Inversion from ECG Foundation Model

Junseok Lee[0009−0003−6922−5929], Seungwoo Jung[0009−0002−8668−2892], Gyeongsik Yang[0000−0003−4560−2972], and Chuck Yoo[0000−0002−1115−1862]

Korea University, Seoul 02841, Republic of Korea
{jslee,swjung}@os.korea.ac.kr,
g_yang@korea.ac.kr, chuckyoo@os.korea.ac.kr

Abstract. Electrocardiogram (ECG) foundation models (FMs) are increasingly deployed as cloud services. For downstream tasks such as disease diagnosis, they return compact embeddings from input ECG signals. In this paper, we find that such embeddings pose significant privacy risks: adversaries can use the exposed embeddings and invert the original ECG, which leads to the inference of sensitive clinical attributes. To assess this vulnerability, we introduce SHINE, a new ECG inversion method that inverts ECG signals from the exposed embeddings. SHINE consists of ECG mapper and ECG inverter. Specifically, SHINE trains ECG mapper to approximate the embedding space of victim ECG FM, which is then used to synthesize the necessary data to train ECG inverter from scratch. Our evaluation against four state-of-the-art ECG FMs shows that SHINE achieves high-fidelity inversion. Importantly, the inverted ECGs retain sensitive clinical attributes, including cardiovascular disease indicators up to 99.6%, almost identical to the original ECG. These results demonstrate that the exposed ECG embeddings preserve substantial physiological information so that when deploying ECG FM, stronger security countermeasures need to be devised. To our knowledge, this paper is the first to report that the exposed embeddings can be inverted to the original ECG signals.

Keywords: ECG foundation models · ECG inversion · privacy · query-efficient attack

1 Introduction

Electrocardiogram (ECG) is a recording of the heart's electrical activity, typically represented as a standard 12-lead signal capturing complementary views of cardiac activity, and is widely used for screening and diagnosing cardiac conditions. Because expert-level ECG interpretation requires nearly a decade of medical and specialty training [3], recent studies have proposed ECG foundation models (FMs) that pre-train large neural networks on massive ECG corpora (e.g., MIMIC-IV-ECG [7] and Harvard–Emory ECG Database [10]). Due to their high computational cost, ECG FMs are commonly deployed as cloud services [16, 9]. In this setup, a client uploads an ECG signal to the cloud, and the ECG FM in the cloud maps the signal to an embedding—a compact numerical vector summarizing clinically relevant patterns in the ECG signal [16]. The embedding is

returned to the client and used in downstream applications such as diagnosis, cohort retrieval, and longitudinal monitoring [16, 24, 11].

However, cloud-based ECG FMs can introduce a critical privacy vulnerability. Outputs of cloud services can be exposed to adversaries (e.g., via vector database leaks [5] or network eavesdropping [6]). Once exposed, the returned embedding becomes vulnerable to embedding inversion, where an adversary inverts the embedding to the original ECG signal. This risk is especially severe for ECG. Unlike resettable credentials such as passwords, ECG signals are immutable biometric traits encoding sensitive attributes of individuals, including gender, age, and clinical diagnoses [22, 30]. Thus, inverting ECG signals from leaked embeddings causes irreversible privacy loss.

Despite these risks, embedding inversion against ECG FMs remains largely unexplored. In computer vision (CV) and natural language processing (NLP), embedding inversion has been demonstrated, where adversaries invert the original image or text from exposed embeddings [32]. Such attacks are feasible because widely available pretrained decoders can be used to map embeddings back to the input space [15, 26]. These decoders are trained on general-purpose datasets and should be fine-tuned to a specific victim model's domain. To achieve this, adversaries collect query-response pairs (direct QR pairs) by submitting inputs to the victim model and storing the returned embeddings. The direct QR pairs are used to align the pretrained decoder with the victim-specific domain.

However, directly applying CV and NLP approaches to ECG FMs is infeasible due to two fundamental barriers. First, unlike CV and NLP [32], there are no widely established pretrained decoders for ECG that can serve as a starting point; thus, adversaries must build an inversion model from scratch, which requires substantially more direct QR pairs to achieve reliable alignment.

Second, acquiring large direct QR pairs is infeasible in realistic cloud settings [18]. For example, platforms such as Hugging Face limit access to 500 requests per 5 minutes [1]. Many services also flag anomalous behavior after approximately 1.6K queries and may block further requests to cap the attainable direct QR pair budget [17]. In contrast, existing CV/NLP inversion methods require $\sim$220K direct QR pairs [15], which is unrealistic to get in real-world deployments.

In this work, we demonstrate that, even with the two barriers, high-fidelity ECG inversion is achievable. Here, we introduce SHINE, a query-efficient method for inverting ECG signals from leaked embeddings. Rather than relying on massive real-world queries, SHINE learns ECG mapper from limited direct QR pairs to approximate the victim embedding space, and uses it to synthesize pseudo QR pairs. We then train ECG inverter from scratch on these pseudo QR pairs without requiring any pretrained decoders. Once trained, the ECG inverter takes exposed embeddings from the victim ECG FM as input and inverts the original ECG signal. By revealing the vulnerability, we aim to provoke a critical discussion on the necessity of securing biometric data in ECG FMs.

Contributions. We make the following contributions:

- **First Evidence of ECG Inversion.** We show that embeddings returned by ECG FMs can be inverted to clinically meaningful ECG signals.

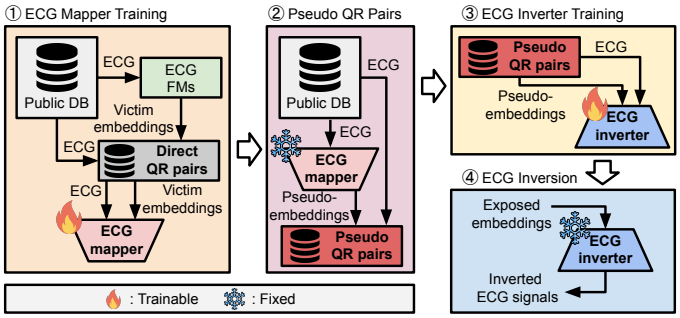


Fig. 1: SHINE method.

– **Query-efficient Inversion Method.** We introduce SHINE that enables inversion with small direct QR pairs without any pretrained decoder.
– **Comprehensive Evaluation.** SHINE achieves high-fidelity inversion on ECG waveform and disease diagnosis (∼99.6%), while remaining robust under realistic query limits (500–1000 direct QR pairs).

2 Method

In this section, we first define the threat model for ECG inversion (§2.1) and then introduce SHINE method (§2.2).

2.1 Attack Model

Adversary's Knowledge and Capabilities: We denote a 12-lead ECG signal as x_i and its corresponding embedding from the ECG FM as z_i. The adversary has query access to the target ECG FM (victim model) and obtains a limited number of direct QR pairs $\{(x_i, z_i)\}_{i=1}^{N}$. Specifically, the adversary samples ECG signals x_i from a publicly available collection of ECG signals $\mathcal{D}_{\text{pub}}$ (e.g., PTB-XL [29]), queries the victim model, and stores the returned embedding to form each pair. Following practical cloud-service constraints [17], we assume a strict direct QR pair budget of at most $N \leq 1000$. The adversary then uses these direct QR pairs to train an inversion model that inverts an ECG signal from an embedding.
Adversary's Objective: The adversary obtains an exposed embedding from the ECG FM via leakage in a cloud-based service [5, 6]. Given a target embedding z_{tgt}, the adversary inverts the corresponding ECG signal $\hat{x}_{\text{tgt}}$ using the inversion model trained above, without access to the original ECG signal x_{tgt}. The inversion is considered meaningful if it preserves clinically sensitive attributes of the individual (e.g., gender, age, and clinical diagnoses).

2.2 SHINE method

Fig. 1 shows the overall workflow of SHINE that consists of four steps. First, SHINE trains ECG mapper M_{ecg} to approximate the victim model's embedding

generation under a limited number of direct QR pairs (Step 1 in Fig. 1). Second, SHINE synthesizes pseudo QR pairs by encoding a large number of ECG signals with M_{ecg} and treating the resulting embeddings as responses (Step 2). Third, SHINE trains ECG inverter I_{ecg} on the pseudo QR pairs to invert ECG signals from embeddings (Step 3). Finally, given an exposed target embedding z_{tgt}, SHINE uses trained I_{ecg} to produce the inverted signal $\hat{x}_{\text{tgt}}$ (Step 4). We explain the details of each step below.

Step 1—ECG Mapper Training: We train ECG mapper M_{ecg} on direct QR pairs $\{(x_i, z_i)\}_{i=1}^{N}$ (explained in §2.1) by minimizing the mean squared error (MSE): $\min_{M_{\text{ecg}}} \sum_{i=1}^{N} \text{MSE}(M_{\text{ecg}}(x_i), z_i)$.

We design M_{ecg}'s architecture to approximate the behavior of ECG FMs, and thus design M_{ecg} to be similar to the architecture of ECG FMs. Existing ECG FMs are either convolutional neural networks (CNNs) or transformer-based models [16, 19, 27, 21]. Based on this, we prepare two architectures—CNN-based M_{ecg} and transformer-based M_{ecg}—and select the corresponding one for each victim model. The detailed architectures are as follows. We follow standard designs commonly used in prior work: ResNet-style 1D CNN [8] for the CNN-based M_{ecg} and transformer with multi-head self-attention [28] for the transformer-based M_{ecg}.

Because M_{ecg} is trained from limited direct QR pairs, directly adopting large standard architectures may cause overfitting. Thus, we customize the standard designs to improve stability under small direct QR pairs [2]. For the CNN-based M_{ecg}, we use a ResNet-style 1D CNN with 5 residual blocks and replace BatchNorm with GroupNorm to stabilize small-batch training [31]. For the transformer-based M_{ecg}, we use a 1-block transformer encoder with eight attention heads. We analyze the impact of the architecture choices in the ablation study (§3.2).

Note that between the two architectures, we select one according to the victim model. Although the detailed architecture of the victim FM is not publicly disclosed, a prior study showed that an adversary can infer architectural characteristics (CNN-based vs. transformer-based) from a small set of input–output pairs [23]. In our setting, we take the limited direct QR pairs to identify the likely architecture of the victim model and instantiate M_{ecg} accordingly.

Step 2—Pseudo QR Pairs Generation: Let $\mathcal{D}_{\text{pub}} = \{x_j^{\text{pub}}\}_{j=1}^{M}$ denote a collection of ECG signals from real individuals, sourced from a publicly available dataset (e.g., PTB-XL [29]), where M is the number of signals. We apply the trained M_{ecg} to each x_j^{pub} to obtain an approximated embedding $z_j^{\text{pub}} = M_{\text{ecg}}(x_j^{\text{pub}})$, and form pseudo QR pairs $\{(z_j^{\text{pub}}, x_j^{\text{pub}})\}_{j=1}^{M}$.

Step 3—ECG Inverter Training: We train I_{ecg} on both direct and pseudo QR pairs by minimizing the MSE: $\min_{I_{\text{ecg}}} \sum_{(z,x)\in\mathcal{D}_{\text{dir}}\cup\mathcal{D}_{\text{pse}}} \text{MSE}(I_{\text{ecg}}(z), x)$, where $\mathcal{D}_{\text{dir}} = \{(z_i, x_i)\}_{i=1}^{N}$ is direct QR pairs and $\mathcal{D}_{\text{pse}} = \{(z_j^{\text{pub}}, x_j^{\text{pub}})\}_{j=1}^{M}$ is pseudo QR pairs.

We design I_{ecg} as an inverse of M_{ecg}, to invert embeddings back to the ECG signal. We consider two architectures: CNN-based I_{ecg} and transformer-based I_{ecg}. For the CNN-based I_{ecg}, we use a ResNet-style 1D CNN with 5 residual blocks, replacing Conv1D with ConvTranspose1D to invert ECG signals

from embeddings, and use GroupNorm for small-batch stability [31]. For the transformer-based I_{ecg}, we use 1-block transformer decoder with eight attention heads and a lightweight projection head to match the output to the ECG signal dimensionality. We select the I_{ecg} architecture for each victim model using the same criteria as for M_{ecg} (details in Step 1).
Step 4—ECG Inversion: Given an exposed embedding z_{tgt}, I_{ecg} inverts the corresponding ECG signal ($\hat{x}_{\text{tgt}}$) as $\hat{x}_{\text{tgt}} = I_{\text{ecg}}(z_{\text{tgt}})$.

3 Experiments and Results

3.1 Setting

Victim Models: We consider four state-of-the-art ECG FMs as victim models: ECGFounder [16], MERL [19], ECGFM-KED [27], and ST-MEM [21]. Among them, ST-MEM is Transformer-based, while the others are CNN-based.
Baseline: To our knowledge, there is no established method for ECG inversion. Thus, to evaluate the inversion accuracy of SHINE, we implement the closest available alternative, following inversion methods in CV/NLP domains that use a pretrained decoder to invert ECG signals from embeddings. As no such decoder exists for ECG FMs, we use DiffuSETS [12], an open-source ECG generator that takes clinical text as its input. Because DiffuSETS does not take embeddings as input, we follow prior work that adapts existing models to take different inputs [26]. Specifically, we attach a new adapter to DiffuSETS and train the adapter so that the model takes embeddings as its input; we denote this approach as "baseline." We note that the baseline is not directly practical, as it requires full modification of DiffuSETS, and DiffuSETS is not designed for taking FM embeddings; however, it serves as a best-effort reproduction of inversion in CV and NLP domains for comparison with SHINE.
Metrics: We evaluate the following items: 1) inversion fidelity, 2) clinical attribute leakage, 3) inversion under practical direct QR pairs, and 4) ablation study.

- **Inversion Fidelity:** We evaluate how accurately the inversion matches the original ECG signal. We report ECG waveform similarity by 1) mean absolute error (MAE; average difference in signal values), 2) Pearson correlation coefficient (PCC; how similarly the two signals rise and fall over time), and 3) dynamic time warping (DTW; similarity of overall waveform shape). All metrics are computed per lead for each ECG signal, and then averaged over the 12 leads and over all ECG samples. We also report fidelity in terms of clinically meaningful intervals, ΔRR and ΔQT, defined as the absolute differences between RR intervals and QT durations extracted from the original and inverted ECGs. These metrics assess whether physiologically meaningful cardiac intervals are preserved. For example, RR reflects rhythm irregularities such as arrhythmias, and QT reflects repolarization abnormalities such as long QT. RR intervals and QT durations are computed following [14]. We

measure all metrics with 500 and 1000 direct QR pairs (both below the practical limit) and report both results. Since the results are consistent across query budgets within the practical limit, we use these two as representatives.
- **Clinical Attribute Leakage:** We evaluate whether inverted ECGs leak sensitive attributes through downstream tasks—1) age and gender classification and 2) cardiovascular disease classification [8, 22]. For each task, we train a classifier [8] on the original ECGs and evaluate the same classifier on both the original and inverted ECGs, reporting AUROC for each input; AUROC measures how accurately the attributes are predicted. We report the normalized AUROC (r; %), defined as the AUROC of the inverted ECG divided by that of the original ECG [4]. An r closer to 100% indicates that the inverted ECGs retain nearly the same clinical attributes as the originals.
- **Inversion under Practical Direct QR Pairs:** We evaluate the fidelity of ECG inversion under diverse direct QR pair budgets within the practical limit. We vary the number of direct QR pairs from 250 to 8000 and analyze whether the fidelity remains stable and high across budgets.
- **Ablation Study:** Lastly, we analyze how our design choices affect inversion performance. Specifically, we evaluate inversion fidelity across different depths of M_{ecg} and I_{ecg} to validate our design choice. We also test the effectiveness of M_{ecg} by comparing the fidelity between full SHINE method and SHINE without M_{ecg}.

Training Details: Unless stated otherwise, SHINE and baseline are trained on PTB-XL (21837 ECGs) and evaluated on CPSC2018 (6877 ECGs) [29, 20], following the respective data-use licenses and citation requirements. To ensure attacks on data unseen by the victim models, we use PTB-XL and CPSC2018, neither of which was included in their pretraining data. For each experiment, we follow the direct QR pair budgets specified in the corresponding metric as explained above. We randomly sample ECG signals from PTB-XL and query the victim model to form direct QR pairs. We feed the remaining signals in the dataset to M_{ecg} to obtain embeddings, which form pseudo QR pairs. I_{ecg} is trained on pseudo QR pairs and direct QR pairs. Also, we use direct QR pairs to train the baseline adapter. For clinical attribute leakage, we train the downstream classifiers on PTB-XL and evaluate them on CPSC2018 for both original and inverted ECGs.

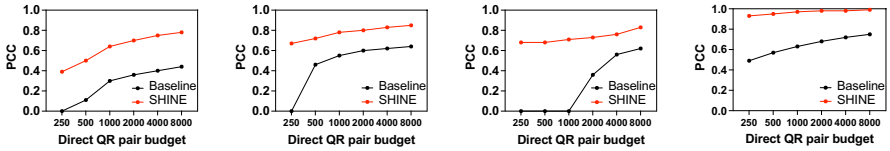
We optimize all models with AdamW (learning rate 1×10^{-3}, batch size 32, weight decay 1×10^{-5}) and apply early stopping based on validation loss (minimum delta 1×10^{-3}). We train M_{ecg} and I_{ecg} for up to 300 epochs (patience 20). We empirically choose the maximum epochs that are sufficient for convergence, with validation loss plateauing before the maximum epoch.

3.2 Results

Inversion Fidelity: Table 1 shows inversion fidelity under the specified practical budgets. For ECG waveform similarity, SHINE consistently achieves the highest fidelity between the inverted and the original ECG signals. Specifically,

Table 1: Inversion fidelity results (mean ± std, bold: best).

Victim Model	Query budget	Method	ECG waveform similarity			Fidelity on clinically meaningful intervals	
			MAE↓	DTW↓	PCC↑	ΔRR↓	ΔQT↓
ECG-Founder	500	Baseline	0.49 ± 0.01	0.76 ± 0.00	0.11 ± 0.00	0.22 ± 0.00	0.11 ± 0.00
		SHINE	$\mathbf{0.47 \pm 0.01}$	$\mathbf{0.31 \pm 0.01}$	$\mathbf{0.51 \pm 0.01}$	$\mathbf{0.12 \pm 0.00}$	$\mathbf{0.08 \pm 0.00}$
	1000	Baseline	0.52 ± 0.00	0.53 ± 0.01	0.30 ± 0.00	0.20 ± 0.00	0.09 ± 0.00
		SHINE	$\mathbf{0.39 \pm 0.01}$	$\mathbf{0.23 \pm 0.01}$	$\mathbf{0.64 \pm 0.01}$	$\mathbf{0.09 \pm 0.01}$	$\mathbf{0.07 \pm 0.00}$
ECGFM-KED	500	Baseline	0.44 ± 0.00	0.38 ± 0.01	0.47 ± 0.01	0.19 ± 0.01	0.08 ± 0.00
		SHINE	$\mathbf{0.34 \pm 0.00}$	$\mathbf{0.20 \pm 0.00}$	$\mathbf{0.73 \pm 0.00}$	$\mathbf{0.14 \pm 0.01}$	$\mathbf{0.07 \pm 0.00}$
	1000	Baseline	0.41 ± 0.01	0.33 ± 0.01	0.55 ± 0.01	0.17 ± 0.01	0.07 ± 0.00
		SHINE	$\mathbf{0.31 \pm 0.00}$	$\mathbf{0.16 \pm 0.00}$	$\mathbf{0.77 \pm 0.01}$	$\mathbf{0.13 \pm 0.01}$	$\mathbf{0.07 \pm 0.00}$
MERL	500	Baseline	0.07 ± 0.01	0.01 ± 0.01	0.01 ± 0.01	0.70 ± 0.11	0.11 ± 0.00
		SHINE	$\mathbf{0.05 \pm 0.00}$	$\mathbf{0.00 \pm 0.00}$	$\mathbf{0.68 \pm 0.01}$	$\mathbf{0.05 \pm 0.00}$	$\mathbf{0.07 \pm 0.00}$
	1000	Baseline	0.06 ± 0.00	0.01 ± 0.01	0.01 ± 0.00	0.61 ± 0.10	0.10 ± 0.01
		SHINE	$\mathbf{0.05 \pm 0.00}$	$\mathbf{0.00 \pm 0.00}$	$\mathbf{0.71 \pm 0.02}$	$\mathbf{0.05 \pm 0.00}$	$\mathbf{0.07 \pm 0.00}$
ST-MEM	500	Baseline	0.47 ± 0.00	0.35 ± 0.00	0.58 ± 0.01	0.07 ± 0.00	0.04 ± 0.00
		SHINE	$\mathbf{0.26 \pm 0.00}$	$\mathbf{0.09 \pm 0.00}$	$\mathbf{0.93 \pm 0.00}$	$\mathbf{0.03 \pm 0.00}$	$\mathbf{0.02 \pm 0.00}$
	1000	Baseline	0.45 ± 0.01	0.31 ± 0.00	0.63 ± 0.01	0.06 ± 0.00	0.04 ± 0.00
		SHINE	$\mathbf{0.22 \pm 0.00}$	$\mathbf{0.08 \pm 0.00}$	$\mathbf{0.94 \pm 0.00}$	$\mathbf{0.03 \pm 0.00}$	$\mathbf{0.02 \pm 0.00}$



(a) ECGFounder. (b) ECGFM-KED. (c) MERL. (d) ST-MEM.

Fig. 2: Inversion under practical direct QR pairs (measured by PCC).

SHINE improves MAE by 4.1–51.1% (avg. 27.2%), and DTW by 47.4–100.0% (avg. 70.4%). Similarly, SHINE achieves an absolute PCC gain of 0.22 to 0.70 (avg. 0.41). SHINE also improves the fidelity of clinically meaningful intervals, reducing ΔRR by 23.5–92.9% (avg. 55.3%) and ΔQT by 0.0–50.0% (avg. 28.5%).

Clinical Attribute Leakage: Table 2 shows clinical attribute leakage from inverted ECG signals on downstream tasks (age, gender, and disease). SHINE preserves downstream task performance to a high degree, achieving r of $\sim$99.6%, indicating that the inverted ECG preserves nearly the same level of clinically relevant attributes as the original ECG. SHINE is consistently higher than the baseline by 11.8–39.3% (avg. 21.1%) for age, 13.7–59.6% (avg. 31.3%) for gender, and 7.8–82.6% (avg. 36.2%) for disease.

Inversion under Practical Direct QR Pair Budgets: Fig. 2 compares inversion fidelity (PCC) as the number of direct QR pairs increases from 250 to

Table 2: Clinical attribute leakage (mean ± std, bold: best).

Victim Model	Query budget	Method	Age r (%) ↑	Gender r (%) ↑	Disease r (%) ↑
ECG-Founder	500	Baseline	73.1±3.3	63.7±0.8	66.5±2.5
		SHINE	**88.7±0.8**	**88.0±1.3**	**94.0±0.9**
	1000	Baseline	78.1±1.3	65.2±2.5	73.1±0.2
		SHINE	**91.5±1.2**	**92.1±1.1**	**97.6±0.1**
ECGFM-KED	500	Baseline	78.9±2.9	77.6±1.9	79.0±0.7
		SHINE	**89.0±1.1**	**88.4±2.9**	**92.4±0.5**
	1000	Baseline	81.4±1.4	79.1±4.4	83.8±1.7
		SHINE	**91.0±1.2**	**91.1±0.8**	**94.5±1.1**
MERL	500	Baseline	60.6±2.8	55.2±5.8	52.2±2.7
		SHINE	**84.4±0.4**	**88.1±1.1**	**95.3±0.7**
	1000	Baseline	62.9±2.4	59.0±2.9	53.9±1.6
		SHINE	**86.5±1.8**	**90.1±0.1**	**96.5±0.4**
ST-MEM	500	Baseline	86.3±0.8	81.3±2.1	86.0±2.7
		SHINE	**99.0±0.3**	**94.2±1.2**	**99.1±0.3**
	1000	Baseline	87.0±1.3	83.0±2.2	92.4±1.0
		SHINE	**99.2±0.3**	**94.4±1.3**	**99.6±0.1**

Table 3: Ablation study (fidelity measured by PCC, bold: best).

Depth	CNN	Transformer
1	0.45	**0.93**
2	**0.63**	0.92
3	0.61	0.91

(a) Model depth choice based on fidelity.

Victim Model	SHINE (-M_{ecg})	SHINE
ECG-Founder	0.44	**0.50**
ECGFM-KED	0.60	**0.72**
MERL	0.25	**0.68**
ST-MEM	0.90	**0.93**

(b) Effectiveness of M_{ecg}.

8000. While SHINE maintains consistently high fidelity across all victim models and budgets, its robustness is particularly evident under practical query constraints (e.g., 1600 pairs). In such settings, SHINE achieves an average absolute PCC gain of 0.39 over the baseline, exhibiting substantially smaller performance degradation. Specifically, SHINE approximates victim ECG FM's embedding well enough by using only 500 direct QR pairs. The results indicate that SHINE is robust even with limited queries, enabling high-fidelity ECG inversion under realistic attack settings.

Ablation Study: We evaluate the design choices of SHINE by measuring inversion fidelity (PCC) under 500 direct QR pairs. Table 3a reports the fidelity for different numbers of blocks in M_{ecg} and I_{ecg}.[1] The depth values 1, 2, and 3 correspond to models using three/five/seven CNN blocks and one/two/three transformer blocks, each. The results show that depth-2 for the CNN and depth-1 for the transformer achieve the highest fidelity; thus, we use these depths.

Second, Table 3b reports the effectiveness of M_{ecg} by comparing 1) SHINE (-M_{ecg})—SHINE without M_{ecg}—and 2) the full SHINE model. The full SHINE achieves an absolute PCC gain of 0.03–0.43 (avg. 0.16) over SHINE (-M_{ecg}), showing that M_{ecg} plays a critical role in high-fidelity inversion.

[1] We use similar layer depths for M_{ecg} and I_{ecg}, a widely used design choice in reconstruction-oriented architectures [25, 13].

4 Conclusion

We present SHINE, a query-efficient ECG inversion method for ECG FMs. By training ECG mapper under limited direct QR pairs and subsequently training ECG inverter with pseudo QR pairs, SHINE successfully inverts ECG signals with high fidelity, preserving physiologically relevant attributes. By demonstrating that ECG FM embeddings retain sensitive clinical information, our findings highlight critical privacy risks in real-world deployments. We leave broader evaluations across diverse ECG FMs and defense mechanisms against SHINE for future work.

Acknowledgments. Corresponding authors: Gyeongsik Yang and Chuck Yoo. This research was partly supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (RS-2021-NR060143), by the NRF grant funded by the Korea government (MSIT) (RS-2024-00336564), by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by MSIT (RS-2024-00405128), by ICT Creative Consilience Program by IITP grant funded by MSIT (IITP-2026-RS-2020-II201819), and by Google Cloud Research Credits program.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Rate limits. Hugging Face Hub Documentation, https://huggingface.co/docs/hub/en/rate-limits, accessed: 2026-02-13
2. Arpit, D., Jastrzębski, S., Ballas, N., Krueger, D., Bengio, E., Kanwal, M.S., Maharaj, T., Fischer, A., Courville, A., Bengio, Y., et al.: A closer look at memorization in deep networks. In: International conference on machine learning. pp. 233–242. PMLR (2017)
3. Berkaya, S.K., Uysal, A.K., Gunal, E.S., Ergin, S., Gunal, S., Gulmezoglu, M.B.: A survey on ecg analysis. Biomedical Signal Processing and Control **43**, 216–235 (2018)
4. Cai, Y., Jin, C., Ma, J., Zhou, F., Xu, Y., Guo, Z., Wang, Y., Zhang, Z., Liang, L., Tan, Y., et al.: A deployment-friendly foundational framework for efficient computational pathology. arXiv preprint arXiv:2602.14010 (2026)
5. Chen, Y., Xu, Q., Bjerva, J.: Algen: Few-shot inversion attacks on textual embeddings using alignment and generation. arXiv preprint arXiv:2502.11308 (2025)
6. Chen, Y., Yang, Q., Shi, Z., Chen, J.: The model inversion eavesdropping attack in semantic communication systems. In: GLOBECOM 2023-2023 IEEE Global Communications Conference. pp. 5171–5177. IEEE (2023)
7. Gow, B., Pollard, T., Nathanson, L.A., Johnson, A., Moody, B., Fernandes, C., Greenbaum, N., Waks, J.W., Eslami, P., Carbonati, T., et al.: Mimic-iv-ecg: Diagnostic electrocardiogram matched subset. Type: dataset **6**, 13–14 (2023)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)

9. Hong, S., Fu, Z., Zhou, R., Yu, J., Li, Y., Wang, K., Cheng, G.: Cardiolearn: a cloud deep learning service for cardiac disease detection from electrocardiogram. In: Companion Proceedings of the Web Conference 2020. pp. 148–152 (2020)
10. Koscova, Z., Li, Q., Robichaux, C., Junior, V.M., Ghanta, M., Gupta, A., Rosand, J., Aguirre, A., Hong, S., Albert, D.E., et al.: The harvard-emory ecg database. medRxiv pp. 2024–09 (2025)
11. Lai, J., Tan, H., Wang, J., Ji, L., Guo, J., Han, B., Shi, Y., Feng, Q., Yang, W.: Practical intelligent diagnostic algorithm for wearable 12-lead ecg via self-supervised learning on large-scale dataset. Nature Communications **14**(1), 3741 (2023)
12. Lai, Y., Chen, J., Zhao, Q., Zhang, D., Wang, Y., Geng, S., Li, H., Hong, S.: Diffusets: 12-lead ecg generation conditioned on clinical text reports and patient-specific information. Patterns (2025)
13. Lee, J., Yoo, Y., Kim, J., Lim, D., Yang, G., Yoo, C.: Parameter-efficient 12-lead ecg reconstruction from a single lead. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 431–441. Springer (2025)
14. Lence, A., Granese, F., Fall, A., Hanczar, B., Salem, J.E., Zucker, J.D., Prifti, E.: Ecgrecover: a deep learning approach for electrocardiogram signal completion. In: Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1. pp. 2359–2370 (2025)
15. Li, H., Xu, M., Song, Y.: Sentence embedding leaks more information than you expect: Generative embedding inversion attack to recover the whole sentence. arXiv preprint arXiv:2305.03010 (2023)
16. Li, J., Aguirre, A.D., Junior, V.M., Jin, J., Liu, C., Zhong, L., Sun, C., Clifford, G., Brandon Westover, M., Hong, S.: An electrocardiogram foundation model built on over 10 million recordings. NEJM AI **2**(7), AIoa2401033 (2025)
17. Li, Z., Pu, Y., Zhang, X., Li, Y., Li, J., Ji, S.: Protecting object detection models from model extraction attack via feature space coverage. In: IJCAI. pp. 431–439 (2024)
18. Li, Z., Zhang, H., Wang, J., Chen, M., Hu, H., Yi, W., Xu, X., Yang, M., Ma, C.: From head to tail: Efficient black-box model inversion attack via long-tailed learning. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 29288–29298 (2025)
19. Liu, C., Wan, Z., Ouyang, C., Shah, A., Bai, W., Arcucci, R.: Zero-shot ecg classification with multimodal learning and test-time clinical knowledge enhancement. arXiv preprint arXiv:2403.06659 (2024)
20. Liu, F., Liu, C., Zhao, L., Zhang, X., Wu, X., Xu, X., Liu, Y., Ma, C., Wei, S., He, Z., et al.: An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. Journal of Medical Imaging and Health Informatics **8**(7), 1368–1373 (2018)
21. Na, Y., Park, M., Tae, Y., Joo, S.: Guiding masked representation learning to capture spatio-temporal relationship of electrocardiogram. arXiv preprint arXiv:2402.09450 (2024)
22. Nolin-Lapalme, A., Avram, R., Julie, H.: Privecg: Generating private ecg for end-to-end anonymization. In: Machine Learning for Healthcare Conference. pp. 509–528. PMLR (2023)
23. Oh, S.J., Schiele, B., Fritz, M.: Towards reverse-engineering black-box neural networks. In: Explainable AI: interpreting, explaining and visualizing deep learning, pp. 121–144. Springer (2019)

24. Qiu, J., Zhu, J., Liu, S., Han, W., Zhang, J., Duan, C., Rosenberg, M.A., Liu, E., Weber, D., Zhao, D.: Automated cardiovascular record retrieval by multimodal learning between electrocardiogram and clinical report. In: Machine Learning for Health (ML4H). pp. 480–497. PMLR (2023)
25. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
26. Shahreza, H.O., George, A., Marcel, S.: Face reconstruction from face embeddings using adapter to a face foundation model. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 5584–5593 (2025)
27. Tian, Y., Li, Z., Jin, Y., Wang, M., Wei, X., Zhao, L., Liu, Y., Liu, J., Liu, C.: Foundation model of ecg diagnosis: Diagnostics and explanations of any form and rhythm on ecg. Cell Reports Medicine **5**(12) (2024)
28. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
29. Wagner, P., Strodthoff, N., Bousseljot, R.D., Kreiseler, D., Lunze, F.I., Samek, W., Schaeffter, T.: Ptb-xl, a large publicly available electrocardiography dataset. Scientific data **7**(1), 1–15 (2020)
30. Wang, Z., Khatibi, E., Kazemi, K., Azimi, I., Mousavi, S., Malik, S., Rahmani, A.M.: Transecg: Leveraging transformers for explainable ecg re-identification risk analysis. arXiv preprint arXiv:2503.13495 (2025)
31. Wu, Y., He, K.: Group normalization. In: Proceedings of the European conference on computer vision (ECCV). pp. 3–19 (2018)
32. Zhou, Z., Zhu, J., Yu, F., Li, X., Peng, X., Liu, T., Han, B.: Model inversion attacks: A survey of approaches and countermeasures. arXiv preprint arXiv:2411.10023 (2024)