In addition, machine learning models may memorize information from the training dataset, and through reconstruction attacks, it is possible to recover training samples, thereby violating user privacy within the dataset [Zhang and Zhang 2025]. Electrocardiogram (ECG) signals are considered sensitive personal data because they contain biometric and health-related information that can reveal an individual's medical condition and history [Vásquez-Iturralde et al. 2024].

According to the Brazilian General Data Protection Law (LGPD), sensitive personal data includes information related to health, genetics, or biometrics when associated with an identifiable individual [Moura et al. 2024]. Similarly, the General Data Protection Regulation (GDPR) classifies health data as a special category of personal data under Article 9, requiring strict safeguards to prevent misuse, discrimination, or harm to the data subject [Phang and Kaabi 2025].

The proposed model combines convolutional layers and attention-based mechanisms to analyze ECG signals. Convolutional layers focus on capturing local waveform patterns, while the attention component models longer temporal relationships within the signal. Global information from the sequence is then summarized into a compact latent representation, which is used by a final classifier to predict the heartbeat category. During training, we employ Focal Loss, a modification of the standard cross-entropy loss that addresses class imbalance by emphasizing difficult samples. In addition, the model is trained using Differentially Private Stochastic Gradient Descent (DP-SGD), which adds controlled noise to the learning process to protect sensitive information in the training data.

The main design principles of the proposed DP-ECG-HC architecture are summarized as follows:

- We propose a hybrid Conformer-based architecture that combines convolutional layers and self-attention with multiple CLS tokens and a dense latent representation. The CLS tokens aggregate global contextual information across the sequence. During training, Focal Loss replaces the standard Cross-Entropy loss to address severe class imbalance by emphasizing hard-to-classify samples. This design enables the joint extraction of local morphological and global rhythmic ECG features, improving the detection of rare pathologies.
- We introduce a DP-SGD training pipeline designed to mitigate biometric reidentification risks, such as heartprint extraction and record-linkage attacks. By compressing information into a latent bottleneck, the model preserves strong clinical utility for multiclass classification while enforcing Differential Privacy guarantees. In this framework, the privacy level is controlled by the parameter ϵ, where

smaller values indicate stronger privacy protection; in our experiments, we adopt a strict setting of $\epsilon = 1.0$.

The remainder of this paper is organized as follows. Section II presents the background concepts, including ECG signal processing, Transformer architectures, and Differential Privacy. Section III reviews related work on privacy-preserving ECG analysis. Section IV introduces the proposed DP-ECG-HC architecture. Section V describes the experimental setup and analyzes the results. Finally, Section VI concludes the paper.

2. Background

This section reviews the key background concepts relevant to this work, including ECG signal processing, Transformer architectures, and Differential Privacy applied to deep learning models.

2.1. Electrocardiogram (ECG) Signal Processing

Electrocardiogram (ECG) signal analysis relies on the identification of characteristic cardiac waveforms, including the P wave, which reflects atrial depolarization; the QRS complex, which represents ventricular depolarization; and the T wave, corresponding to ventricular repolarization. Although ECG signals are sometimes represented as images for analysis, they are inherently one-dimensional time series describing voltage variations over time [Ghazarian et al. 2024]. Processing the signal in its native 1D format is generally more appropriate, as converting it to a 2D representation may complicate feature extraction and reduce computational efficiency.

In the context of deep learning, ECG analysis presents several challenges. One major issue is the high inter-patient variability caused by anatomical and physiological differences between individuals. Additionally, ECG datasets often exhibit significant class imbalance, as pathological events such as arrhythmias are much rarer than normal heartbeats, making the detection of abnormal conditions more difficult [Cui et al. 2021].

2.2. Transformer Architectures

The Transformer model proposed by [Vaswani et al. 2017] revolutionized sequence processing by introducing the self-attention mechanism, which enables the model to capture long-range dependencies by relating different positions within the sequence and modeling global contextual relationships.

The Conformer architecture extends this capability by integrating convolutional modules into a block that alternates feed-forward, attention, and convolution layers [Gulati et al. 2020]. This hybrid design is well-suited for complex signals such as ECG. Convolutional layers effectively extract local morphological features of heartbeats through small receptive fields, while the self-attention mechanism captures global rhythmic dependencies and long-range temporal interactions.

For classification, these architectures commonly employ a token-based representation in which a Class Token (CLS) is appended to the beginning of the input sequence [Marouani et al. 2025]. This token serves as an aggregator that learns to summarize the global information across all sequence elements using the attention mechanism. Recent studies suggest that specialized processing of these tokens, treating them through distinct computational paths, can reduce interference between local and global feature learning, resulting in richer representations for the final classification task.

2.3. Differential Privacy

Differential Privacy (DP) [Dwork et al. 2006, Dwork and Roth 2014] has become a widely adopted standard for ensuring privacy in a variety of contexts. Through a rigorous mathematical framework, DP provides formal privacy guarantees. When applied to databases, DP limits the influence of any individual record on the outcome of a query, thereby preserving data utility while protecting the anonymity of individuals. Formally, an algorithm M satisfies (ϵ, δ)-differential privacy if, for any two neighboring datasets D and D' (which differ by a single record), the probability of producing a given output S is similar for both datasets, satisfying the following inequality:

$$P(M(D) \in S) \leq e^{\epsilon} P(M(D') \in S) + \delta \tag{1}$$

In this definition, the privacy budget ϵ quantifies the maximum allowable privacy loss; smaller values correspond to stronger privacy protection and typically require adding more noise. The relaxation parameter δ represents the probability that the privacy guarantee may fail and should ideally be smaller than the inverse of the dataset size.

2.4. Differentially Private SGD (DP-SGD)

Differentially Private Stochastic Gradient Descent (DP-SGD) [Abadi et al. 2016] is a widely used method for training deep learning models under differential privacy constraints. In this approach, gradients are first computed at the level of individual training examples within a randomly sampled mini-batch. Each per-example gradient is then clipped so that its ℓ_2 norm does not exceed a predefined threshold C, limiting the maximum influence of any single data point on the training process. The clipped gradients are aggregated and perturbed through Gaussian noise injection, producing a privatized gradient that is used to update the model parameters. Because training occurs over many iterations, each step contributes to the cumulative privacy loss. To track this loss and ensure that the total privacy budget (ϵ, δ) is not exceeded, privacy accounting techniques such as the Moments Accountant or Rényi Differential Privacy (RDP) are employed, providing tighter bounds on privacy loss across multiple training steps.

Because neural network training involves multiple iterations, the privacy loss accumulates over time. Privacy accounting mechanisms are therefore used to track the total privacy cost during training. Techniques such as the Moments Accountant and Rényi Differential Privacy (RDP) provide tighter estimates of the cumulative privacy loss as a function of the sampling probability q, the number of iterations T, and the privacy parameters (ϵ, δ), ensuring that the training process remains within a predefined privacy budget.

3. Related Work

The growing demand for privacy-preserving machine learning in ECG analysis aims to protect cardiac signals as unique biometric identifiers (heartprints). Techniques such as differential privacy and cryptographic methods play a key role in preventing re-identification and data reconstruction attacks, ensuring that diagnostic utility does not compromise the confidentiality and integrity of patient information.

PrivECG was proposed by [Nolin-Lapalme et al. 2023] and focuses on generating synthetic data for end-to-end anonymization using deep generative models. Although it

achieves good results in creating datasets that preserve essential characteristics, the work does not rely on the mathematical guarantees of Differential Privacy; instead, it limits its protection to that derived from synthetic data. Following a similar line of research based on synthetic data, [Ma et al. 2023] proposed RDP-GAN to formalize the generation of ECG data under Differential Privacy guarantees. The approach employs Rényi Differential Privacy to train GANs that generate synthetic heartbeats with formal privacy guarantees.

[Ghazarian et al. 2024] conducted an extensive review and a case study focused on arrhythmia classification. However, the primary focus is on releasing queries over extracted structured data, such as the mean/median of the QRS complex, heart rate, sex, and age, rather than processing the data in its raw time-series form. The study shows that Differential Privacy protects individual identity but makes the classification of rare conditions more challenging due to the noise introduced into the data.

More recently, [ElKomy et al. 2025] investigated the application of DP-SGD across several neural network architectures (CNN, LSTM, GRU, and RNN) for heartbeat anomaly detection. Using multimodal data, their results indicate that GRU achieves the best performance under privacy constraints. However, the study focuses on binary anomaly detection rather than multiclass classification of specific pathologies and operates under relatively mild privacy constraints without a strict epsilon budget. [Agrawal et al. 2024] combines federated learning with DP-SGD applied to a ResNet architecture. Their results show that, even with large-scale datasets, achieving utility comparable to a non-private model requires a very large epsilon budget ($\epsilon \geq 1000$), highlighting the challenge of maintaining clinical utility under strong privacy guarantees.

Unlike other noise-based approaches, [Barni et al. 2011] proposed ECG classification in the encrypted domain. The work employs Homomorphic Encryption (HE) and Garbled Circuits (GC) to enable classification without ever accessing the original data. Although it provides strong security without introducing noise, this technique introduces substantially higher computational and communication overhead than Differential Privacy-based methods.

Table 1 summarizes the landscape of privacy-preserving ECG analysis and highlights the critical gaps our proposed architecture addresses. A formal Differential Privacy guarantee ensures that the privacy protection is mathematically bounded by the budget ϵ, rather than relying on empirical obfuscation heuristics. While recent studies such as [Ghazarian et al. 2024] and [ElKomy et al. 2025] successfully achieve formal ϵ-DP, their scopes remain limited. [Ghazarian et al. 2024] focus on releasing statistical queries over pre-extracted tabular features rather than training predictive models on raw temporal signals. Conversely, [ElKomy et al. 2025] apply DP-SGD directly to signals but restrict their evaluation to simpler binary anomaly detection tasks. In contrast, scaling DP to complex multiclass heartbeat classification on raw signals often forces models to operate under impractically large privacy budgets (e.g., $\epsilon \geq 1000$) to maintain utility, as seen in [Agrawal et al. 2024]. To the best of our knowledge, DP-ECG-HC is the first approach to effectively couple a Conformer-based architecture with DP-SGD to overcome this barrier. By extracting both local morphological features and global rhythmic patterns from raw 1D signals, our method successfully preserves robust predictive utility for multiclass classification while satisfying strict, mathematically sound privacy guarantees ($\epsilon \leq 5.0$).

Table 1. Comparison of state-of-the-art privacy-preserving machine learning approaches for ECG data.

Reference	Privacy Mechanism	DP Guarantee	Input Data	Task Level
[Nolin-Lapalme et al. 2023]	Synthetic Gen. Models	$\times$	Raw Signal	Generation
[Ghazarian et al. 2024]	Statistical Queries	✓(ϵ-DP)	Tabular/Features	Query Release
[ElKomy et al. 2025]	DP-SGD (CNN/RNN)	✓(ϵ-DP)	Multimodal	Binary Anomaly
[Barni et al. 2011]	Homomorphic Encrypt.	$\times$ (Crypto)	Features	Binary Class.
[Agrawal et al. 2024]	FL + DP-SGD (ResNet)	✓($\epsilon \geq 1000$)	Raw Signal	Multiclass
This Work (DP-ECG-HC)	**DP-SGD + Conformer**	✓($\epsilon \leq 5.0$)	**Raw Signal**	**Multiclass**

4. The DP-ECG-HC Architecture

This section presents the Differentially Private ECG Heartbeat Classifier (DP-ECG-HC). The architecture is designed to perform heartbeat classification from raw ECG signals while preserving patient privacy through Differential Privacy guarantees. By combining a Conformer-based feature extractor with a latent representation bottleneck and DP-SGD training, the model captures both local morphological patterns and long-range temporal dependencies while preventing the memorization of sensitive data.

4.1. ECG-HC Architecture (Base Model)

The Electrocardiogram Heartbeat Classifier (ECG-HC) is the base architecture underlying our approach. The differentially private model presented in the following subsections is obtained by training this same network under DP constraints, and its non-private version also serves as the upper-bound baseline in our experiments (Section 5). ECG-HC is a hybrid convolutional-attention (Conformer-based) network that maps a raw 1D ECG segment to a predicted heartbeat class. The input signal is first processed by a 1D convolutional layer, which slides small filters along the time axis to detect local waveform patterns such as the QRS complex, a max-pooling operation then downsamples the resulting feature maps, shortening the sequence and reducing computational cost while retaining the most salient activations. Each position of this feature sequence is projected into a fixed-size embedding and augmented with a positional encoding, which preserves the temporal order of the heartbeat information that the self-attention mechanism would otherwise disregard. A set of learnable *class tokens* is then concatenated to the embedded sequence, and the combined sequence is fed to a Transformer encoder. Through self-attention, each position is related to all others, capturing the long-range dependencies that characterize the cardiac rhythm, while each class token aggregates the information across the sequence that is most relevant to its corresponding class (this multi-token scheme is detailed in Section 4.4). Finally, the class tokens feed a lightweight feed-forward (MLP) classifier that outputs the predicted category, so that ECG-HC jointly models local morphology and global rhythm.

4.2. Overview of the Privatized Pipeline

Building on the ECG-HC architecture introduced above, DP-ECG-HC retains the same network structure but replaces standard training with a Differentially Private pipeline designed to maximize predictive utility under stringent privacy constraints. Figure 1 illustrates the end-to-end pipeline. The architecture processes raw 1D ECG signals by extracting local morphological features and global rhythmic patterns, which are then compressed

into a dense, lower-dimensional latent space $\mathbf{z}$. This representation bottleneck prevents the direct memorization of raw input data. During the backward pass, the model mitigates the impact of data imbalance via a Focal Loss mechanism, while the DP-SGD optimizer sanitizes gradients before updating the network weights, ensuring the model does not memorize patient-specific patterns.

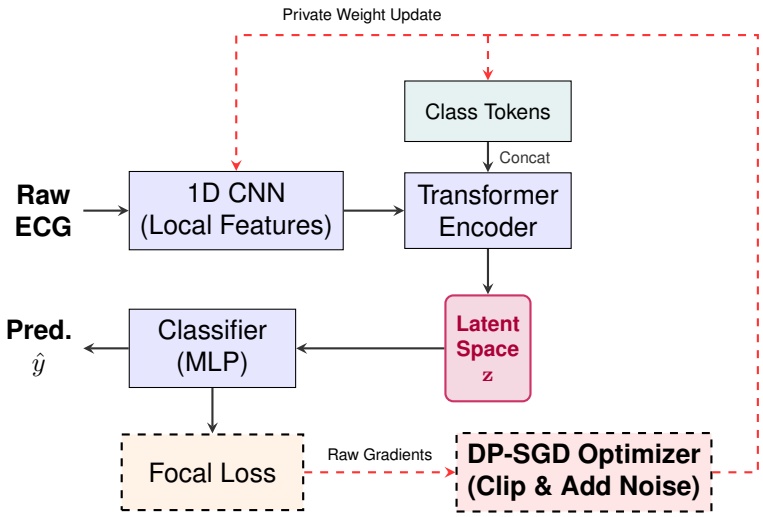


Figure 1. Architecture and training pipeline of DP-ECG-HC. The forward pass (solid lines) extracts local and global features, compressing the sequence into a dense latent space $\mathbf{z}$ bottleneck before classification. The backward pass (dashed lines) uses Focal Loss to address imbalance, while the DP-SGD optimizer injects noise to ensure differential privacy before updating the network weights.

4.3. Conformer-based Feature Extraction

Raw ECG signals are characterized by highly localized morphological features (such as the QRS complex and P/T waves) embedded within a broader temporal rhythm. To capture both aspects effectively, DP-ECG-HC employs a Conformer-based architecture.

First, the input signal passes through a 1D Convolutional Neural Network (1D-CNN) module. The convolutional layers act as local feature extractors, utilizing adjusted kernels and strides specifically tuned for 1D time-series data to downsample the signal while capturing transient morphological anomalies. Subsequently, the extracted feature maps are fed into a Transformer Encoder. The self-attention mechanism in Transformer models shapes long-range global dependencies among previously extracted local features, allowing the network to understand the overall cardiac rhythm. This local-global fusion creates a highly discriminative latent manifold even before privacy constraints are applied.

4.4. Multi-token Processing Technique

Traditional Transformer architectures for classification often rely on a single global average pooling operation or a single classification token. However, to better preserve the semantics of rare arrhythmias under DP noise, our architecture integrates a multi-token processing technique.

Specific class tokens are generated and concatenated to the sequence of features produced by the CNN before entering the Transformer Encoder. Through the successive layers of self-attention, these tokens act as information aggregators, continuously interacting with the sequence to dynamically extract and summarize the most relevant morphological patterns for each respective class. Ultimately, these tokens form the dense latent representation $\mathbf{z}$, which serves as input to the final Multi-Layer Perceptron (MLP) classifier head.

4.5. Learning under Differential Privacy Constraints

Training deep learning models with Differential Privacy typically causes a pronounced drop in utility. This degradation is highly asymmetric: the Gaussian noise injected during training disproportionately overwhelms the learning signals of minority classes (e.g., rare arrhythmias), driving the model to collapse into predicting only the majority class.

To address this challenge, we integrate the DP-SGD algorithm with a Focal Loss mechanism. Unlike standard Cross-Entropy, Focal Loss dynamically scales the error based on the prediction confidence, focusing the learning process on "hard" examples. It is formally defined as:

$$FL(p_t) = -(1 - p_t)^{\gamma} \log(p_t)$$

where p_t is the model's estimated probability for the ground-truth class, and $\gamma \geq 0$ is a tunable focusing parameter. In our privatized pipeline, γ plays a pivotal role. By heavily penalizing misclassifications of minority classes, the Focal Loss generates larger gradient signals for rare arrhythmias. This ensures that their gradient updates are strong enough to withstand the clipping and noise injection steps of DP-SGD, preserving the model's discriminative capacity for critical cardiac events.

4.6. Formal Privacy Guarantees

The formal privacy guarantees of DP-ECG-HC are based on the rigorous application of the Gaussian mechanism within the DP-SGD optimizer. To ensure that no single heartbeat dictates the model's behavior, a sensitivity analysis is implicitly enforced via gradient clipping. Before aggregation, the per-example gradients $\mathbf{g}_i$ are clipped to a maximum ℓ_2-norm threshold C:

$$\bar{\mathbf{g}}_i = \frac{\mathbf{g}_i}{\max\left(1, \frac{\|\mathbf{g}_i\|_2}{C}\right)}$$

Following the clipping step, Gaussian noise proportional to the sensitivity threshold C and a noise multiplier σ is added to the aggregated gradients: $\mathcal{N}(0, \sigma^2 C^2 \mathbf{I})$. To strictly track the privacy budget consumed across multiple training epochs, our implementation uses Rényi Differential Privacy (RDP) accounting. RDP provides a tight upper bound on the privacy loss, ensuring that the cumulative ϵ budget is analytically guaranteed at the end of training, thereby establishing a mathematical foundation for the model's privacy compliance.

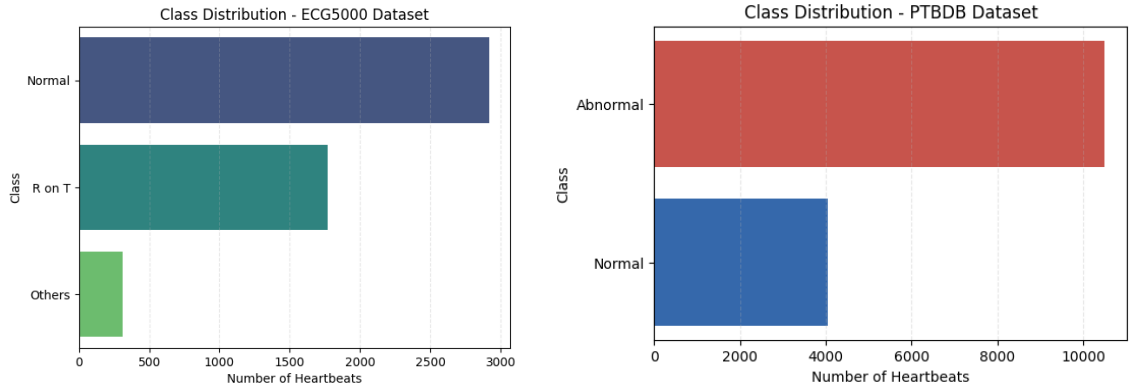
5. Experiments

We evaluate DP-ECG-HC on two ECG heartbeat classification datasets, ECG5000 and PTBDB, analyzing the privacy-utility trade-off and the model's robustness to the noise

introduced by DP-SGD. It is compared against CNN, Encoder-Transformer, and Random Forest baselines using Accuracy and Macro-averaged F1, Recall, and Precision across different privacy budgets.

5.1. Datasets

Experiments were conducted on two ECG datasets commonly used for heartbeat classification: ECG5000 and PTBDB. ECG5000 contains 5,000 segmented heartbeat signals derived from the MIT-BIH Arrhythmia Database and originally includes five classes representing normal and arrhythmic beats. Due to the strong class imbalance in the dataset, the three minority classes were grouped into a single category, yielding a three-class classification problem with one majority normal class and two abnormal classes. PTBDB is derived from the PTB Diagnostic ECG Database and contains 14,552 heartbeat segments for binary classification between normal and pathological heartbeats associated with myocardial infarction. The class distributions of both datasets are illustrated in Figure. 2.



(a) ECG5000 dataset. **(b) PTBDB dataset.**

Figure 2. Class distribution of the datasets used in the experiments.

5.2. DP Hyperparameters

The training configurations used in this work and in related studies are summarized in Table 2. The table highlights key hyperparameters, including dataset splits, optimization settings, and Differential Privacy parameters, such as gradient clipping and the noise multiplier, enabling a direct comparison between the proposed approach and previously reported methods.

Table 2. Training hyperparameters used in this work.

Parameter	Value
Epochs	200
Training Split	70%
Validation Split	15%
Test Split	15%
Optimizer	DP-SGD
Batch Size	512
Learning Rate	10^{-3}
Max Grad Norm (C)	1.0
Noise Multiplier (σ)	1.0

5.3. Analysis of Results

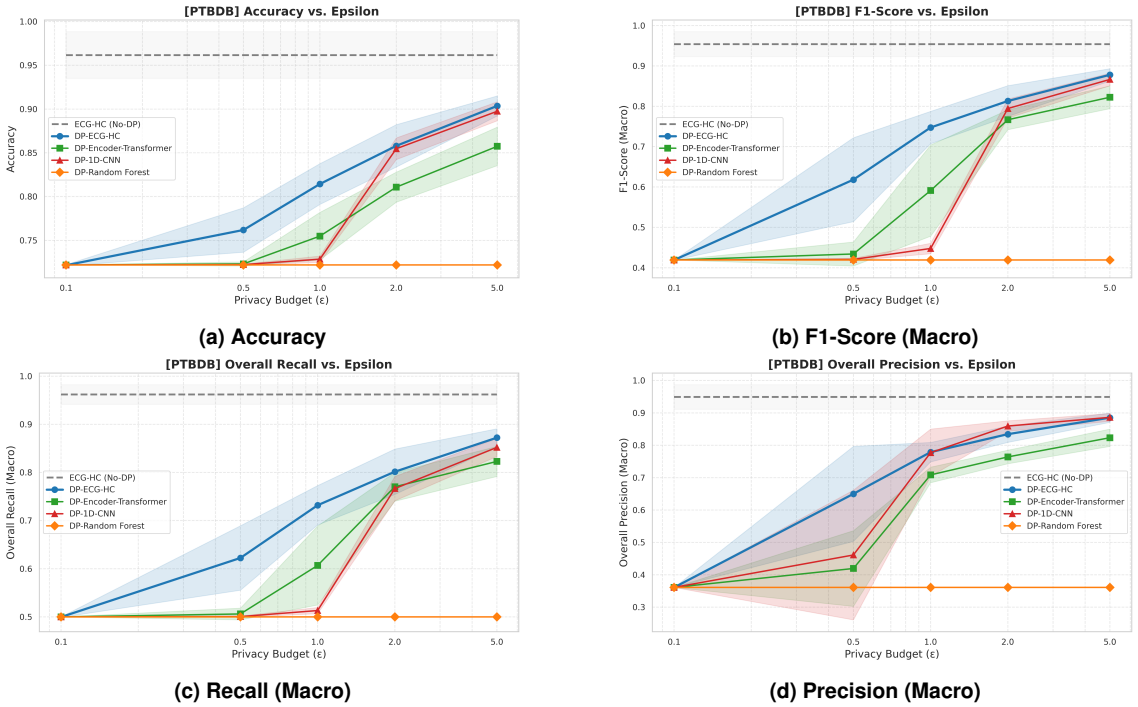
To assess the privacy-utility trade-off across privacy levels (ϵ), we report standard metrics derived from the confusion matrix (TP, TN, FP, FN). Accuracy is $(TP + TN)/(TP + TN + FP + FN)$. Since ECG datasets are imbalanced, we also report macro-averaged metrics, which weight every class equally and thus prevent dominant classes from masking poor performance on rare pathologies:

$$\text{Macro-}P = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FP_i}, \qquad \text{Macro-}R = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FN_i}, \tag{2}$$

$$\text{Macro-}F_1 = \frac{1}{N}\sum_{i=1}^{N}\frac{2\,P_i R_i}{P_i + R_i}. \tag{3}$$

Lower ϵ yields stronger privacy but more training noise, which may reduce performance. Across both datasets, DP-ECG-HC consistently achieved the best trade-off, showing greater robustness in low-privacy regimes ($\epsilon \leq 1.0$), where competing architectures degrade more.

On the PTBDB dataset (Figure 3), which poses a binary classification task between normal and myocardial-infarction heartbeats, DP-ECG-HC reaches 0.90 accuracy and 0.88 macro F1 at $\epsilon = 5$, outperforming both DP-Encoder-Transformer and DP-1D-CNN across all evaluated privacy budgets. The advantage of the proposed model is most pronounced in the moderate-privacy regime ($\epsilon \in [0.5, 2]$), while the convolutional and pure-Transformer baselines degrade sharply as the budget tightens. In contrast, DP-Random Forest remains largely insensitive to ϵ but fails to achieve competitive performance.

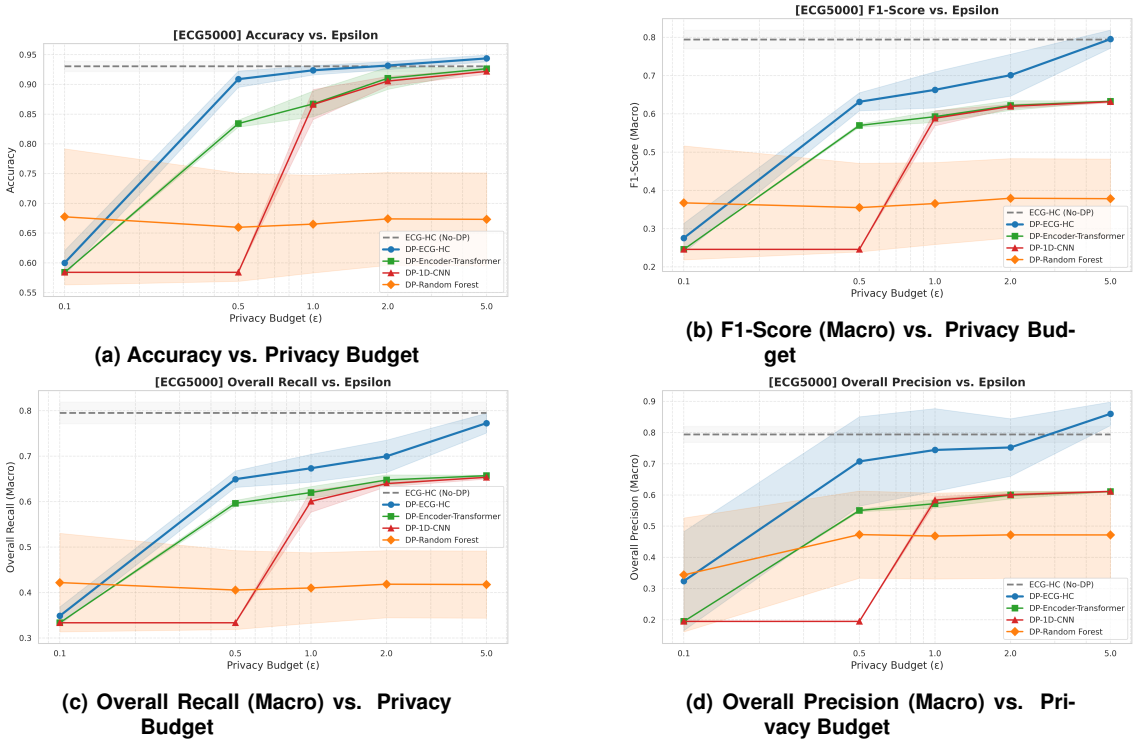


(a) Accuracy **(b) F1-Score (Macro)**

(c) Recall (Macro) **(d) Precision (Macro)**

Figure 3. Privacy-utility trade-off analysis on the PTBDB dataset.

Similar behavior is observed on ECG5000, refers to Figure 4, where the proposed model maintains superior recall and macro-F1, indicating improved detection of minority classes even under strong privacy constraints. These results suggest that the hybrid architecture of DP-ECG-HC better mitigates the instability introduced by noisy gradient updates in DP-SGD, enabling more stable representation learning compared to baseline architectures.



(a) Accuracy vs. Privacy Budget

(b) F1-Score (Macro) vs. Privacy Budget

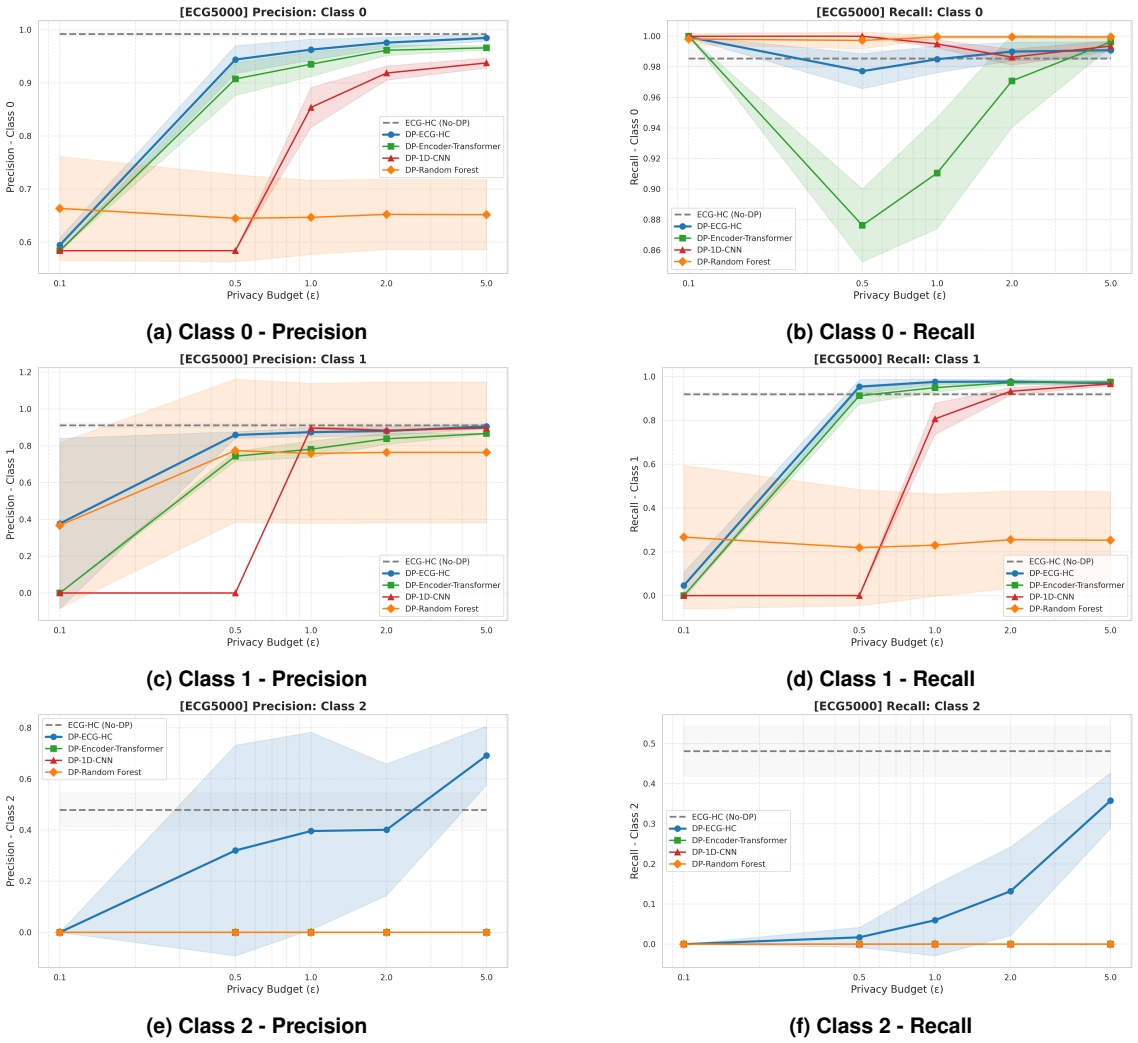
(c) Overall Recall (Macro) vs. Privacy Budget

(d) Overall Precision (Macro) vs. Privacy Budget

Figure 4. Privacy-utility trade-off on the ECG5000 dataset.

Overall, the results demonstrate that DP-ECG-HC preserves discriminative capacity under privacy noise, enabling reliable ECG classification while maintaining formal privacy guarantees.

To complement the aggregate results, we also provide a finer-grained, per-class analysis on the ECG5000 dataset, examining how Differential Privacy affects each heartbeat category individually rather than the dataset as a whole. The recall and precision metrics for Classes 0, 1, and 2 (Figure 5) show that the impact of Differential Privacy varies according to the difficulty and representativeness of each class. Class 0 exhibited high robustness, maintaining recall and precision close to those of the non-private model for $\epsilon \geq 0.5$, indicating that its patterns are easily preserved even under privacy constraints. Class 1 suffered significant degradation only at $\epsilon = 0.1$, but quickly recovered for higher values of ϵ, reaching results close to the baseline. Class 2, by contrast, was the most affected: it exhibited low recall across all privacy levels, while its precision increased gradually, suggesting that the model becomes more conservative when predicting this category. Overall, DP-ECG-HC proved more robust than the other evaluated architectures, particularly on the harder classes, and was the only model able to preserve meaningful discriminative power for Class 2 when trained under Differential Privacy.



(a) Class 0 - Precision (b) Class 0 - Recall

(c) Class 1 - Precision (d) Class 1 - Recall

(e) Class 2 - Precision (f) Class 2 - Recall

Figure 5. Performance metrics for heartbeat classes under privacy budgets.

6. Conclusion

This work presented DP-ECG-HC, a hybrid convolutional-Transformer architecture for privacy-preserving ECG classification under Differential Privacy, jointly capturing local morphological patterns and long-range temporal dependencies. Experiments on the ECG5000 and PTBDB datasets show that, when trained with DP-SGD, it achieves a favorable privacy–utility trade-off, remaining more robust to gradient noise while maintaining competitive performance, highlighting the feasibility of privacy-preserving learning for ECG-based diagnostic systems.

Future work includes exploring DP-protected generative models for synthetic ECG generation, enabling pretraining strategies that maintain clinical utility under strict privacy budgets ($\epsilon \approx 1.0$), as well as fairness-aware DP techniques, such as group-aware gradient clipping, to mitigate the disproportionate impact of privacy noise on rare diseases and patient subgroups.

Acknowledgements We acknowledge the support of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) grant 308635/2025-6, and of Lenovo, as part of its R&D investment under Brazil's Informatics Law.

References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., and Zhang, L. (2016). Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 308–318. ACM.

Agrawal, V., Kalmady, S. V., Malipeddi, V. M., Manthena, M. V., Sun, W., Islam, S., Hindle, A., Kaul, P., and Greiner, R. (2024). Federated learning and differential privacy techniques on multi-hospital population-scale electrocardiogram data. *arXiv preprint arXiv:2405.00725*.

Al-Jibreen, A., Al-Ahmadi, S., Islam, S., and Artoli, A. M. (2024). Person identification with arrhythmic ecg signals using deep convolution neural network. *Scientific Reports*, 14(1):4460.

Barni, M., Failla, P., Lazzeretti, R., Sadeghi, A.-R., and Schneider, T. (2011). Privacy-preserving ecg classification with branching programs and neural networks. *IEEE Transactions on Information Forensics and Security*, 6(2):452–468.

Cui, J., Wang, L., He, X., De Albuquerque, V. H. C., AlQahtani, S. A., and Hassan, M. M. (2021). Deep learning-based multidimensional feature fusion for classification of ecg arrhythmia. *Neural Comput. Appl.*, 35(22):16073–16087.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In Halevi, S. and Rabin, T., editors, *Theory of Cryptography*, pages 265–284. Springer Berlin Heidelberg.

Dwork, C. and Roth, A. (2014). The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3–4):211–407.

ElKomy, O. M., Rushdy, E., Nasser, S., and Khashaba, M. M. (2025). Exploring differential privacy in cnns, lstms, grus, and rnns for heartbeat detection from multimodal data. *Journal of Big Data*, 12(1):216.

Ghazarian, A., Zheng, J., and Rakovski, C. (2024). Privacy-preserving ecg data analysis with differential privacy: A literature review and a case study. *arXiv preprint arXiv:2406.13880*.

Gulati, A., Qin, J., Chiu, C.-C., Parmar, N., Zhang, Y., Yu, J., Han, W., Wang, S., Zhang, Z., Wu, Y., and Pang, R. (2020). Conformer: Convolution-augmented transformer for speech recognition. In *Proc. Interspeech 2020*.

Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., and Qadir, J. (2023). Privacy-preserving artificial intelligence in healthcare: Techniques and applications. *Computers in Biology and Medicine*, 158:106848.

Ma, C., Li, J., Ding, M., Liu, B., Wei, K., Weng, J., and Poor, H. V. (2023). Rdp-gan: A rényi-differential privacy based generative adversarial network. *IEEE Transactions on Dependable and Secure Computing*, 20(6):4838–4852.

Marouani, A., Siméoni, O., Jégou, H., Bojanowski, P., and Vo, H. V. (2025). Revisiting [cls] and patch token interaction in vision transformers. *arXiv preprint*.

Moura, L., Coutinho, E., Moreira, L., and Costa, I. (2024). Desafios do tratamento de dados da lgpd em aplicações web. In *Anais do Workshop em Engenharia de Requisitos (WER24)*. UFC/ITA.

Nolin-Lapalme, A., Avram, R., and Julie, H. (2023). Privecg: generating private ecg for end-to-end anonymization. In Deshpande, K., Fiterau, M., Joshi, S., Lipton, Z., Ranganath, R., Urteaga, I., and Yeung, S., editors, *Proceedings of the 8th Machine Learning for Healthcare Conference*, volume 219 of *Proceedings of Machine Learning Research*, pages 509–528. PMLR.

Phang, K. and Kaabi, J. (2025). Privacy in flux: A 35-year systematic review of legal evolution, effectiveness, and global challenges (u.s./e.u. focus with international comparisons). *Journal of Cybersecurity and Privacy*, 5(4):103.

Pinto, J. R. and Cardoso, J. S. (2020). Explaining ecg biometrics: Is it all in the qrs? In *2020 International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE.

Rafie, N., Kashou, A. H., and Noseworthy, P. A. (2021). Ecg interpretation: Clinical relevance, challenges, and advances. *Hearts*, 2(4):505–513.

Vásquez-Iturralde, F., Flores-Calero, M. J., Grijalva, F., and Rosales-Acosta, A. (2024). Automatic classification of cardiac arrhythmias using deep learning techniques: A systematic review. *IEEE Access*, 12.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA.

Zanchi, B., Monachino, G., Fiorillo, L., Conte, G., Auricchio, A., Tzovara, A., and Faraci, F. D. (2025). Synthetic ecg signals generation: A scoping review. *Computers in Biology and Medicine*, 184:109453.

Zhang, X. and Zhang, Q. (2025). Defending against attacks in deep learning with differential privacy: a survey. *Artificial Intelligence Review*, 58(347).