May 1-5, 2023 - Kigali Convention Centre

ICLR International Conference On Learning Representations

BANOOK GROUP

UNIVERSITÉ DE LORRAINE

IADI

Inserm La science pour la santé From science to health

Post-hoc Uncertainty Quantification for QT Interval Measurements with Ensembles of Electrocardiographic Leads and Deep Models

Mously D. Diaw[1,2], Stéphane Papelier[1], Alexandre Durand-Salmon[1], Jacques Felblinger[2], Julien Oster[2]

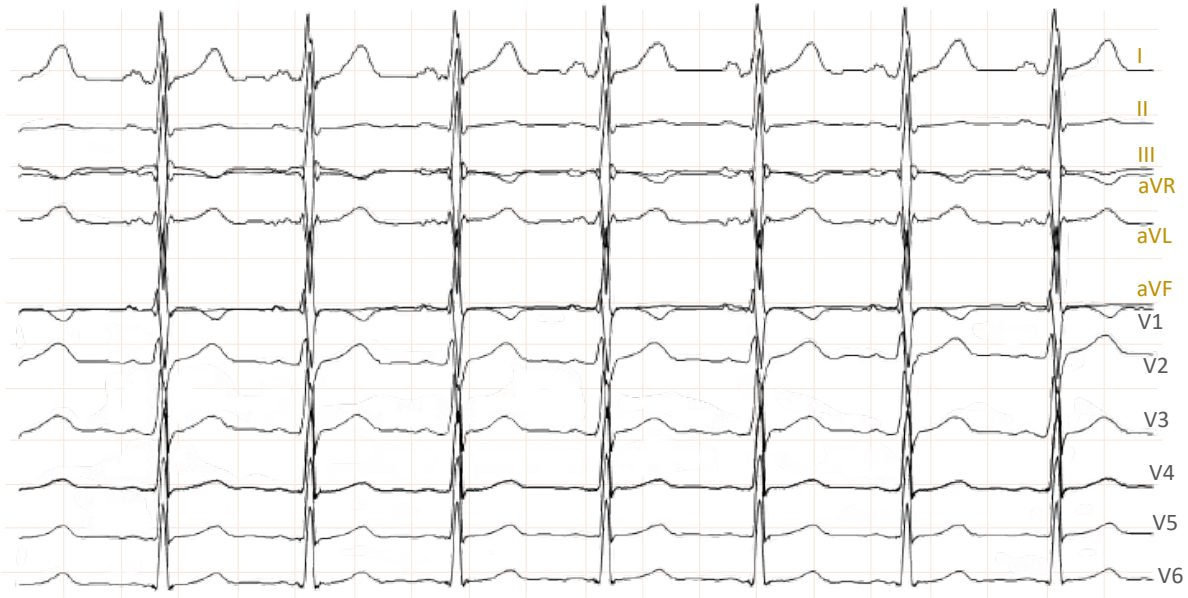
[1]Cardiabase, Banook Group [2]IADI, U1254, Inserm, Université de Lorraine, Nancy, France

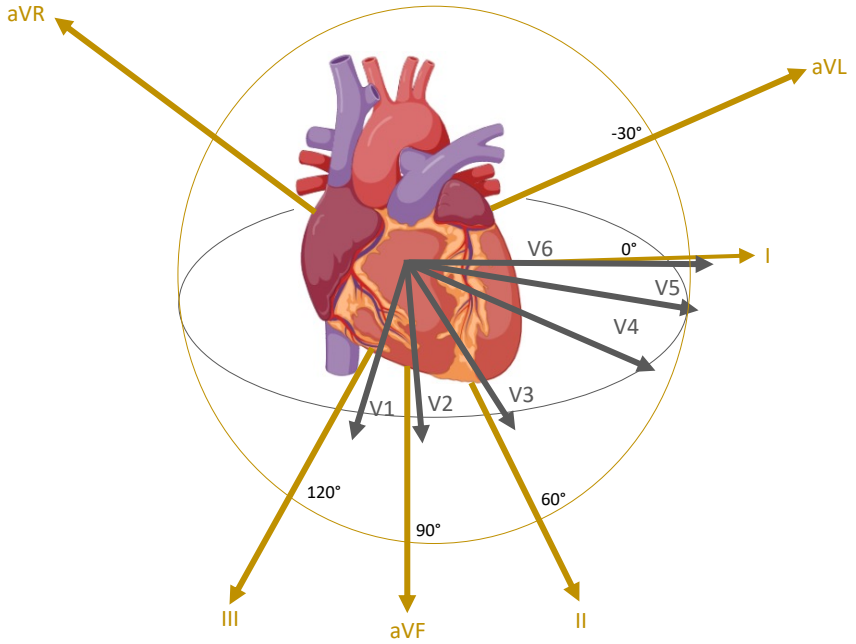
Workshop on Time Series Representation Learning for Health

May 5th, 2023

Electrocardiogram (ECG)

- Electrical activity of the heart
- 12 standard **leads**: same activity sensed from a different spatial viewpoint

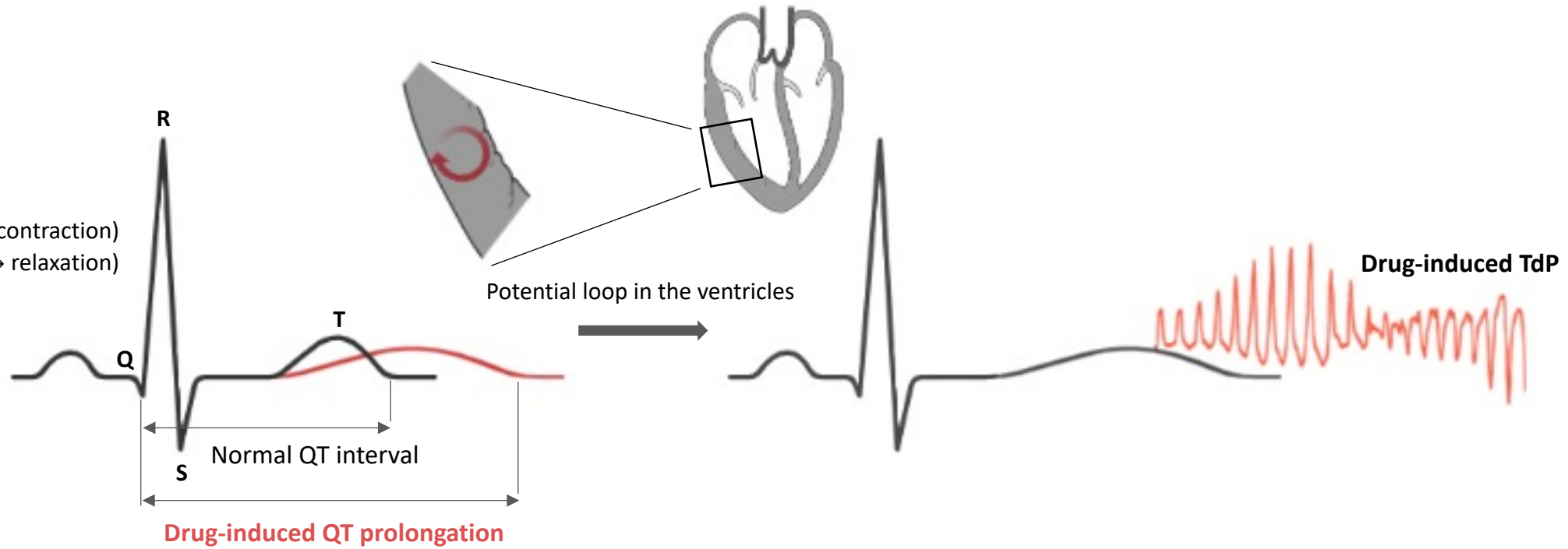




QT prolongation: ECG predictor of TdP risk

- Drug-induced torsades de pointes (TdP), a life-threatening arrhythmia

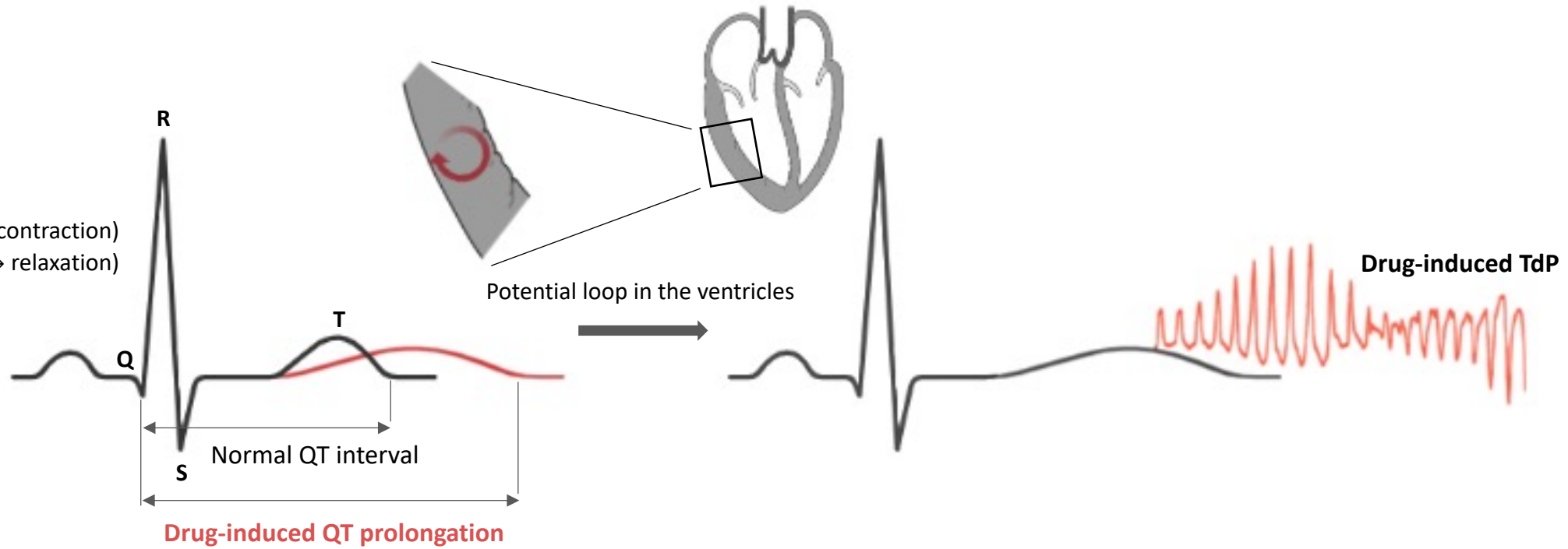
QRS: Ventricular depolarization (→ contraction)
T: Ventricular repolarization (→ relaxation)



QT prolongation: ECG predictor of TdP risk

- Drug-induced torsades de pointes (TdP), a life-threatening arrhythmia

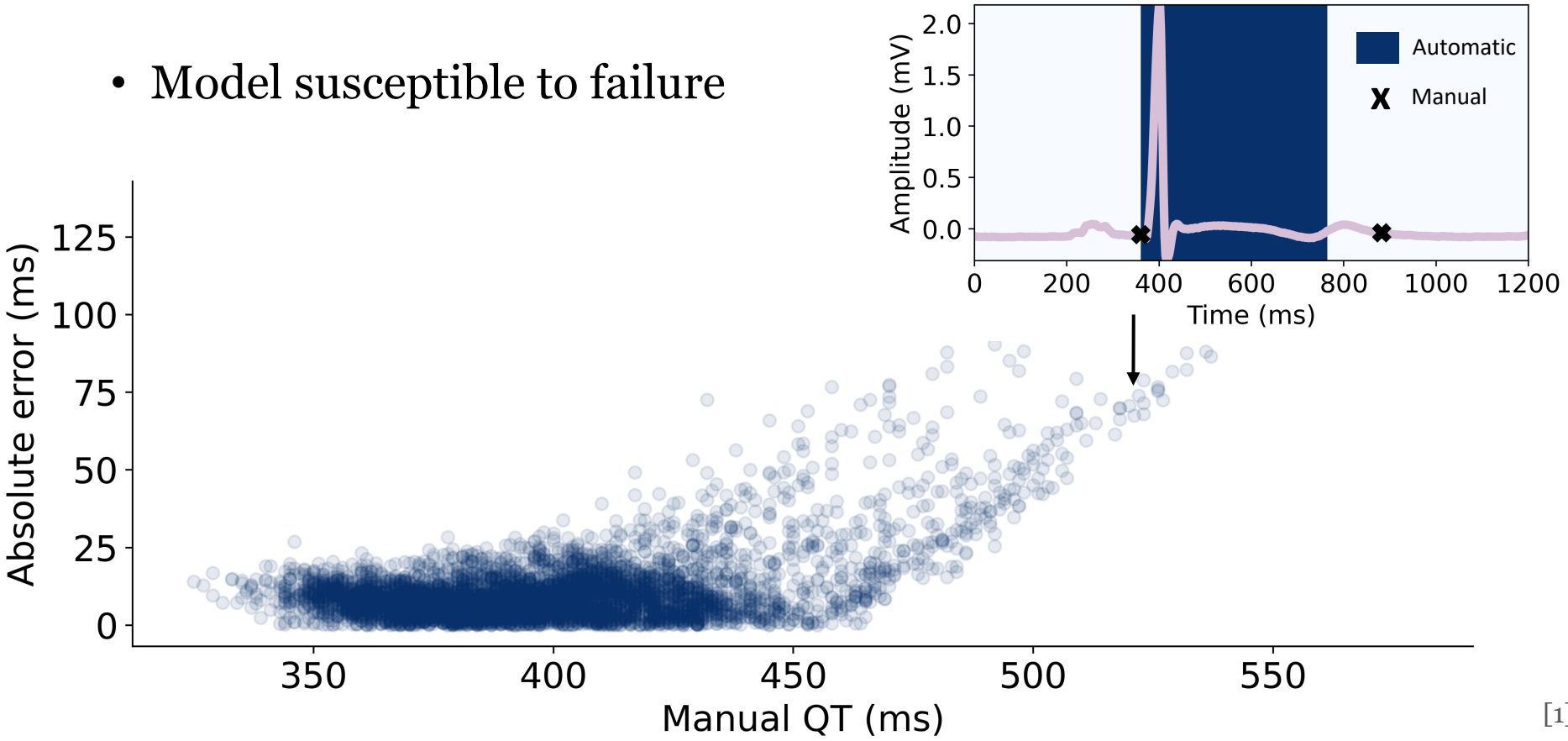
QRS: Ventricular depolarization (→ contraction)
T: Ventricular repolarization (→ relaxation)

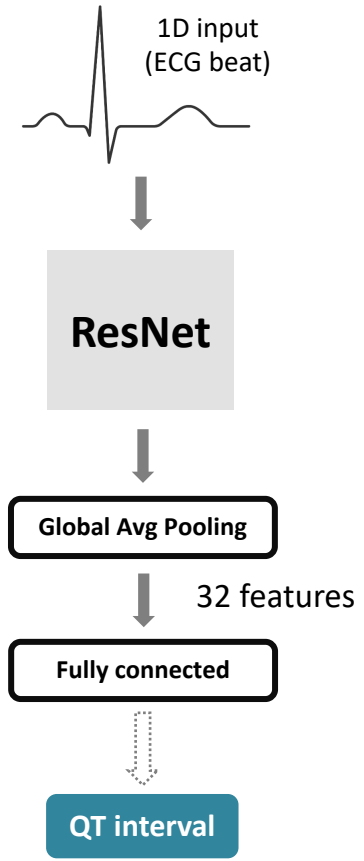


- QT monitoring guidelines to anticipate TdP risk (American Heart Association)

Automated QT interval measurement

- Deep learning approach
- Model susceptible to failure





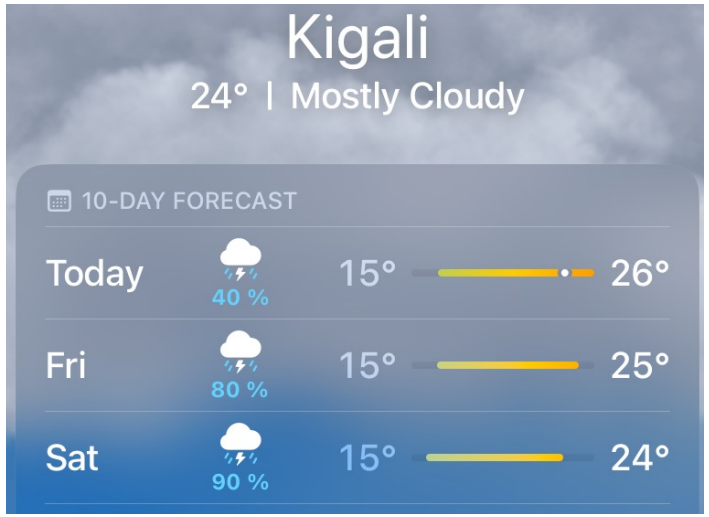
[1] Diaw, et al., *IEEE TBME*, 2022

Uncertainty quantification (UQ)

- Convey model reliability
- Regression: prediction interval (PI), e.g., $\mathbb{P}\{Y \in C(X)\} \approx 90\%$

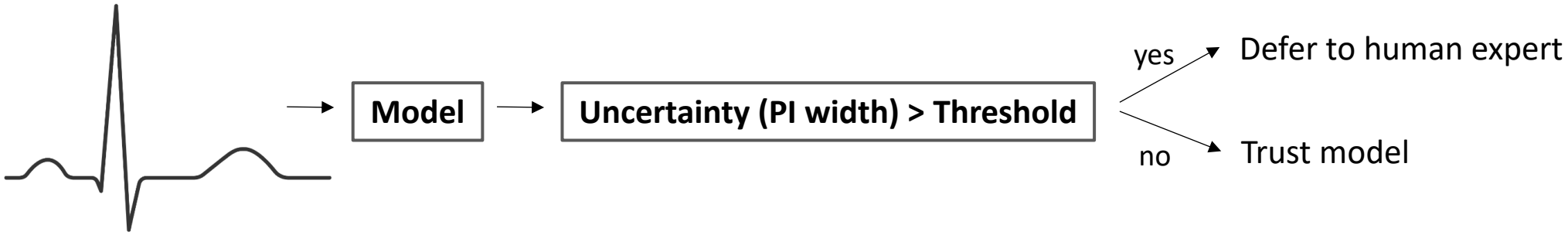
Uncertainty quantification (UQ)

- Convey model reliability
- Regression: prediction interval (PI), e.g., $\mathbb{P}\{Y \in C(X)\} \approx 90\%$



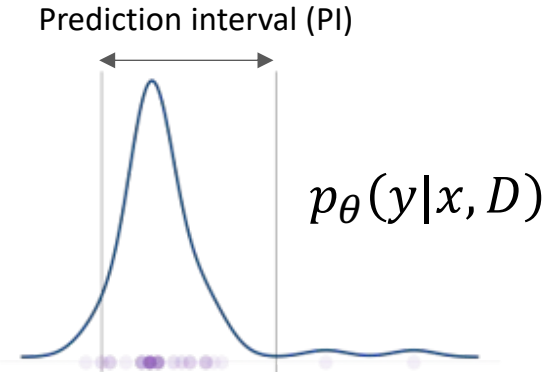
Uncertainty quantification (UQ)

- Convey model reliability
- Regression: prediction interval (PI), e.g., $\mathbb{P}\{Y \in C(X)\} \approx 90\%$
- Human-in-the-loop QT measurement



UQ in deep learning

- Approximate the posterior predictive distribution
- Focus on model diversity: posterior over model weights $p(\theta|D)$



Bayesian Neural Networks

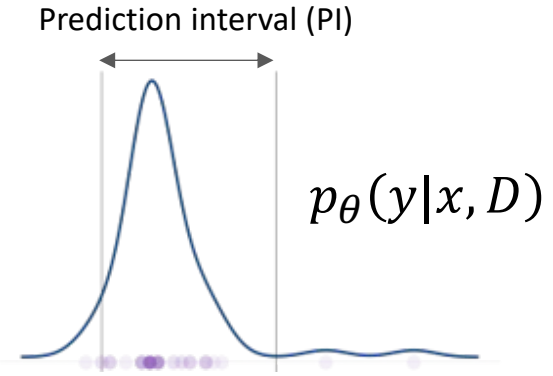
Monte Carlo Dropout[2]

Deep ensembles[3]

[2] Gal, et al., *ICML*, 2016
[3] Lakshminarayanan, et al., *NeurIPS*, 2017

UQ in deep learning

- Approximate the posterior predictive distribution
- Focus on model diversity: posterior over model weights $p(\theta|D)$



Bayesian Neural Networks

Monte Carlo Dropout[2]

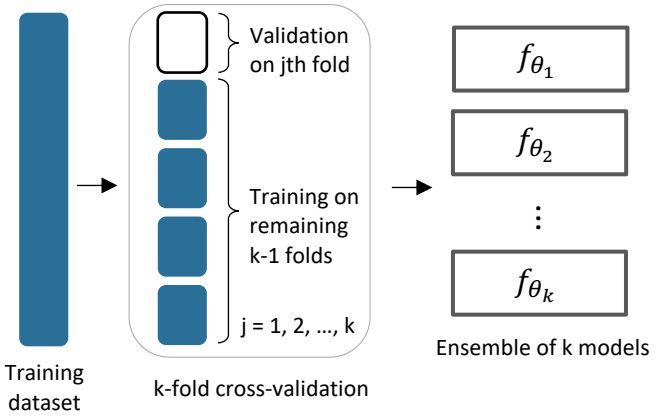
Deep ensembles[3]

[2] Gal, et al., *ICML*, 2016
[3] Lakshminarayanan, et al., *NeurIPS*, 2017

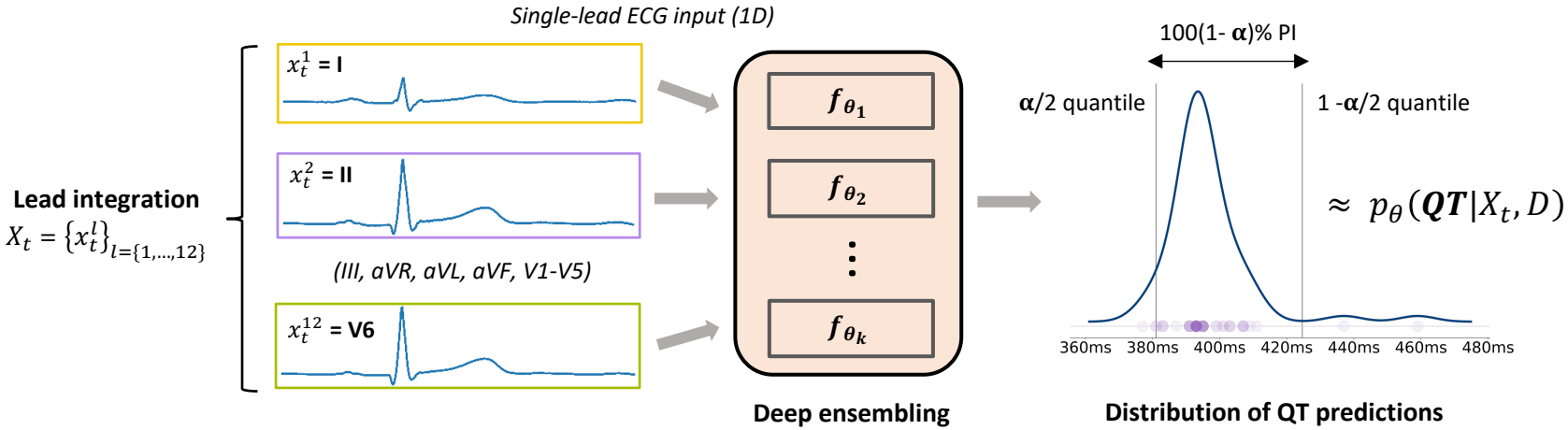
Can we leverage the **diversity in multi-lead ECGs** $X_t = \{x_t^l\}_{l=\{1,\dots,L\}}$ to approximate $p_\theta(QT|X_t, D)$?

Proposed post-hoc UQ method

Training of a cross-validation ensemble

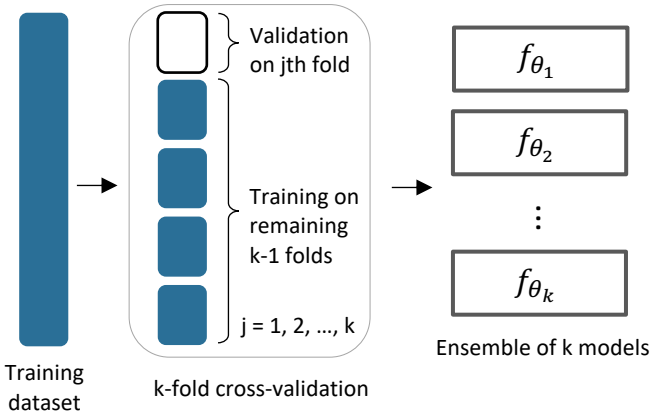


Approximate Bayesian inference

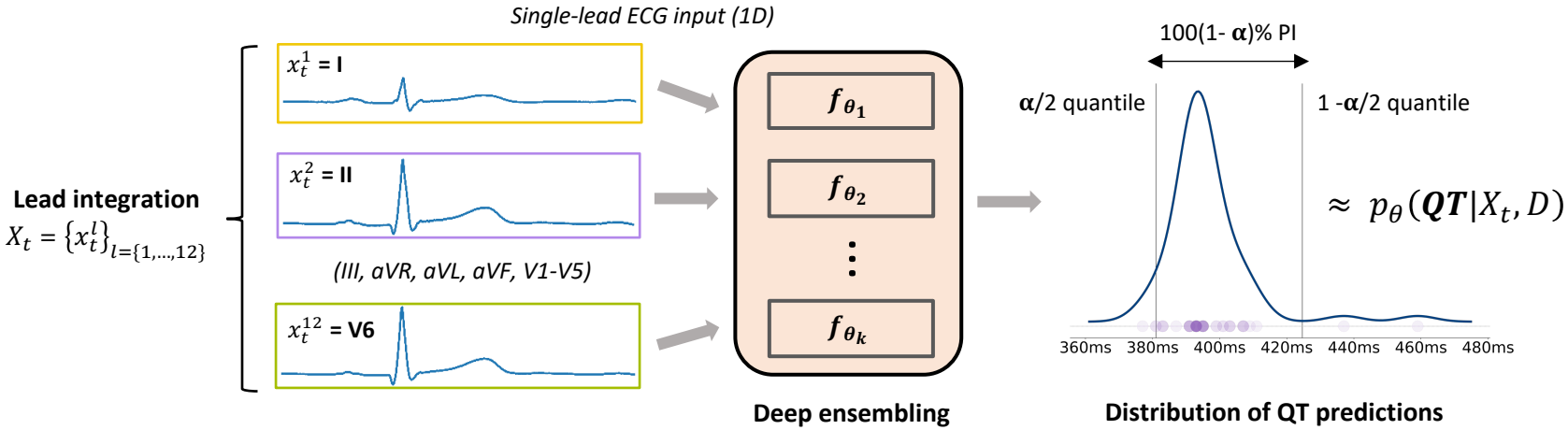


Proposed post-hoc UQ method

Training of a cross-validation ensemble



Approximate Bayesian inference



- **UQ-ELM** (UQ using an Ensemble of Leads and Models): $\boldsymbol{k} \times \boldsymbol{L}$ estimates $f_{\theta_k}(x_t^l)$
- **UQ-EL** (UQ using an Ensemble of Leads): $\boldsymbol{L}$ estimates $f_\theta(x_t^l)$, with $f_\theta = \frac{1}{k}\sum_{j=1}^{k} f_{\theta_k}$

Locally Adaptative Split Conformal Prediction (LASCP)

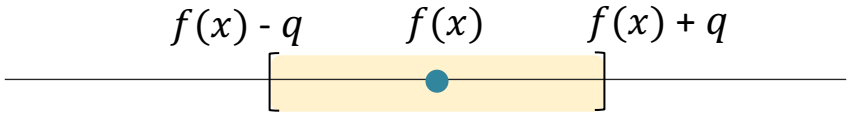
- Given n past observations $(X_1, Y_1), \dots, (X_n, Y_n)$ construct 90% PI for $(X_{n+1}, ?)$

[4] Vovk, et al., *Algorithmic learning in a random world*, 2005
[5] Papadopoulos, et al., *AIA*, 2008
[6] Papadopoulos, et al., *JAIR*, 2011
[7] Lei, et al., *JASA*, 2018

Locally Adaptative Split Conformal Prediction (LASCP)

- Given n past observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ construct 90% PI for $(X_{n+1}, ?)$

- SCP consists in splitting the n samples into:
  - D_{train} to fit predictive model $f(x)$
  - D_{calib} to compute q, the 90th percentile of the residual scores $|y - f(x)|$

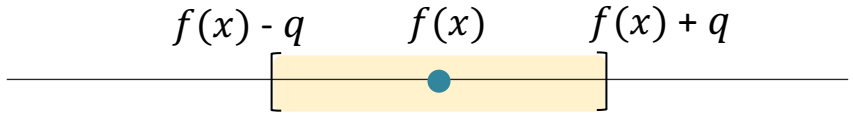


[4] Vovk, et al., *Algorithmic learning in a random world*, 2005
[5] Papadopoulos, et al., *AIA*, 2008
[6] Papadopoulos, et al., *JAIR*, 2011
[7] Lei, et al., *JASA*, 2018

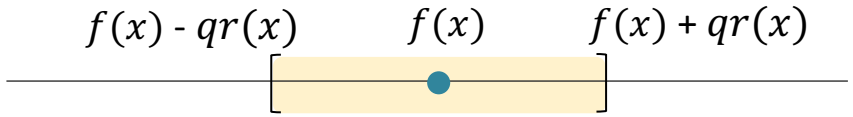
Locally Adaptative Split Conformal Prediction (LASCP)

- Given n past observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ construct 90% PI for $(X_{n+1}, ?)$

- SCP consists in splitting the n samples into:
  - D_{train} to fit predictive model $f(x)$
  - D_{calib} to compute q, the 90th percentile of the residual scores $|y - f(x)|$



- With LASCP, construct adaptative PIs by further splitting D_{calib} into:
  - D_1 to fit residual (error) predictor $r(x)$
  - D_2 to compute q, the 90th percentile of the new non-conformity scores $\frac{|y - f(x)|}{r(x)}$



[4] Vovk, et al., *Algorithmic learning in a random world*, 2005
[5] Papadopoulos, et al., *AIA*, 2008
[6] Papadopoulos, et al., *JAIR*, 2011
[7] Lei, et al., *JASA*, 2018

Experimental setup

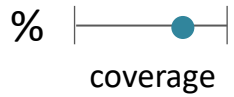
- **12-lead** ECGs from 2 drug safety clinical studies[8,9] (available at PhysioNet.org)
- Patient-stratified **5-fold** cross-validation of ResNet
- LASCP residual predictor r(x) - r: Gradient boosting regressor
  x: ResNet-extracted ECG features

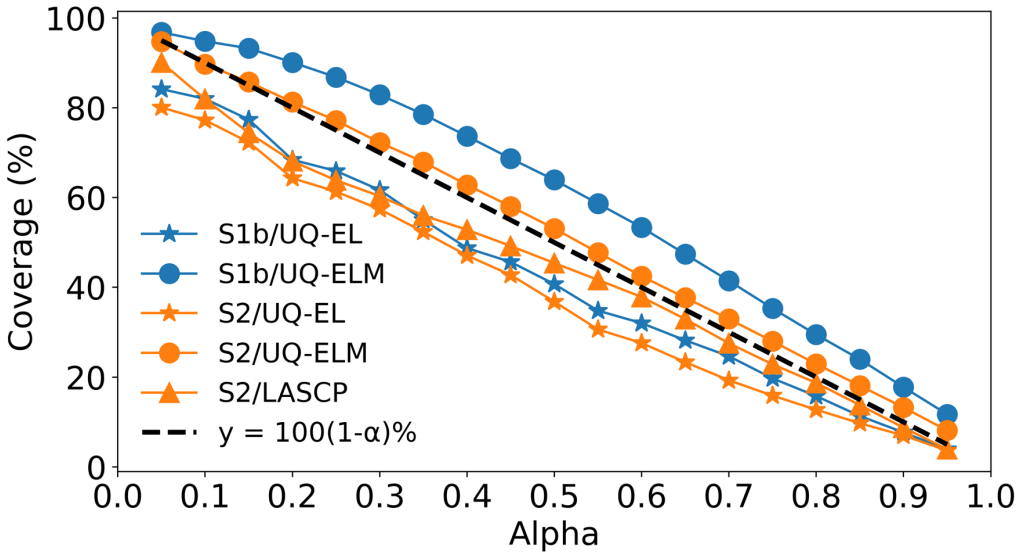
Dataset	Purpose	Subjects/12-lead ECGs
S1a	Cross-validation	22/2056
S1b	LASCP (and evaluation of UQ-ELM/EL)	14/2014 (D_1) 8/1149 (D_2)
S2	Evaluation (UQ-ELM/EL, LASCP)	22/4211

[8] Johannesen, et al., *Clin Pharmacol Ther*, 2014
[9] Johannesen, et al., *Clin Pharmacol Ther*, 2016

Good coverage-based calibration

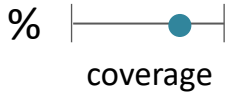
Coverage ideally $\approx 100(1-\alpha)\%$

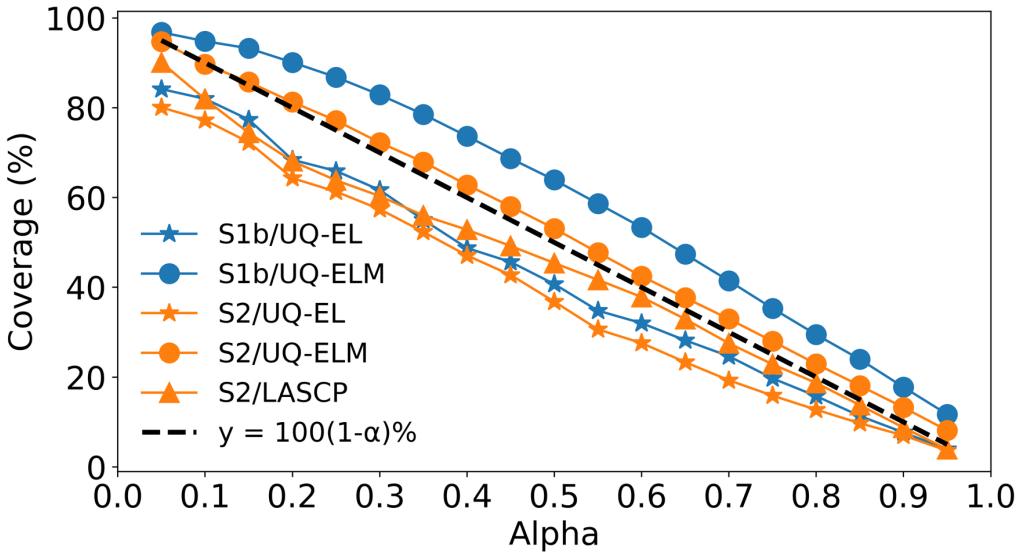




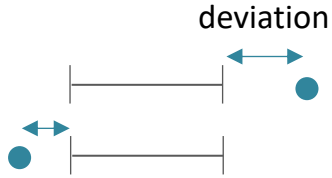
Good coverage-based calibration

Coverage ideally $\approx 100(1-\alpha)\%$





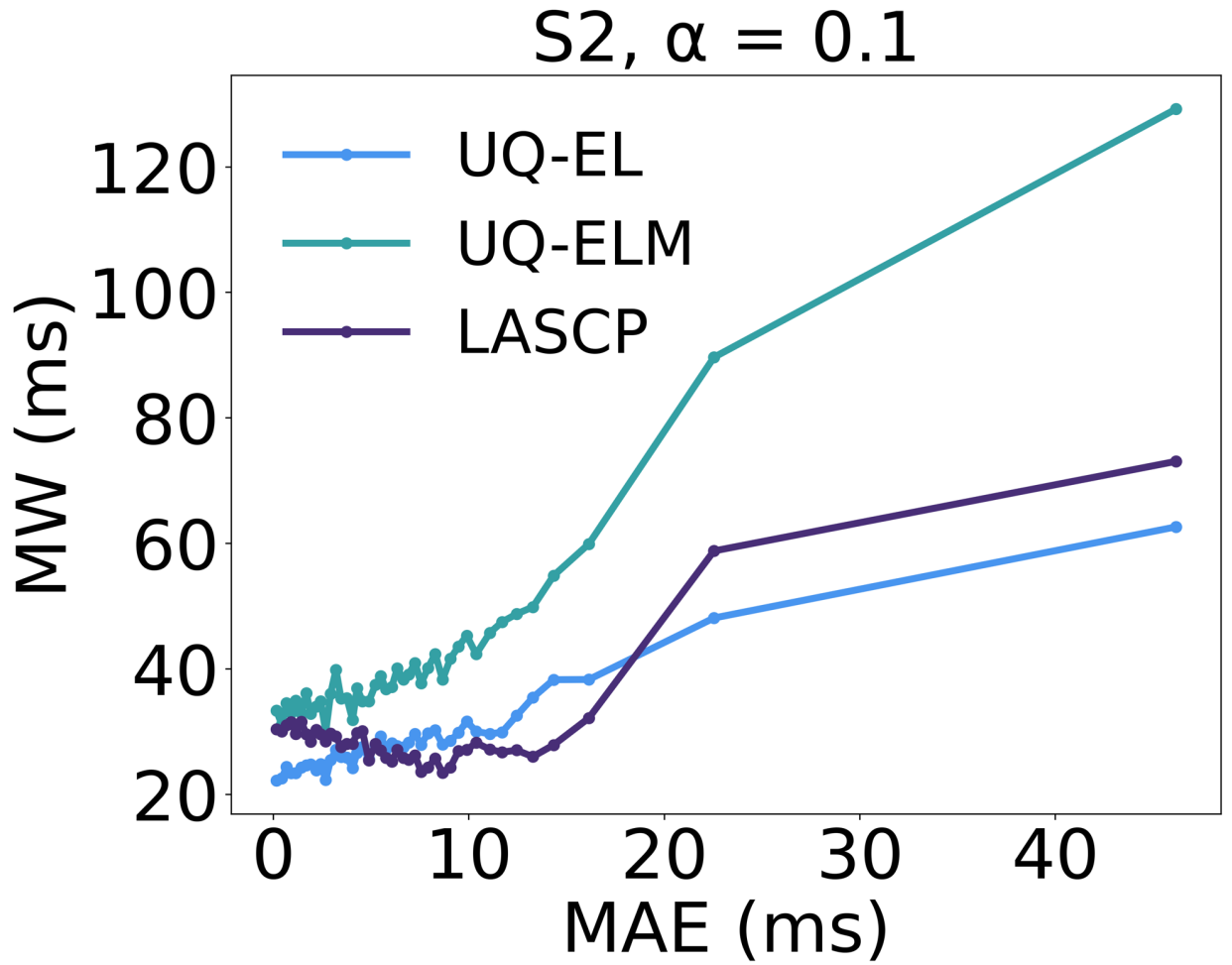
MW: Mean PI Width
MAD: Mean Absolute Deviation



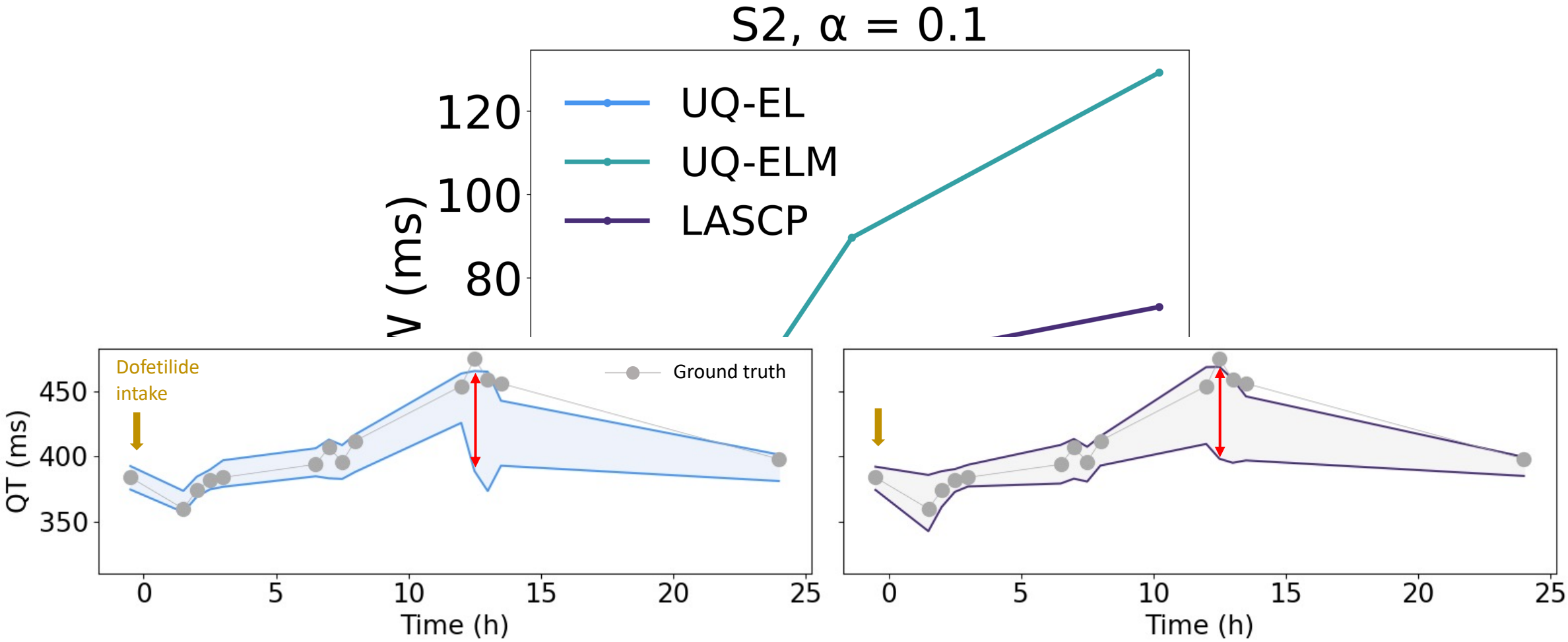
90% PIs ($\alpha = 0.1$)

Dataset	Method	Coverage	MW (ms)	MAD (ms)
S1b	UQ-ELM	95%	43.91	9.81
	UQ-EL	82%	29.49	7.20
S2	UQ-ELM	90%	40.47	3.26
	UQ-EL	77%	28.27	3.95
	LASCP	82%	28.67	3.50

The higher the model error, the higher the uncertainty



The higher the model error, the higher the uncertainty



Future work

- Other ensembling techniques
- UQ for improved predictive performance

Future work

- Other ensembling techniques
- UQ for improved predictive performance

Take-away

By leveraging the inherent diversity in health time series, we could build simple and reliable UQ tools

Thanks !

Contact
mously.diaw@univ-lorraine.fr