>>> **Privacy-preserving Data Generation:**
>>> Towards generating privacy-preserving, synthetic and useful time series ECG data for anomaly detection

KTH x RISE

Sijun John Tu

March 6, 2024

>>> Outline

\>>> Outline

1. Project introduction
2. Heartbeat Arrhythmia
3. Privacy-preserving Time Series Data Generation
4. Results
5. Summary
6. References

>>> Machine Learning Pipeline

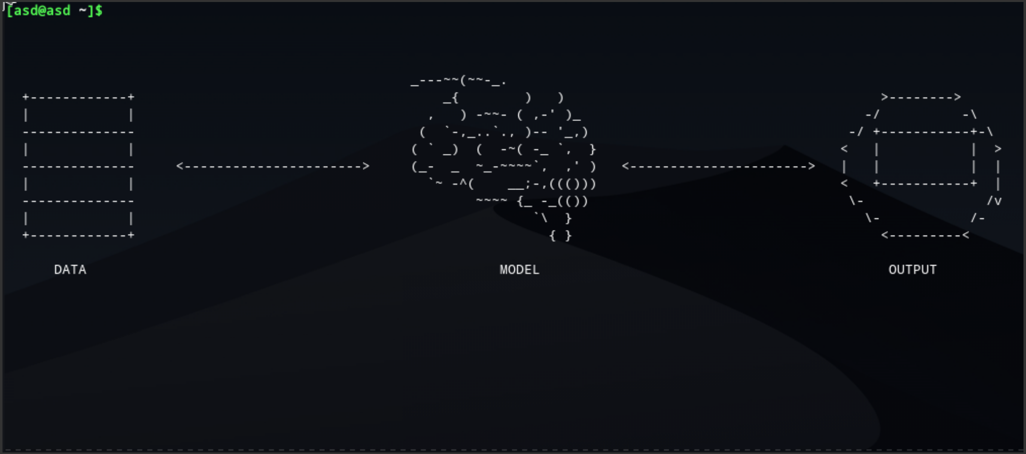

Figure: High-level machine learning pipeline

>>> Anomaly detection using privacy-preserving, synthetic time series data

$ Problem

- ML models are very data hungry.
- In many cases sharing data comes with privacy risks.

$ Solution:

- Promising solution: `synthetic data` with privacy guarantees!
- Synthetic data with `differential private` (DP) guarantees is a promising solution to ensure privacy independent of downstream task.

$ BUT:

- `Privacy-Utility-Tradeoff`: Commonly, a gain in privacy results in a loss of utility.
- For `anomaly detection` this might not be the case (?).

Goal: generate useful and privacy-preserving ECG data for anomaly detection (heartbeat arrhythmia).

>>> Structure

1. Train baseline model for anomaly detection only on regular heartbeat data using an LSTM-AE.
2. Generate heartbeat data (without DP) using two approaches:
   - AE-MERF
   - RTSGAN
3. Train LSTM-AE for anomaly detection on synthetic data and test on real (TSTR).
4. Add DP noise and repeat:
   - AE-DPMERF
   - DP-RTSGAN
5. Contaminate training data with anomalous heartbeats and repeat

>>> Outline

>>> Heartbeat Arrhythmia

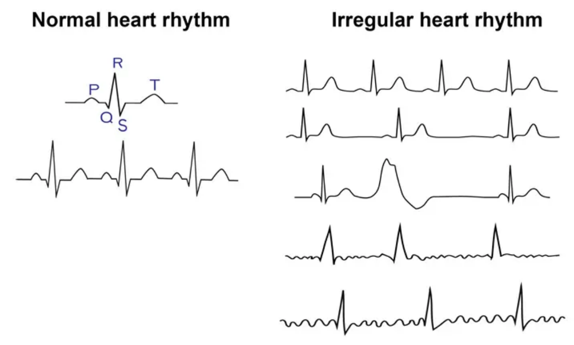

Figure: Different heartbeat arrhythmias [1]

[1] source: https://www.parkwayshenton.com.sg/health-plus/article/arrhythmia-guide

>>> Arrhythmia Detection as an Anomaly Detection Problem

We treat the problem of detecting irregular heartbeats as an anomaly detection problem from machine learning based on the reconstruction error:

- $ We train a model on regular heartbeats that is able to reconstruct that regular heartbeat.
- $ Given an irregular heartbeat the model should give higher reconstruction error.
- $ Based on an optimal threshold for that error we classify this heartbeat as either regular or irregular.

Two reasons for this semi-supervised approach: high class imbalancy and no need for labelling.

>>> Baseline Model

Model is a LSTM-AE that is trained only on normal samples with the goal to reconstruct normal samples. The classification is made based on the reconstruction error.

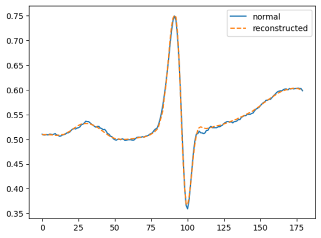

Figure: reconstruction on normal sample

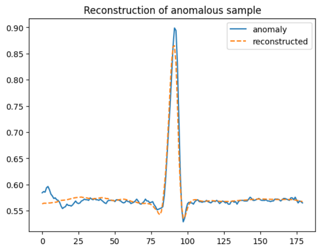

Figure: reconstruction on anomalous sample

\>>> Classification based on Reconstruction Error

>>> Outline

>>> Review: Differential Privacy

Idea. Hide the influence of one particular sample on the output of the model by adding randomness.

>>> Review: Differential Privacy

Idea. Hide the influence of one particular sample on the output of the model by adding randomness.

Definition (Differential Privacy)

A randomised algorithm $\mathcal{M}$ is (ϵ, δ)- differentially private if for all set of outcomes $S \subset ran\mathcal{M}$ and for all databases x, y, such that they **only differ in one element**, we have

$$\mathbb{P}(\mathcal{M}(x) \in S) \leq e^{\epsilon} \cdot \mathbb{P}(\mathcal{M}(y) \in S) + \delta \quad , \tag{1}$$

where the probability is taken over the randomness of $\mathcal{M}$.

>>> Review: Differential Privacy

Definition (Differential Privacy)

A randomised algorithm $\mathcal{M}$ is (ϵ, δ)- differentially private if for all set of outcomes $S \subset ran\mathcal{M}$ and for all databases x, y, such that they **only differ in one element**, we have

$$\mathbb{P}(\mathcal{M}(x) \in S) \leq e^{\epsilon} \cdot \mathbb{P}(\mathcal{M}(y) \in S) + \delta \quad , \tag{1}$$

where the probability is taken over the randomness of $\mathcal{M}$.

Informally. Replacing one record in the data will not change the outcome of algorithm $\mathcal{M}$ *too much* (specified via privacy budget ϵ). The lower ϵ the stricter the privacy guarantees.

>>> Models

AE-(dp)MERF

- $ AE-(dp)MERF is based on DP-MERF (best state of the art generator for tabular data).

>>> Models

AE-(dp)MERF

- $ AE-(dp)MERF is based on DP-MERF (best state of the art generator for tabular data).
- $ Simple architecture with mathematically sophisticated loss function.

>>> Models

AE-(dp)MERF

- $ AE-(dp)MERF is based on DP-MERF (best state of the art generator for tabular data).
- $ Simple architecture with mathematically sophisticated loss function.
- $ Does not work with time series data out of the box, but we will modify it so it works.

>>> Models

AE-(dp)MERF

- $ AE-(dp)MERF is based on DP-MERF (best state of the art generator for tabular data).
- $ Simple architecture with mathematically sophisticated loss function.
- $ Does not work with time series data out of the box, but we will modify it so it works.

AE-(dp)WGAN

\>>> **Models**

AE-(dp)MERF

- $ AE-(dp)MERF is based on DP-MERF (best state of the art generator for tabular data).
- $ Simple architecture with mathematically sophisticated loss function.
- $ Does not work with time series data out of the box, but we will modify it so it works.

AE-(dp)WGAN

- $ Model based on GAN network, which are commonly used in image generation.

>>> Models

AE-(dp)MERF

- $ AE-(dp)MERF is based on DP-MERF (best state of the art generator for tabular data).
- $ Simple architecture with mathematically sophisticated loss function.
- $ Does not work with time series data out of the box, but we will modify it so it works.

AE-(dp)WGAN

- $ Model based on GAN network, which are commonly used in image generation.
- $ Delivers state of the art performance for time series data.

\>>> Models

AE-(dp)MERF

- $ AE-(dp)MERF is based on DP-MERF (best state of the art generator for tabular data).
- $ Simple architecture with mathematically sophisticated loss function.
- $ Does not work with time series data out of the box, but we will modify it so it works.

AE-(dp)WGAN

- $ Model based on GAN network, which are commonly used in image generation.
- $ Delivers state of the art performance for time series data.
- $ No private counterpart, hence we will implement our own private version.

>>> Outline

1. Project introduction
2. Heartbeat Arrhythmia
3. Privacy-preserving Time Series Data Generation
4. Results
5. Summary
6. References

>>> AE-(DP)MERF

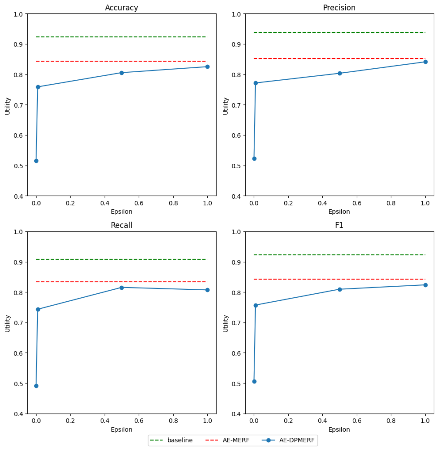

Figure: Results of AE-(DP)MERF with different privacy budgets (lower epsilon means higher privacy)

>>> (DP-)RTSGAN

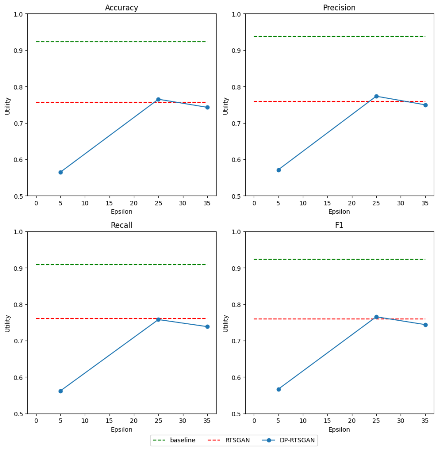

Figure: Results of (DP-)RTSGAN with different privacy budgets (lower epsilon means higher privacy)

>>> Conclusion

- $ DPMERF performs best and is very efficient computationally.
- $ DPMERF can work in lower epsilon ranges, which translates to higher privacy guarantees.
- $ DP-RTSGAN gives worse generative performance and can only work with meaningless privacy budgets epsilon.
- $ Adding privacy does not impact the utility for anomaly detection too much until too much noise is added.

\>>> Contamination

We contaminate the train set that only consists of normal samples with 1%, 2%, 5% anomalous samples (the percentage of heartbeat arrhythmias is estimated to be around max. 5%).

>>> Contamination: AE-(DP)MERF

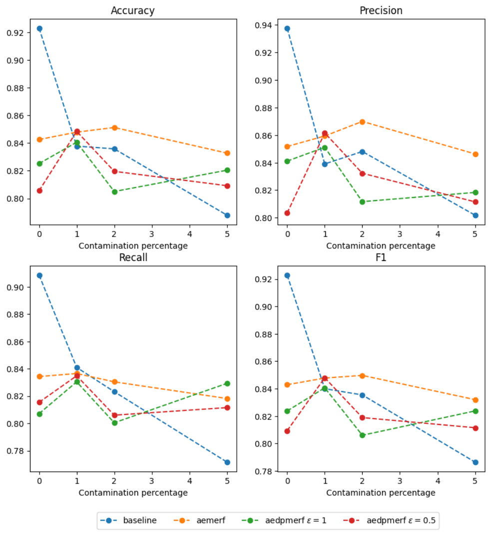

Figure: Contaminated training set: AE-(DP)MERF

>>> Contamination: AE-(DP)MERF

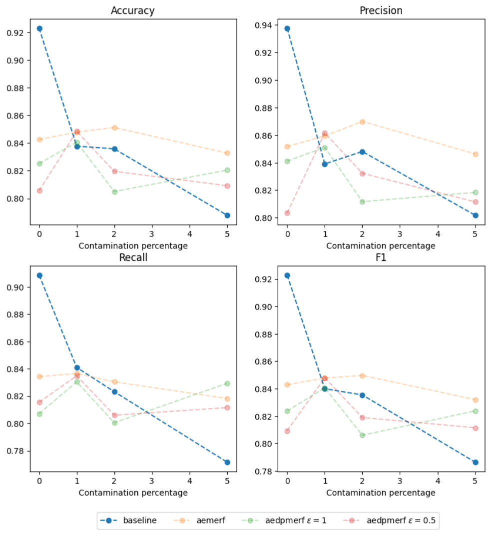

Figure: Contaminated training set: AE-(DP)MERF

>>> Contamination: AE-(DP)MERF

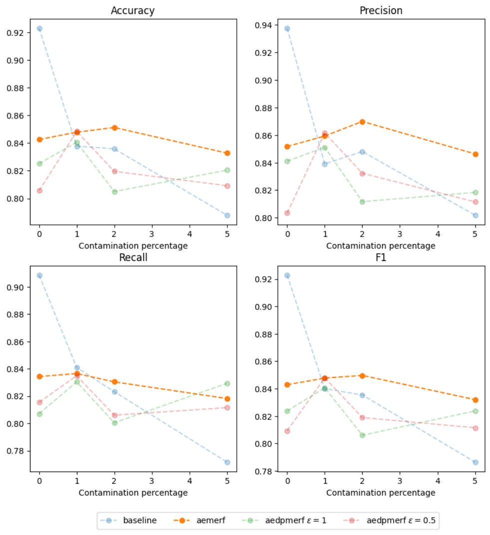

Figure: Contaminated training set: AE-(DP)MERF

>>> Contamination: AE-(DP)MERF

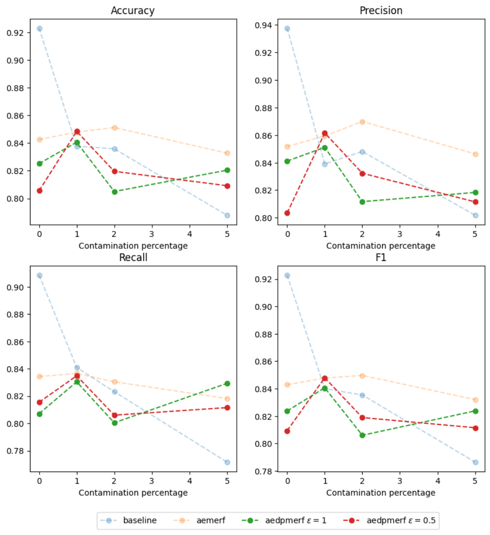

Figure: Contaminated training set: AE-(DP)MERF

>>> Contamination: AE-(DP)MERF

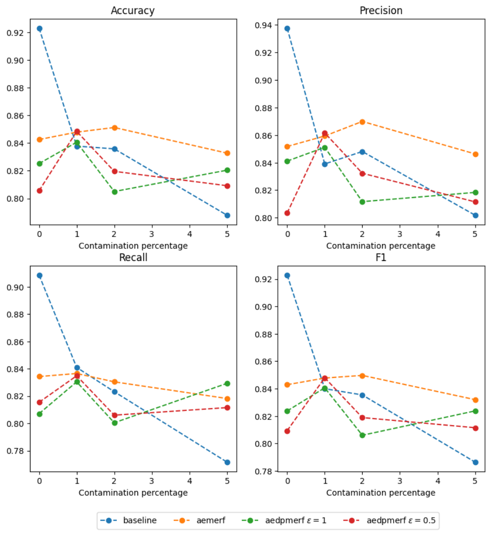

Figure: Contaminated training set: AE-(DP)MERF

>>> Outline

>>> Summary

tba

>>> Privacy-Preserving Acknowledgement

Thank you Aflsono, Apslotsuo, Hna, Sihahd!

>>> Outline

1. Project introduction
2. Heartbeat Arrhythmia
3. Privacy-preserving Time Series Data Generation
4. Results
5. Summary
6. References

Ang, Yihao et al. (2023). TSGBench: Time Series Generation Benchmark. arXiv: 2309.03755 [cs.LG].

Giomi, Matteo et al. (2022). A Unified Framework for Quantifying Privacy Risk in Synthetic Data. arXiv: 2211.10459 [cs.CR].

Greenwade, George D. (1993). "The Comprehensive Tex Archive Network (CTAN)". In: TUGBoat 14.3, pp. 342–351.

Hu, Yuzheng et al. (2023). SoK: Privacy-Preserving Data Synthesis. arXiv: 2307.02106 [cs.CR].

Kotelevskii, Nikita et al. (2022). "FedPop: A Bayesian Approach for Personalised Federated Learning". In: Advances in Neural Information Processing Systems. Ed. by S. Koyejo et al. Vol. 35. Curran Associates, Inc., pp. 8687–8701. URL: https://proceedings.neurips.cc/paper_files/paper/2022/file/395409679270591fd2a70abc694cf5a1-Paper-Conference.pdf.

Lin, Zinan et al. (2020). “Using GANs for Sharing Networked Time Series Data”. In: Proceedings of the ACM Internet Measurement Conference. ACM. DOI: 10.1145/3419394.3423643. URL: https://doi.org/10.1145%2F3419394.3423643.

Pei, Hengzhi et al. (2021). “Towards Generating Real-World Time Series Data”. In: Proceedings of the 2021 IEEE International Conference on Data Mining (ICDM), Auckla

Prediger, Lukas et al. (2023). Collaborative Learning From Distributed Data With Differentially Private Synthetic Tw arXiv: 2308.04755 [cs.LG].

Shokri, Reza et al. (2017). Membership Inference Attacks against Machine Learning Models. arXiv: 1610.05820 [cs.CR].

Stadler, Theresa, Bristena Oprisanu, and Carmela Troncoso (2020). “Synthetic Data - A Privacy Mirage”. In: CoRR abs/2011.07018. arXiv: 2011.07018. URL: https://arxiv.org/abs/2011.07018.

Tan, Alysa Ziying et al. (2022). “Towards Personalized Federated Learning”. In: IEEE Transactions on Neural Networks and Learning Systems, pp. 1–17. DOI: 10.1109/TNNLS.2022.3160699.

Yoon, Jinsung, Daniel Jarrett, and Mihaela van der Schaar (2019). “Time-series Generative Adversarial Networks”. In: Advances in Neural Information Processing Systems. Ed. by H. Wallach et al. Vol. 32. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper_files/paper/2019/file/c9efe5f26cd17ba6216bbe2a7d26d490-Paper.pdf.

>>> BACKUP

>>> Model Architecture