PrivECG: generating private ECG for end-to-end anonymization

Nolin-Lapalme Alexis ALEXIS.NOLIN-LAPALE@UMONTREAL.CA
Montreal Heart Institute,
Université de Montréal,
Montreal, Qc, Canada

Avram Robert ROBERT-CALIN.AVRAM@UMONTREAL.CA
Montreal Heart Institute,
Université de Montréal,
Montreal, Qc, Canada

Hussin Julie JULIE.HUSSIN@UMONTREAL.CA
Montreal Heart Institute,
Université de Montréal,
Montreal, Qc, Canada

Abstract

The electrocardiogram (ECG) remains the cornerstone of diagnosis in cardiology where, pathologies uniquely impact its appearance, permitting the identification of underlying electrical or structural abnormalities. Notably, multiple deep learning approaches have demonstrated that disease prediction could be performed with high accuracy using ECG waveforms. However, this signal-rich modality has also demonstrated the potential to be predictive of a patient's private attributes such as biological sex and age. More importantly, recent research has demonstrated that many medical data modalities could allow patient re-identification with only the modality of interest despite anonymization through current paradigms, raising important privacy concerns. In this paper, we propose a novel approach to anonymize the ECG waveforms themselves while maximizing the privacy-utility trade-off. We describe PrivECG[1], a generative adversarial network (GAN) framework capable of privatizing 12-lead ECGs while conserving their disease-descriptive features. PrivECG significantly decreases patient validation performances by targeting sex-linked features. Our approach reduces sex prediction accuracy from 0.876 to near-random 0.529, by permitting greater variability of the ECG's R-wave morphology, as well as bringing the equal error rate (EER) from 0.098 to 0.251 on individual validation tasks. Moreover, the regenerated ECGs maintain a majority of their disease-predicting potential, with an F1 score of 0.885 from the baseline's 0.931 on a multilabel disease prediction task. We further demonstrate that reintroducing sex-linked information downstream in the network allows recuperating performances with an F1 score of 0.893 proving our loss of performance is due to the privatization of the sex-linked features, as well as serves as a disambiguation tool to evaluate the impact of sex information on prediction performances. Our results suggest that our

1. https://github.com/anolinlapalme/privECG

approach could allow improved anonymization of a large ECG database in minutes without strongly impacting downstream clinically-relevant tasks in a task-independent manner.

1. Introduction

The electrocardiogram (ECG) monitors the heart's electrical activity by measuring the voltage changes across the cardiac muscle. The typical clinical ECG uses 12 electrodes placed on various body sections, including the chest and limbs. Each of these electrodes provides a unique view of the heart's electrical activity. When combined, these 12 leads offer a 3-dimensional representation of the heart's depolarization patterns. This allows a more comprehensive understanding of the heart's activity, essential for accurate diagnosis and treatment of common cardiovascular conditions such as atrial fibrillation or myocardial infarction ("heart attack"). The standard healthy ECG waveform contains a distinctive repeating pattern of P, QRS, and T waves, each representing a key stage in the cardiac cycle of a complete heartbeat. Notably, the R-wave is often the most salient as it represents the depolarization of the ventricles, the bulk of the heart's muscle mass (Becker, 2006) (Figure 1).

Anomalies in the expected ECG waveform can help in identifying underlying cardiological pathologies such as infarction or arrhythmia (Begg et al., 2016) and is often the preferred investigation approach due to its cost-effectiveness and non-invasiveness (Saunders and Lankiewicz, 2019). Thus, the ECG became ubiquitous in cardiac pathology diagnosis and monitoring in the United States (Kligfield et al., 2007). Because of this, large public datasets and challenges emerged quickly, such as the PhysioNet/CinC Challenge in 2000 (Goldberger et al., 2000) fostering the rapid evolution of ECG-based predicting algorithms (Somani et al., 2021). Current architectures generally rely on convolution neural networks (CNN) and are occasionally reported as capable of outperforming medical expert electrophysiologists on certain disease-predictive tasks (Hannun et al., 2019; Hughes et al., 2021).

However, with the rising interest in using ECG data as a biometric authentication modality (Melzi et al., 2022) and in medical data-sharing (Flanagin et al., 2022), a growing emphasis has been put on mitigating privacy concerns (Murdoch, 2021). Indeed, although difficult for the naked eye, it has been shown that private attributes such as biological sex and age could be accurately inferred from ECGs using deep learning approaches (Attia et al., 2019). Thus, unprotected data could potentially lead to a breach of confidentiality and allow external agents to identify the patient's defining characteristics from only an ECG waveform. As underlined by Kaissis et al. (2020), current anonymization strategies targeting metadata might not be sufficient to prevent deep learning models from overcoming current anonymization approaches and re-identifying a patient's identity from the modality of interest itself (X-ray images, ECG waveforms, biological laboratory results). In fact, it has been reported that private data-mining companies have focused on building re-identifying models using deep learning models (Tanner, 2017). In the literature, examples such as Packhäuser et al. (2022) have demonstrated that levels of re-identification could reach an impressive 95% on cohorts of chest X-rays despite the anonymization of the data. In another study, Schwarz et al. (2019) were able to recreate and re-associate pictures of faces with only cranial MRI data with an accuracy of 83%. Additionally, a recent paper by Ghazarian et al. (2022) explored the ability to predict patient ID from their database using ECG waveforms

and a CNN with output neurons corresponding to each patient. Their results demonstrated that re-identification achieved an accuracy of 99.7%. Finally, studies, such as Sweeney et al. (2017), have stated that even the Health Insurance Portability and Accountability Act (HIPAA) Safe Harbour standards might not be sufficient to hinder re-identification attacks. These examples thus underlie the dire importance of going beyond current anonymization paradigms and proposing easy-to-implement and highly private approaches that transform the data itself when handling and sharing healthcare-derived data.

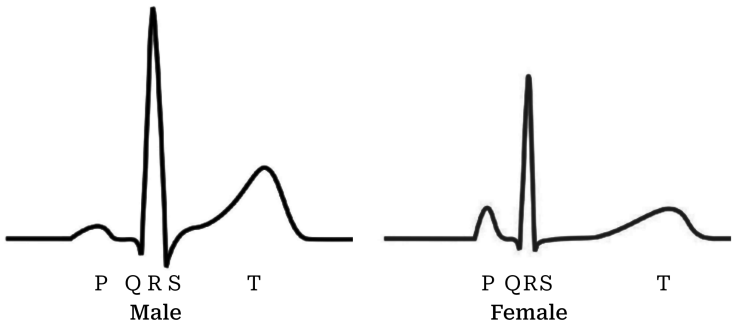

Figure 1: Differences in the ECG voltage for one heartbeat between the average male and female.

Generalizable Insights about Machine Learning in the Context of Healthcare

As the private sector increasingly seeks to employ re-identification strategies, and health insurance companies view such data as a valuable target, we contend that enhancing healthcare legislation to prioritize improved anonymization methods is essential, given the insufficiency of current approaches (Kaissis et al., 2020). However, as research on deep learning strategies for privacy and medical data remains scarce, we believe that this paper opens the door to modality-specific anonymization strategies that allow the sharing of private healthcare data while retaining the utility of the data itself. Moreover, we believe that sharing these findings will establish a foundation for both healthcare professionals and computer scientists to collaboratively foster enhanced patient data anonymization.

Our proposed approach constructs a generative adversarial network (GAN)-based and differs from past works by proposing for the first time a flexible ECG-focused privatization approach that is agnostic to the downstream task.

- We demonstrate for the first time that clinical 12-lead ECG data could be used for re-identification of patients with a high fidelity regardless of disease status in a database-independent manner with a siamese network.
- We propose a novel reconstruction loss, MSE_{ECG}, that leverages R-wave morphological alterations to induce sex-ambiguity.
- We demonstrate that GAN-based approach can increase privacy levels on ECG waveforms while maintaining high levels of utility.

- We present novel morphological analysis methods able to investigate ECG morphology.
- We demonstrate that sex information can be reinserted downstream of networks to both validate our method, but also as a potential tool to investigate the impact of sex-linked features on deep learning models.

2. Related Work

Although nascent, the field of privacy has begun to tackle ECG data. For example, Son et al. (2017) proposed to use a wearable sensor connected to a distant analysis server and induce privacy through a public key cryptosystem and signal scrambling. More recently, Noh et al. (2022) approached the problem by investigating so-called fiducial points representing key ECG features such as R-wave amplitude or the interval between the Q-wave and the T-wave (QT interval) and identified 9 that didn't induce privacy leakage, in other words, didn't yield any information about the private attribute of the individuals. They subsequently used these features for an ECG-based authentication scheme. Also, Yang and Wang (2022) tackled the ECG privacy issue by generating a novel privacy-preserving data transformation, the manipulable Haar transform, for wireless body sensors. On the other hand, Kang et al. (2021) have demonstrated that GAN-based approaches could be used to remove arrhythmia-linked features from ECG while preserving patient-defining characteristics on ambulatory ECGs to create biometric waveforms without apparent pathologies. Nevertheless, these methods primarily emphasize utilizing ECG as a biometric modality rather than a medical diagnostic instrument, and they seldom consider the comprehensive 12-lead decomposition employed in clinical settings. Consequently, these approaches do not aim to optimize the privacy-utility balance within a cardiological context.

In parallel, the field of privacy has progressed in the area of automatic speech recognition. Indeed, research groups have employed generative adversarial networks (Goodfellow et al., 2014) to train a generator able to regenerate a voice-derived spectrogram while minimizing the probability of a discriminator network to identify the speaker's sex (Sisman et al., 2021; Nelus and Martin, 2019; Stoidis and Cavallaro, 2022). However, these approaches often train on the real component of the short-time Fourier transform of the original voice, inducing potential data loss. Nonetheless, various groups demonstrated that these approaches could achieve high levels of anonymity while optimizing the trade-off with utility, here measured as an ability to identify the word pronounced in the resulting signal (Stoidis and Cavallaro, 2022; Ericsson et al., 2020).

3. Methods

3.1. Network architectures

The GAN is a generative model trained by a two-player minimax game that learns to generate new data from sample data. The GAN is composed of two models: a generator G and a discriminator D. In the proposed architecture, D takes the shape of a 1D CNN. Its task is to predict whether the input ECG belongs to a biological male or female, while G, a U-Net (Ronneberger et al., 2015), takes a sex-defined ECG E and learns to remove or obscure the sex-defining features to generate E' rendering it sex ambiguous. To improve

ambiguity in the generated E', a small Gaussian noise $\mathcal{Z} \sim \mathcal{N}(0, 0.1)$ was added to the bottleneck (Figure 2).

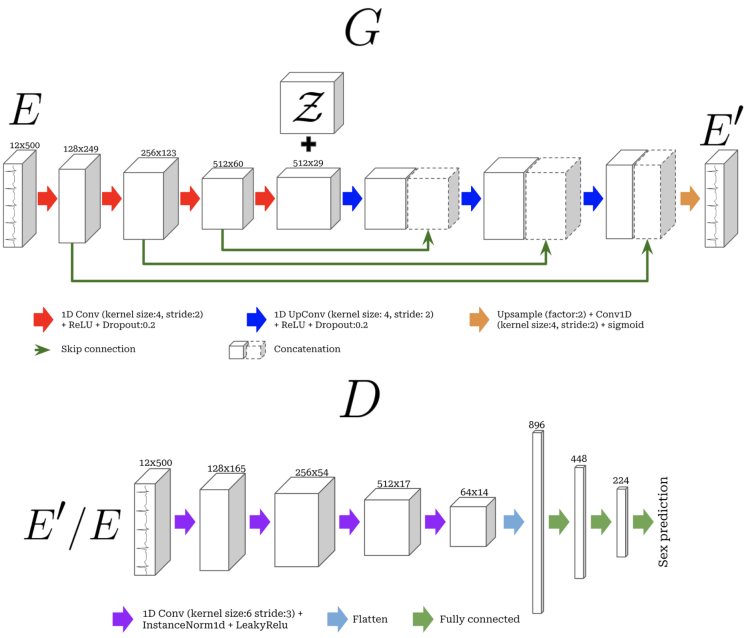

Figure 2: Architecture of the U-Net generator (top) and discriminator (bottom).

3.2. Generative adversarial training

Each ECG E is paired with its known private attribute, in this case, biological sex, in the label vector y. To generate a private sex target vector y_p, we based our approach on Stoidis and Cavallaro (2022)'s work, where we randomly sampled a Gaussian centred at the mean between the ground truth sex labels (0,1) with a small variance σ^2 of 0.05 such that $y_p \sim \mathcal{N}(0.5, 0.05)$ to promote the attenuation of sex-linked features and the amplification of sex-ambiguous ones by G. To further maximize sex ambiguity, y_p was regenerated at each training epoch.

The discriminator loss $\mathcal{L}_D$ can be expressed as the combination of the binary cross-entropy BCE between original label y and the predicted sex y' when presented with real examples E (BCE_E) and with the BCE between ambiguous sex label y_p and the predicted sex y' on generated E' examples ($BCE_{E'}$):

$$\mathcal{L}_D = BCE_E(y, y') + BCE_{E'}(y_p, y') \tag{1}$$

Thus, the discriminator learns to predict the original sex from the original waveforms while also predicting y_p the sex-ambiguous ECGs.

The generator loss $\mathcal{L}_G$ is composed of the BCE between the y' and y_p on the generated E', plus a variation of the mean squared error loss, we defined as MSE_{ECG}. Indeed, as

described by Carbone et al. (2020), one of the major differences between sexes on ECGs is the morphology of the R-wave (Figure 1).

Thus, to minimize the penalty attributed to the reconstruction of this region, we weighted R-wave and non-R-wave regions differently. The R-wave boolean mask $mask_R$ was obtained using Ivaylo I. Christov's QRS detection algorithm (Christov, 2004), with a ± 5 window around the detection point. The non-R-wave mask $mask_{NR}$ was considered as the inverse of $mask_R$ (Figure 3).

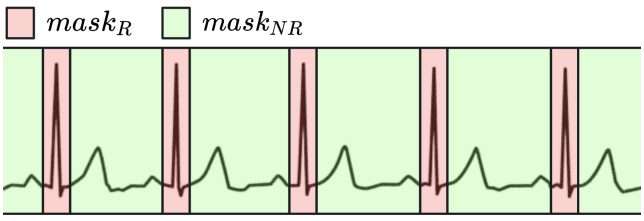

Figure 3: Representation of the placement of $mask_R$ and $mask_{NR}$ on the ECG waveform

We modified MSE into MSE_{ECG} by separating it into two parts, a weighted non-R-wave region by a factor λ and an unweighted R-wave region such that:

$$MSE_{ECG}(E, E') = MSE(E \cdot mask_R, E' \cdot mask_R) + MSE(E \cdot mask_{NR}, E' \cdot mask_{NR}) \cdot \lambda \quad (2)$$

The resulting $\mathcal{L}_G$ thus can be expressed as:

$$\mathcal{L}_G = MSE_{ECG}(E, E') + BCE(y_p, y') \cdot \epsilon \quad (3)$$

Thus, the generator needs to balance both generating ECGs as close as possible to the original waveform while being as close as possible to sex-ambiguous. Additionally, ϵ represents the dis-utility budget in the privacy-utility trade-off as presented by Tripathy et al. (2019). This value allows us to tune how severely the anonymization needs to be taken into account compared to the reconstruction quality during training.

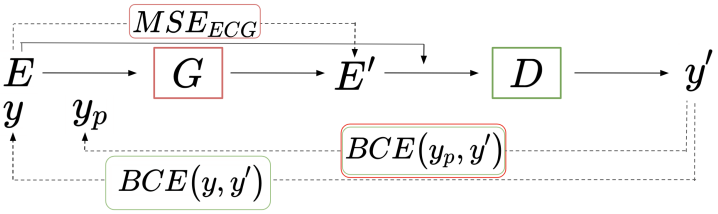

Figure 4: Training procedure of PrivECG-λ with components of $\mathcal{L}_G$ (red) and $\mathcal{L}_D$ (green).

3.3. Experimental models

We compare four approaches:

1. **PrivECG**: The aforementioned approach with the MSE_{ECG} replaced by the normal mean squared error between E and E' with an ϵ=0.0001. To have a MSE within a similar range to PrivECG-λ's for a fair comparison, we multiplied MSE by 100 as well. This acts as a control for MSE_{ECG}.

2. **PrivECG-λ**: The proposed approach using MSE_{ECG} with a λ=100 and an ϵ=0.0001.

3. **GenGAN-ECG**: A control approach derived from GenGAN by Stoidis and Cavallaro (2022). This approach was initially proposed for voice privatization and the original implementation was defined for spectrograms. We adapted it to our context using our GAN architecture. This methodology proposes the inclusion of a random vector $\mathcal{Z}\sim\mathcal{N}(0,1)$ that is inserted at the bottleneck of G as well as modifying the generator's loss to $\mathcal{L}_G = MSE(E, E') + BCE(y', y) \cdot \epsilon$.

4. **controlGAN**: A non-private GAN architecture. In this architecture, G attempts to recreate an ECG waveform from an input waveform and D attempts to discriminate generated against true waveforms. This allows quantifying the impact of the noise added by the generative approach and the addition of the random vector at the bottleneck of G.

All experiments were performed using the Adam optimizer (Gemp, 2019) with its hyperparameters set to $\mathcal{B}_1$=0.5 and $\mathcal{B}_2$=0.99 with a learning rate of 1e-7 for 500 epochs with mini-batches each containing 64 examples.

3.4. Dataset

We generated 3 distinct datasets from ECGs all obtained from the Montreal Heart Institute internal database. The first (dataset 1) was used for the training, validation, and testing of our GAN. Disease prediction included 50,000 10-second 12-lead ECGs with various cardiac illnesses obtained using the MUSE v9 Cardiology Information System at 250Hz with a unit voltage of 4.88 microvolts (see Appendix Table A.1). Once acquired, all examples were filtered using a bandpass filter from 0.1 to 45 Hz, standardized and resample such that each lead had 500 data points. The dataset was subdivided into balanced training, validation, and test sets on non-overlapping datasets based on patient sex, age, and diagnosis using scikit-multilearn's iterative dataset splitting function in a 0.6/0.2/0.2 split (Szymański and Kajdanowicz, 2017). We also ensured that patients with repeated ECG did not have ECG in different stratification to prevent data leakage as well as having a 1:1 female-to-male ratio. The sex label vector y was encoded from the biological sex such that:

$$y = \begin{cases} 0, & \text{if female} \\ 1, & \text{if male} \end{cases}$$

The second dataset (dataset 2) was used to train a siamese network (Koch, 2015) for the patient verification task. Briefly, this task verifies whether a pair of ECGs originated from

the same patient. This dataset was composed of 43850 ECGs from 877 distinct patients. Each patient had 50 ECGs sampled across 15 years. Following the same pre-processing as dataset 1, it was separated into a training set containing 50 true ECG pairs and 50 false pairs per patient where the siamese dataset S, produced from the raw dataset of ECG K, can be described as:

$$S = \{(ECG_i, ECG_j, z_{ij}), (i,j) \in K\} : z_{ij} \in \{0,1\} \tag{4}$$

where a pair is considered true if ECG_i and ECG_j come from the same patient where $z_{ij} = 1$ and $z_{ij} = 0$ if both ECGs are from different patients making a false pair. The 87700 resulting ECG pairs were subsequently subdivided into train, validation, and test sets similar to dataset 1.

To investigate disease prediction, we built a third dataset (dataset 3) containing 653 241 ECGS with 6 potential labels: atrial fibrillation, first-degree atrioventricular block, right bundle branch block, left bundle branch block, infarction and sinus rhythm (see Appendix Table A.2). Similar to the other datasets, we pretreated and split them into training, validation, and testing sets. None of the patients had ECGs shared between datasets to prevent data leakage.

3.5. Evaluation

3.5.1. Metrics

To evaluate this work, we propose three families of metrics, precision metrics, reconstruction metrics, and privacy metrics. The first set contains performance metrics to evaluate both sex-predictive and disease-predictive performances: accuracy, F1 score, area under the precision-recall curve (AUPR) and area under the receiver operating characteristic curve (AUC-ROC). The later is to facilitate the comparison with previously published studies on sex prediction from ECG waveforms. Briefly, the value of AUC-ROC ranges from 0 to 1, where a value under 0.5 indicates a classifier performing worse than "random", 0.5 indicates that it ranks a random positive example higher than a random negative example 50% of the time and 1 indicates perfect discrimination between the classes. Values nearing 1 indicate a better performance of the classifier. For multilabel disease prediction tasks, we propose the use of both the F1 score and the AUPR. Briefly, the F1 score represents the harmonic mean of its precision and recall of the algorithm. On the other hand, the AUPR is obtained by measuring a classifier's recall and precision at various thresholds and, similarly to the AUC-ROC, values close to 1 are favoured.

To evaluate the quality of the reconstruction, we propose a second set of metrics. First, the root mean squared error (RMSE) represents the standard deviation of the errors or distances between equivalent points between E and E'. The closer the RMSE is to 0 the closer E' is to E.

$$RMSE(E, E') = \sqrt{\frac{\sum_{i=1}^{N}(E_i - E'_i)^2}{N}} \tag{5}$$

where E_i and E'_i represent the ith position on the original and reconstructed waveform respectively. Although it theoretically should be optimal to obtain the smallest $RMSE$ pos-

sible, there remains interest to have some key differences between E and the sex-anonymous E'.

We also investigate the Fréchet distance (FD) (Godau, 1991) between the E and E' described as the smallest of the maximum pairwise distances between two curves. Similarly to RMSE, the closer FD is to 0 the closer E' is to E.

$$FD(E, E') = \min\left(\max_{i \in Q}\left(d\left(E_i, E'_i\right)\right)\right), Q = [1, N] \tag{6}$$

where d is the Euclidean distance and N is the length of the ECG vector. Similarly to the analysis of the RMSE, we wish to reproduce an ECG-like waveform with key differences induced by sex anonymization. Thus, although a low FD is desired, an FD of zero is equally undesired.

We also propose to evaluate the Pearson correlation coefficient ρ that measures how E co-varies with E'. The values vary from -1, representing a perfect inverse correlation, and 1 being an ideal correlation.

$$\rho(E, E') = \frac{\sum_{i=1}^{N}(E_i - \bar{E})(E'_i - \bar{E'})}{\sqrt{\sum_{i=1}^{N}(E_i - \bar{E})^2}\sqrt{\sum_{i=1}^{N}(E'_i - \bar{E'})^2}} \tag{7}$$

Where $\bar{E}$ and $\bar{E'}$ represent the mean of the original and reconstructed ECGs respectively. Again, a value of ρ tending towards 1 is desired for reconstruction, however, we wish to have sufficient alterations in E' to yield an imperfect score to permit the inclusion of sex ambiguity.

Finally, to evaluate the privacy-preserving ability of the experimental models, we propose employing the equal error rate (EER). The EER defines the set point where the false acceptance rate (FAR) equals the false rejection rate (FRR). The smaller the EER, the more secure the biometric modality, as fewer misidentifications occur. A highly private modality will have an EER approaching 50% where the likelihood of performing a FAR or FRR is equal.

3.5.2. Evaluation networks

Following the evaluation of the reconstruction from PrivECG and the other comparative approaches, three distinct networks were needed. A 1D-ResNet-50 (He et al., 2015) was trained to predict biological sex from ECG to validate that dataset 1 contained sex-linked information. A second 1D-ResNet-50 was trained on disease prediction to validate the generated data's ability to generate useful waveforms for disease prediction on dataset 3. To ensure the best predictions, this network was trained using focal loss citep(with a gamma of 2 (Lin et al., 2017). To allow a fair comparison of downstream analysis, the precise architecture was the same as for the Re-insertion of sex information.

To evaluate the privacy and disease-prediction ability of the regenerated sex-ambiguous ECG, we trained a siamese network based on two 1D-ResNet-50 feature extractors. The network was trained, validated, and tested on dataset 2. The task of the siamese network (Koch, 2015) was to validate if two ECGs instances came from the same patient to measure the final EER (Figure 5). Results presented are on the respective transformed test sets modified by the trained generators from each method tested.

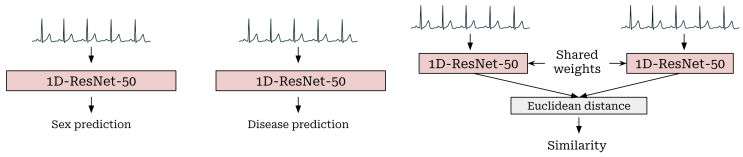

Figure 5: The sex-predicting (left) and disease-predicting (center) networks are both based on a 1D-ResNet-50 architecture trained on datasets 1 and 3 respectively. The siamese network (right) or the patient validation task trained on dataset 2.

3.5.3. R-wave analysis

The R-wave represents the depolarization of the heart's ventricle and, as ventricle size varies across males and females, it is one of the most apparent sex-specific morphological distinctions on the ECG waveform (Moss, 2010). To validate the baseline variation in R-wave morphologies between biological sexes within our dataset we used the Ivaylo I. Christov's QRS detection algorithm (Christov, 2004) and measured their height using scipy's `find_peak` function (Virtanen et al., 2020) in combination with Makowski et al. (2021)'s NeuroKit 2.0 `ecg_delineate` function to find the start and end of the R-wave. These were subsequently used to measure both the average amplitude and width of the baseline ECG cohort.

To evaluate the changes in R-wave morphology in the generated ECGs, we also localized the R-wave using the Ivaylo I. Christov's QRS detection algorithm (Christov, 2004). We calculated two metrics, the first being the average standard variation in R-wave amplitude within a single ECG waveform: σ_{R-wave}. A larger σ_{R-wave} represents a large variation in amplitudes across the ECG while a smaller value represents more consistent amplitudes. To better identify differences in R-wave amplitude from the original waveform, we propose the $\mu_{\Delta R-wave}$ metric which can be measured as the average mean difference from M ECG with P R peaks where Ag and Ao represent the peak's amplitude in the generated and original waveform respectively.

$$\mu_{\Delta R-wave} = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{P}\sum_{j=1}^{P}(Ag_j - Ao_j) \tag{8}$$

A smaller value for $\mu_{\Delta R-wave}$ thus represents a waveform with R-wave amplitude more similar to the original waveform.

To yield a non-generative R-wave normalized baseline, we also propose to normalize the per-lead R-wave amplitude using the per-lead gender-balanced average amplitude.

3.5.4. Re-insertion of sex information

To validate that the loss of performance observed from using the PrivECG-derived data resulted from the absence of the sex-linked features and not the generative process itself, we reinserted the sex information at the level of the dense layers following the 1D-ResNet-50 block. The output of the network was concatenated with a one-hot encoded vector of the

sex. To prevent gradient explosion, the positive class was set to 0.8 while the negative class was set to 0.1. For the control approaches, the same architecture was used, whereby the entry tensor to the dense layer was simply the 2048 features output of the 1D-ResNet-50. The network was also trained using focal loss (Lin et al., 2017) with a gamma of 2 (Figure 6).

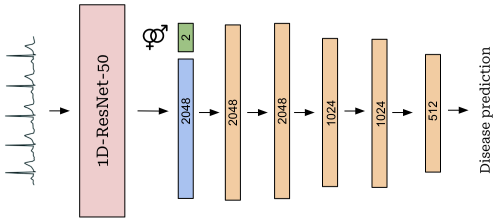

Figure 6: Architecture for evaluating the sex information re-insertion. The blue tensor represents the 1D-ResNet-50 output vector to which we combine the sex-label vector. No concatenation is performed for controls.

4. Results

4.1. Evaluating if sex-linked features were preserved in ECG of patients suffering from multiple illnesses

Works such as Attia et al. (2019) demonstrated that 12-lead ECG waveforms permitted the accurate prediction of the patient's sex. We set out to confirm these results in our dataset as it is composed of a wide range of diseased waveforms thus potentially altering the performance on such a task. We evaluated the ability to predict sex using a 1D-ResNet-50 (He et al., 2015). We compared our results to the declared AUC-ROC value from the original publication, as neither the trained model nor the data were available. From our results, we achieved an AUC-ROC of 0.882 ± 0.022 while Attia et al. (2019) were able to achieve results of 0.969. Thus, although sex seems more difficult to predict in our database, potentially due to our different disease distribution, sex prediction still yields high algorithmic performance and thus permitted us to move forward with our generative approach.

4.2. Evaluating the sex-linked morphological R-wave differences in our dataset

To ensure that R-wave differences indeed existed between biological sexes in our dataset, we investigated the changes in amplitude and width of R-waves between groups (Table 1).

These results suggest that true differences exist between biological sexes regarding the R-waves morphology on the baseline ECG validating the potential of our proposed method.

Table 1: Variation in R-wave amplitude and width in the baseline dataset.

R-wave amplitude		**R-wave width**	
Female	Male	Female	Male
0.439±0.003	**0.464±0.003**	4.902±0.123	**5.174±0.147**

4.3. Generated ECG

Next, we evaluated the quality of the reconstruction to ensure the downstream analysis was based on quality generated data as well as helping guide the hypothesis on obtained results.

Thus, following training, we contrasted the final models comparing the original ECG with the generated ECG according to the RMSE, FD and ρ (Table 2).

Table 2: Reconstruction metrics for tested approach with the 95% confidence interval

Model	**RMSE**	**FD**	$\boldsymbol{\rho}$
PrivECG	0.083 ± 0.001	0.444 ± 0.003	0.813 ± 0.001
PrivECG-λ	0.141 ± 0.001	0.494 ± 0.004	0.736 ± 0.002
GenGAN-ECG	0.112 ± 0.002	0.483 ± 0.005	0.767 ± 0.001
controlGAN	0.071 ± 0.001	0.395 ± 0.004	0.854 ± 0.001

From our reconstruction quality results, we validated that the control, controlGAN, performed as expected with low RMSE and FD values as well as a high ρ when compared to the entry waveform. Both PrivECG and GenGAN-ECG performed similarly with accurate regeneration metrics. Notably, PrivECG-λ offered the worse RMSE and FD while providing the lowest ρ within the compared approaches. However, GenGAN-ECG provided very close results. Indeed visual analysis of Figure 7 demonstrates that both methods are characterized by flared T-wave (Figure 1) and wider R-wave potentially explaining those results.

Indeed, all methods generated waveforms qualitatively similar to the original (Green). Intriguingly, both the PrivECG-λ and GenGAN-ECG have a wide T-wave appearance while PrivECG and controlGAN produced more similar morphologies to the original data. Moreover, both PrivECG methods seem to generate "wavelets" in the segment joining each PQRST complex (TP segment). These might be the result of the addition of the random Gaussian vector at the bottleneck of G, while the smaller amplitude wavelets of GenGAN-ECG are most likely obtained from the inclusion of the random vector yielding this smaller effect. Even if there is no clear sex-linked biological correlate, this might be a strategy employed by G to further trick D. Uniquely, PrivECG-λ appears to generate R-waves with less defined borders with a fusion with the P-wave. This could be a strategy to further blur the actual duration of that section of the cardiac cycle, improving sex ambiguity. As described by Carbone et al. (2020), not only does the R-wave amplitudes but also their widths vary, further helping to confuse D. To quantify the changes to the R-wave's amplitude, we

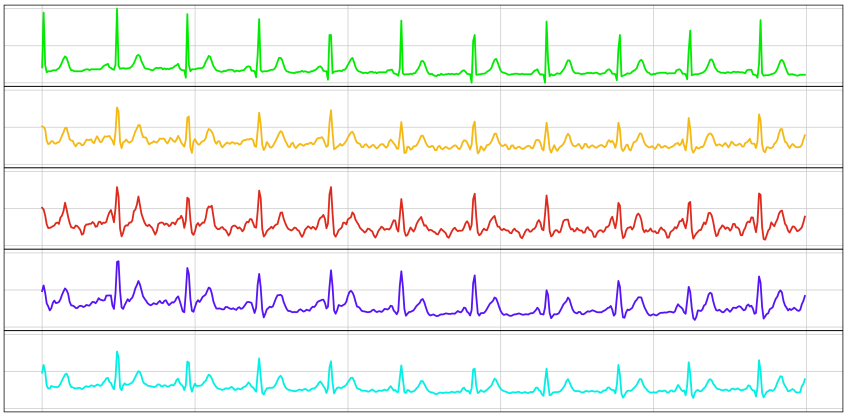

Figure 7: Example of lead I waveforms. The green waveform represents the original data while the orange and red represent PrivECG and PrivECG-λ respectively. The purple and cyan waveforms represent the GenGAN-ECG and controlGAN results.

investigated the changes to the uniformity of the R-wave's amplitude σ_{R-wave} and also the difference from the original $\mu_{\Delta R-wave}$ (Table 3).

Table 3: R-wave analysis where σ_{R-wave} represents the average intra-ECG standard deviation in amplitude while $\mu_{\Delta R-wave}$ represents the average difference in amplitude with the original ECG's

Model	$\boldsymbol{\sigma_{R-wave}}$	$\boldsymbol{\mu_{\Delta R-wave}}$
Original	0.174 ± 0.001	NA
PrivECG	0.137 ± 0.001	0.144 ± 0.001
PrivECG-λ	0.198 ± 0.001	0.160 ± 0.001
GenGAN-ECG	0.162 ± 0.001	0.102 ± 0.001
controlGAN	0.142 ± 0.001	0.112 ± 0.001

Similar to reconstruction metrics, controlGAN produced results approximating the original input as expected. However, PrivECG-λ showed to produce the R-waves with the greatest difference with the original waveform and generating the greatest variation within the waveform. Similar effects are seen in PrivECG and GenGAN-ECG but to a lesser extent. This could further demonstrate the strategy that the networks use to confuse the discriminator by modifying the R-waves amplitude but also increasing its variation across the generated ECG, thus making the sex prediction more difficult. However, controlGAN demonstrates to yield ECG waveforms with $\mu_{\Delta R-wave}$ values in between those of PrivECG

and GenGAN-ECG. This could be explained by the difficulty to recreate tall spindle-like R-waves thus yielding the short, yet uniform, R-waves as present in Figure 7.

To validate the impact of these reconstructions on the sex-prediction task we have subsequently tested these waveforms with their original sex labels on our sex-predicting network.

Table 4: Validation results of the reconstructions

Model	**Sex prediction accuracy**	**EER**	**F1 score**	**AUPR**
Original	0.876±0.004	0.098±0.005	0.931±0.006	0.987±0.002
PrivECG	0.686±0.012	0.177±0.003	0.900±0.009	0.932±0.004
PrivECG-λ	0.529±0.014	0.251±0.008	0.885±0.005	0.877±0.003
GenGAN-ECG	0.563±0.009	0.176±0.006	0.909±0.004	0.945±0.007
controlGAN	0.771±0.007	0.132±0.002	0.921±0.003	0.960±0.006
non-generative	0.512±0.002	0.107±0.005	0.956±0.004	0.872±0.007

These results, presented in Table 4, demonstrate that both PrivECG, GenGAN-ECG and PrivECG-λ significantly decreased the ability to predict the patient's sex. However, only PrivECG-λ made the sex prediction impossible with an accuracy of 0.529. These results translated to the patient validation task where the EER diminished for GenGAN-ECG and PrivECG-λ while staying, as expected, only slightly affected by the generative process as demonstrated by the results for the controlGAN condition. Furthermore, the results obtained with controlGAN demonstrate the importance of noise-sensitive features for accurate sex-prediction potentially further explaining our lower baseline accuracy compared to Attia et al. (2019). Additionally, the non-generative control demonstrates that the adjustment of the R-wave amplitude causes sex-ambiguization however does little effect on the privacy illustrated by the small change induced on the EER. The significant loss in AUPR could be both linked with the resulting R-wave deformation causing issues with the detection of diseases such as infarction.

4.4. Rescuing sex-linked features

To validate that the impact on performance was actually linked to the lack of sex-linked features, we attempted to "rescue" the network's accuracy by re-injecting those features by concatenating them with the convolution features before the dense layers of the CNN. Our results show that the initial PrivECG-λ with an F1 score of 0.885±0.005 jumped to 0.893±0.004. Although small, this change demonstrates that the sex-linked information can be rescued as well as telling of the small effect that sex-linked information has on overall prediction performance. This demonstrates that sex-linked tasks could potentially be accomplished by also sharing an encrypted sex-linked vector and reinserting downstream of the network without inducing major performance loss.

5. Discussion

In this work, we present a methodology to generate ECG-specific privatization methods. Our approach investigates the generation of privatized ECG waveforms through the privatization of sex-linked features without severely affecting disease-dependent ones using generative approaches.

We initially validated that the patient's sex was accurately detectable using convolution methods. Moreover, our premise of targeting the R-wave as a potential target for sex reidentification appeared to be valid in our dataset, with a significant difference in terms of width and amplitude existing in our dataset (Table 1). Thus, our strategy seemed appropriate as a potential privatization methods

Indeed, as presented in table 4, most methods attempting to regenerate ECG waveforms without sex-identifying information produced similar levels of utility. However, the biggest difference was in the EER obtained where PrivECG-λ outshined competitive approaches. This could be explained by the flared T-wave as well as the larger R-wave that often connects with the neighbouring P-wave (Figure 7) generated through the usage of our MSE_{ECG} loss. Interestingly similar results were obtained by PrivECG and GenGAN-ECG demonstrating the usefulness of those morphologies to better trick the discriminator. Furthermore, morphological analysis of the obtained ECGs demonstrates that PrivECG-λ strives towards a greater difference from the input (Table 2) mainly through a larger variation in R-wave amplitude within the waveform (σ_{R-wave}) and a larger difference between the input ($\mu_{\Delta R-wave}$). Indeed, these most likely further confuse the ability of an attacker to distinguish the biological sex of the patient as R-wave amplitude is linked to the heart's ventricle mass that is, in part, related to the biological sex of the patient Carbone et al. (2020). However, these could worsen predictive performances on diseases that are identified through R-wave width or the PR segment such as atrial infarction (Zbiciak and Markiewicz, 2023).

We subsequently investigate the ability to reintegrate the sex information later in the network to validate that indeed the drop in performance was due to the ambiguization or removal of sex-linked features. To do so, we opted to concatenate a one-hot encoded vector of the sex of the patient, as presented in Figure 6. We demonstrated that performance improved by this methodology, indicating that roughly 0.01 points in accuracy were directly lost due to the lack of clear sex information in the network. Although not significatively better, this could act as a proxy method to disentangle the impact of sex-relevant features on potentially more sex-dependant deep learning tasks and could be used to further investigate the impact of such private attributes in a typical ECG-based prediction algorithm.

Thus, our results suggest that using MSE_{ECG} combined with GAN-based approach could be a promising avenue to generate private ECG waveforms that still permit the training of accurate deep learning models.

Limitations Although we deem our approach interesting, we recognize various limitations. First, our utility has been validated on a handful of ECG-bound conditions and thus would need to validate the true utility on a much wider array of cardiac illnesses where the artifact introduced could have a more major impact on the prediction quality. Also, although we demonstrated that the loss of utility was minimal, it is likely that sex $\times$ disease interactions exist, and it is unclear how these would impact the generalization capacity on

external datasets. Thus, potentially privatizing another attribute, such as age or ethnicity, might be a safer objective to guarantee a better privacy-utility trade-off. However, it must be stated that age affects, the ECG by a slowdown in heart rate but also by the accumulation of heart abnormalities. Therefore, introducing age ambiguity would necessitate the generator to incorporate disease-related features into younger patients' ECGs, resulting in a less desirable privacy-utility tradeoff.

Future work As we believe that improving privacy in healthcare is a key issue, we have the intention to further update these approaches to test the prioritization of a combination of various attributes including sex, age and ethnicity in a 12-lead format to improve the safety of ECG data-sharing and further prevent re-identification risks beyond the current paradigms. Moreover, we aim to include a re-identification network within the adversarial framework to further increase the resulting EER and thus further improve the privacy of the algorithmic solution.

References

Zachi I. Attia, Paul A. Friedman, Peter A. Noseworthy, Francisco Lopez-Jimenez, Dorothy J. Ladewig, Gaurav Satam, Patricia A. Pellikka, Thomas M. Munger, Samuel J. Asirvatham, Christopher G. Scott, Rickey E. Carter, and Suraj Kapa. Age and sex estimation using artificial intelligence from standard 12-lead ECGs. *Circulation: Arrhythmia and Electrophysiology*, 12(9), September 2019. doi: 10.1161/circep.119.007284. URL https://doi.org/10.1161/circep.119.007284.

Daniel E. Becker. Fundamentals of electrocardiography interpretation. *Anesthesia Progress*, 53(2):53–64, June 2006. doi: 10.2344/0003-3006(2006)53[53:foei]2.0.co;2. URL https://doi.org/10.2344/0003-3006(2006)53[53:foei]2.0.co;2.

Gordon Begg, Kathryn Willan, Keith Tyndall, Chris Pepper, and Muzahir Tayebjee. Electrocardiogram interpretation and arrhythmia management: a primary and secondary care survey. *British Journal of General Practice*, 66(646):e291–e296, March 2016. doi: 10.3399/bjgp16x684781. URL https://doi.org/10.3399/bjgp16x684781.

Vincenzo Carbone, Franco Guarnaccia, Giovanni Carbone, Giovanni Battista Zito, Ugo Oliviero, Silvia Soreca, and Francesca Carbone. Gender differences in the 12-lead electrocardiogram: clinical implications and prospects. *Italian Journal of Gender-Specific Medicine*, (2020September-December), September 2020. doi: 10.1723/3432.34217. URL https://doi.org/10.1723/3432.34217.

Ivaylo I Christov. Real time electrocardiogram QRS detection using combined adaptive threshold. *BioMedical Engineering OnLine*, 3(1), August 2004. doi: 10.1186/1475-925x-3-28. URL https://doi.org/10.1186/1475-925x-3-28.

David Ericsson, Adam Östberg, Edvin Listo Zec, John Martinsson, and Olof Mogren. Adversarial representation learning for private speech generation, 2020. URL https://arxiv.org/abs/2006.09114.

Annette Flanagin, Gregory Curfman, and Kirsten Bibbins-Domingo. Data sharing and the growth of medical knowledge. *JAMA*, 328(24):2398, December 2022. doi: 10.1001/jama.2022.22837. URL https://doi.org/10.1001/jama.2022.22837.

Ian M. Gemp. The unreasonable effectiveness of adam on cycles. 2019.

Arin Ghazarian, Jianwei Zheng, Daniele Struppa, and Cyril Rakovski. Assessing the re-identification risks posed by deep learning algorithms applied to ECG data. *IEEE Access*, 10:68711–68723, 2022. doi: 10.1109/access.2022.3185615. URL https://doi.org/10.1109/access.2022.3185615.

Michael Godau. A natural metric for curves — computing the distance for polygonal chains and approximation algorithms. In *STACS 91*, pages 127–136. Springer-Verlag, 1991. doi: 10.1007/bfb0020793. URL https://doi.org/10.1007/bfb0020793.

Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. PhysioBank, PhysioToolkit, and PhysioNet. *Circulation*, 101(23), June 2000. doi: 10.1161/01.cir.101.23.e215. URL https://doi.org/10.1161/01.cir.101.23.e215.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014. URL https://arxiv.org/abs/1406.2661.

Awni Y. Hannun, Pranav Rajpurkar, Masoumeh Haghpanahi, Geoffrey H. Tison, Codie Bourn, Mintu P. Turakhia, and Andrew Y. Ng. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, 25(1):65–69, January 2019. doi: 10.1038/s41591-018-0268-3. URL https://doi.org/10.1038/s41591-018-0268-3.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. URL https://arxiv.org/abs/1512.03385.

J. Weston Hughes, Jeffrey E. Olgin, Robert Avram, Sean A. Abreau, Taylor Sittler, Kaahan Radia, Henry Hsia, Tomos Walters, Byron Lee, Joseph E. Gonzalez, and Geoffrey H. Tison. Performance of a convolutional neural network and explainability technique for 12-lead electrocardiogram interpretation. *JAMA Cardiology*, 6(11):1285, November 2021. doi: 10.1001/jamacardio.2021.2746. URL https://doi.org/10.1001/jamacardio.2021.2746.

Georgios A. Kaissis, Marcus R. Makowski, Daniel Rückert, and Rickmer F. Braren. Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*, 2(6):305–311, June 2020. doi: 10.1038/s42256-020-0186-1. URL https://doi.org/10.1038/s42256-020-0186-1.

Youngshin Kang, Hyeonbin Lee, Youngioo Shin, and Cheolsoo Park. A secure and privacy-preserving ECG-based personal authentication. In *2021 IEEE International Conference on Smart Internet of Things (SmartIoT)*. IEEE, August 2021. doi:

10.1109/smartiot52359.2021.00072. URL https://doi.org/10.1109/smartiot52359.2021.00072.

Paul Kligfield, Leonard S. Gettes, James J. Bailey, Rory Childers, Barbara J. Deal, E. William Hancock, Gerard van Herpen, Jan A. Kors, Peter Macfarlane, David M. Mirvis, Olle Pahlm, Pentti Rautaharju, and Galen S. Wagner. Recommendations for the standardization and interpretation of the electrocardiogram. *Circulation*, 115 (10):1306–1324, March 2007. doi: 10.1161/circulationaha.106.180200. URL https://doi.org/10.1161/circulationaha.106.180200.

Gregory R. Koch. Siamese neural networks for one-shot image recognition. 2015.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection, 2017. URL https://arxiv.org/abs/1708.02002.

Dominique Makowski, Tam Pham, Zen J. Lau, Jan C. Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and S. H. Annabel Chen. NeuroKit2: A python toolbox for neurophysiological signal processing. *Behavior Research Methods*, 53(4):1689–1696, feb 2021. doi: 10.3758/s13428-020-01516-y. URL https://doi.org/10.3758%2Fs13428-020-01516-y.

Pietro Melzi, Ruben Tolosana, and Ruben Vera-Rodriguez. Ecg biometric recognition: Review, system proposal, and benchmark evaluation, 2022. URL https://arxiv.org/abs/2204.03992.

Arthur J. Moss. Gender differences in ECG parameters and their clinical implications. *Annals of Noninvasive Electrocardiology*, 15(1):1–2, January 2010. doi: 10.1111/j.1542-474x.2009.00345.x. URL https://doi.org/10.1111/j.1542-474x.2009.00345.x.

Blake Murdoch. Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Medical Ethics*, 22(1), September 2021. doi: 10.1186/s12910-021-00687-3. URL https://doi.org/10.1186/s12910-021-00687-3.

Alexandru Nelus and Rainer Martin. Privacy-aware feature extraction for gender discrimination versus speaker identification. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, May 2019. doi: 10.1109/icassp.2019.8682394. URL https://doi.org/10.1109/icassp.2019.8682394.

Seungil Noh, Jaehan Kim, Seokmin Lee, Youngshin Kang, Cheolsoo Park, and Youngjoo Shin. Broken heart: Privacy leakage analysis on ECG-based authentication schemes. *Security and Communication Networks*, 2022:1–14, September 2022. doi: 10.1155/2022/7997509. URL https://doi.org/10.1155/2022/7997509.

Kai Packhäuser, Sebastian Gündel, Nicolas Münster, Christopher Syben, Vincent Christlein, and Andreas Maier. Deep learning-based patient re-identification is able to exploit the biometric nature of medical chest x-ray data. *Scientific Reports*, 12(1), September 2022. doi: 10.1038/s41598-022-19045-3. URL https://doi.org/10.1038/s41598-022-19045-3.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015. URL https://arxiv.org/abs/1505.04597.

Rhodri Saunders and Julie Lankiewicz. The cost effectiveness of single-patient-use electrocardiograph cable and lead systems in monitoring for coronary artery bypass graft surgery. *Frontiers in Cardiovascular Medicine*, 6, May 2019. doi: 10.3389/fcvm.2019.00061. URL https://doi.org/10.3389/fcvm.2019.00061.

Christopher G. Schwarz, Walter K. Kremers, Terry M. Therneau, Richard R. Sharp, Jeffrey L. Gunter, Prashanthi Vemuri, Arvin Arani, Anthony J. Spychalla, Kejal Kantarci, David S. Knopman, Ronald C. Petersen, and Clifford R. Jack. Identification of anonymous MRI research participants with face-recognition software. *New England Journal of Medicine*, 381(17):1684–1686, October 2019. doi: 10.1056/nejmc1908881. URL https://doi.org/10.1056/nejmc1908881.

Berrak Sisman, Junichi Yamagishi, Simon King, and Haizhou Li. An overview of voice conversion and its challenges: From statistical modeling to deep learning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:132–157, 2021. doi: 10.1109/taslp.2020.3038524. URL https://doi.org/10.1109/taslp.2020.3038524.

Sulaiman Somani, Adam J Russak, Felix Richter, Shan Zhao, Akhil Vaid, Fayzan Chaudhry, Jessica K De Freitas, Nidhi Naik, Riccardo Miotto, Girish N Nadkarni, Jagat Narula, Edgar Argulian, and Benjamin S Glicksberg. Deep learning and the electrocardiogram: review of the current state-of-the-art. *EP Europace*, 23(8):1179–1191, February 2021. doi: 10.1093/europace/euaa377. URL https://doi.org/10.1093/europace/euaa377.

Junggab Son, Juyoung Park, Heekuck Oh, Md Bhuiyan, Junbeom Hur, and Kyungtae Kang. Privacy-preserving electrocardiogram monitoring for intelligent arrhythmia detection. *Sensors*, 17(6):1360, June 2017. doi: 10.3390/s17061360. URL https://doi.org/10.3390/s17061360.

Dimitrios Stoidis and Andrea Cavallaro. Generating gender-ambiguous voices for privacy-preserving speech recognition, 2022. URL https://arxiv.org/abs/2207.01052.

Latanya Sweeney, Ji Su Yoo, Laura J. Perovich, Katherine E. Boronow, Phil Brown, and Julia Green Brody. Re-identification risks in hipaa safe harbor data: A study of data from one environmental health study. *Technology science*, 2017.

P. Szymański and T. Kajdanowicz. A scikit-based Python environment for performing multi-label classification. *ArXiv e-prints*, February 2017.

Adam Tanner. Our bodies, our data: How companies make billions selling our medical records, Jan 2017. URL https://books.google.com/books/about/Our_Bodies_Our_Data.html?id=IxuPEAAAQBAJ.

Ardhendu Tripathy, Ye Wang, and Prakash Ishwar. Privacy-preserving adversarial networks. In *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, September 2019. doi: 10.1109/allerton.2019.8919758. URL https://doi.org/10.1109/allerton.2019.8919758.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

Wencheng Yang and Song Wang. A privacy-preserving ECG-based authentication system for securing wireless body sensor networks. *IEEE Internet of Things Journal*, 9(8):6148–6158, April 2022. doi: 10.1109/jiot.2021.3109609. URL https://doi.org/10.1109/jiot.2021.3109609.

Adrian Zbiciak and Tymon Markiewicz. A new extraordinary means of appeal in the polish criminal procedure: the basic principles of a fair trial and a complaint against a cassatory judgment. *Access to Justice in Eastern Europe*, 6(2):1–18, March 2023.

Appendix A.

Table A.1: Top 15 diagnosis on dataset 1 ECGs. (Multiple labels can be associated with the same waveform).

Label	Occurrences
Normal sinusal rhythm	18 015
Sinusal bradycardia	6 241
Inferior wall infarction	6 123
Left axial deviation	5 258
Complete right bundle branch block	3 574
Atrial fibrillation	2 595
Non-specific T wave abnormality	2 426
Complete left bundle branch block	2 294
First degree A-V block	2 238
Prolonged QT	1 680
Left atrial dilatation	1 584
Anterior wall infarction	1 500
Left anterior hemiblock	1 402
Non-specific intraventricular conduction disorder	1 377
Septal infarction	1 351

Table A.2: Distribution of labels in dataset 3 for multilabel classification.

Label	Occurrences
Atrial fibrillation	68 303
First-degree atrioventricular block	64 508
Right bundle branch block	63 999
Left bundle branch block	39 050
Infarction	161 603
Sinus rhythm	482 240