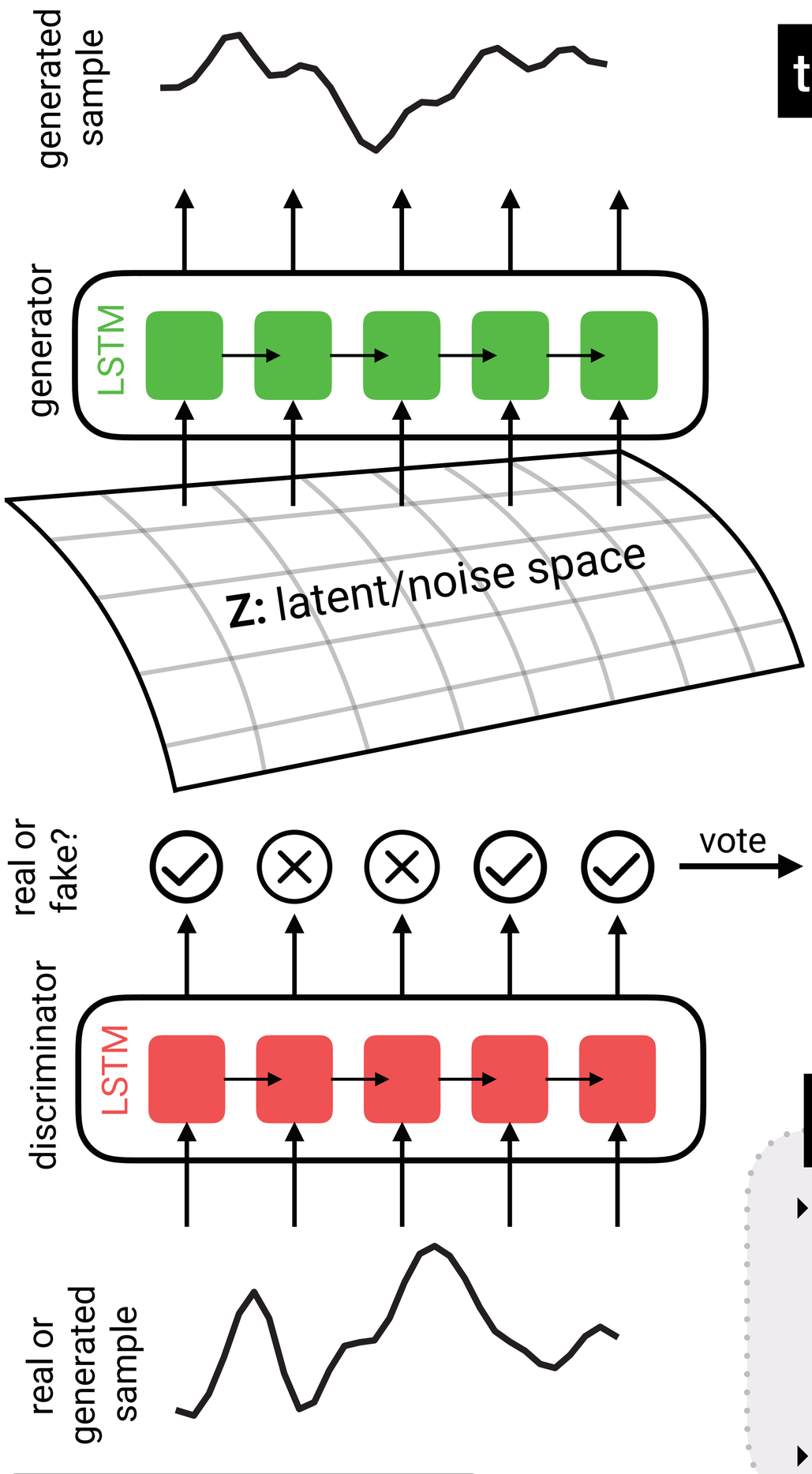
Real-valued (Medical) Time Series Generation with Recurrent Conditional GANs

Stephanie L. Hyland*, Cristóbal Esteban*, Gunnar Rätsch

*contributed equally

ETH zürich
BIOMEDICAL INFORMATICS
Weill Cornell Medicine



this work in a nutshell

- Generative adversarial network using recurrent neural networks (LSTMs)
- Data: **eICU** Collaborative Research Database
- Evaluation: **TSTR** method based on transfer learning
  - Use synthetic data to train classifier
- Privacy: empirical analysis, training GAN with **differential privacy**

the model (RGAN)

- **Discriminator** (LSTM) performs binary classification: real v. synthetic sample
- **Generator** (LSTM) tricks discriminator by generating realistic samples
- RCGAN: include **conditional** information (e.g. label of sequence) to both networks
  - Can generate examples from labels

data and tasks

- **eICU** Collaborative Research Database *(via PhysioNet, Goldberger et al., 2000)*
  - Vitals: MAP, heart rate (HR), SpO2, respiratory rate (RR)
  - Measurements every 15min for first 4 hours
  - Filter missing data: have **17,693 patients**
- Predict vitals becoming abnormal in next hour (see table)

evaluation (TSTR)

- Visual evaluation doesn't work for time-series
- Evaluate sample quality via **TSTR** *(train on synthetic, test on real):*
  - use **synthetic dataset** to train a model
  - test it on **real data**

Epoch was chosen based on validation set performance for SpO2 < 95, HR > 100, RR < 13

		SpO2 < 95	HR < 70	*HR > 100*
AUROC	real	0.9587 ± 0.0004	0.9908 ± 0.0005	0.9919 ± 0.0002
	TSTR	0.88 ± 0.01	0.96 ± 0.01	0.95 ± 0.01
AUPRC	real	0.9059 ± 0.0005	0.9855 ± 0.0002	0.9778 ± 0.0002
	TSTR	0.66 ± 0.02	0.90 ± 0.02	0.84 ± 0.03
	random	0.16	0.26	0.18

		RR < 13	RR > 20	MAP < 70	MAP > 110
AUROC	real	0.9735 ± 0.0001	0.963 ± 0.001	0.9717 ± 0.0001	0.960 ± 0.001
	TSTR	0.86 ± 0.01	0.84 ± 0.02	0.875 ± 0.007	0.87 ± 0.04
AUPRC	real	0.9557 ± 0.0002	0.891 ± 0.001	0.9653 ± 0.0001	0.8629 ± 0.0007
	TSTR	0.73 ± 0.02	0.50 ± 0.06	0.82 ± 0.02	0.42 ± 0.07
	random	0.26	0.1	0.39	0.05

why?

… generate medical data?

- Enable data sharing (if done carefully!)
- Dataset augmentation
- Simulation + training

privacy

- Q: is the GAN **overfitting** to the sensitive training data?
  1. Does synthetic data look more similar to training data than to test *data*? MMD 3-sample test *(Bounliphone et al., 2015)*
  2. Do reconstruction errors on training set look different to test set? Kolmogorov-Smirnov test between error distributions
- Use **differentially private** SGD *(Abadi et al., 2016)* to train discriminator: degrade TSTR performance, gain (ε, δ)-privacy

code and **paper**:

github.com/ratschlab/RGAN/

✨ **We're hiring!** ✨

https://bmi.inf.ethz.ch