UMO
Universitätsmedizin
Oldenburg

D1: Machine-learning for photoplethysmography analysis: Benchmarking feature, image, and signal-based approaches

QUMPHY Stakeholder Meeting 2026

Prof. Dr. Nils Strodthoff
Carl von Ossietzky Universität Oldenburg
Germany

KLINIKUM OLDENBURG
EVANGELISCHES KRANKENHAUS OLDENBURG
Menschen im Mittelpunkt
Psychiatrieverbund Oldenburger Land gGmbH
KARL-JASPERS-KLINIK
pius HOSPITAL OLDENBURG

Biomedical Signal Processing and Control 120 (2026) 109831

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Machine-learning for photoplethysmography analysis: Benchmarking feature, image, and signal-based approaches

Mohammad Moulaeifard [a], Loic Coquelin [b], Mantas Rinkevičius [c], Andrius Sološenko [c], Oskar Pfeffer [d], Ciaran Bench [e], Nando Hegemann [d], Sara Vardanega [f], Manasi Nandi [f], Jordi Alastruey [f], Christian Heiss [g], Vaidotas Marozas [c], Andrew Thompson [e], Philip J. Aston [e,h], Peter H. Charlton [i], Nils Strodthoff [a,*]

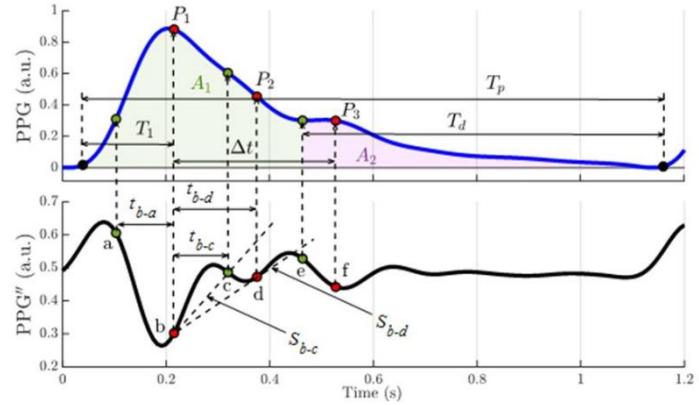
Role in the project

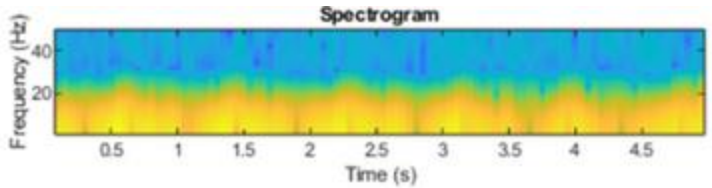
- **Part of work package 1 (model development and uncertainty quantification)**

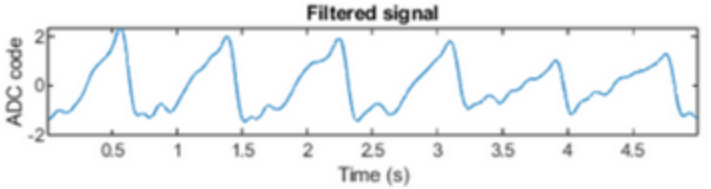
- Three deliverables
  - **D1: Report on machine learning models for classification and regression problems using PPG signals and a comparison of their performance**
  - D2: Report on uncertainty quantification for machine learning classification and regression problems using PPG signals and a comparison of their performance
  - D3: Report on quantitative assessment and comparison of model accuracy and uncertainty estimates

Motivation & Research Question

- **Models can be trained on different input representations**
  - Clinically interpretable features
  - Image representations
    e.g. spectrograms, wavelets, SPAR
    processed by convolutional neural networks
  - raw waveforms
    e.g. convolutional neural networks, transformers, …







- **How do models operating on different input representations compare performance-wise?**

Image credit: Barki et al

Datasets and Tasks

- **Two prototypical datasets (classification & regression)**

- **DeepBeat Atrial fibrillation detection (Torres-Soto & Ashley)**
  - binary classification AF/no-AF

Subset	AF	Non-AF	Data Ratio	AF Ratio
Train (samples / subjects)	40603 / 50	65646 / 38	0.78	0.38
Validation (samples / subjects)	5800 / 19	9456 / 7	0.11	0.38
Test (samples / subjects)	5797 / 19	9580 / 5	0.11	0.37

- **PulseDB(Vital) Blood pressure estimation (Wang et al 2023)**
  - two regression targets (SBP/DBP)
  - „Calibfree“ (test on unseen patients)
  - „Calib“ (test on known patients)

Subset	VitalDB ‘Calib’	VitalDB ‘CalibFree’
Train (samples / subjects)	418986 / 1293	416880 / 1158
Validation (samples / subjects)	40673 / 1293	32400 / 90
Test (samples / subjects)	51720 / 1293	57600 / 144

Results (AF classification)

- best performance achieved by modern convolutional neural networks (not transformers)
- image-based and raw-timeseries methods on par, feature-based approaches inferior
- Remaining doubts about the label quality of the DeepBeat dataset: alternatives needed
- **Conclusion: Strong performance for PPG-based AF classification**

	Model	AUC	F1 (0.5)	Specificity (sensitivity > 0.8)
T	XResNet1d101	0.85	0.66	0.74
	XResNet1d50	**0.85**	0.69	0.75
	Inception1d	**0.85**	0.69	0.74
	LeNet1d	0.74	0.55	0.55
	AlexNet1d	0.82	0.65	0.68
	Minirocket	0.82	0.66	0.69
	iTransformer	0.68	0.40	0.47
	TimesNet	0.79	0.58	0.61
	PPNet	**0.85**	0.69	0.76
	TCN+MLP	**0.85**	0.67	0.75
F	WAVELET + MLP	0.76	0.60	0.58
	CIF+MLP	0.52	0.39	0.20
I	CWT	0.82	0.69	0.72
	Spectrogram-ResNet-18	0.82	0.68	0.68
	Spectrogram-ResNet-50	**0.85**	0.69	0.72

Results (BP estimation, calibfree)

- best performance achieved by raw-timeseries models (CNNs)
- MAEs are large, only slightly better than constant baseline prediction
- 64% in IEEE grade D, therefore not suitable for clinical applications
- Similar for calib:
  - Baseline MAE 9.4 mmHg
  - Best model (LeNet1d): MAE 7.9 mmHg
- **Conclusion: BP estimation from PPG alone remains challenging, in line with literature results**

	Model	SBP
		MAE (MASE)
B	Baseline	14.87 (1.00)
T	XResNet1d101	12.43 (0.83)
	XResNet1d50	12.46 (0.83)
	Inception1d	14.46 (0.97)
	LeNet1d	12.37 (0.83)
	XResNet1d50+GNLL	**12.27 (0.82)**
	AlexNet1d	12.51 (0.84)
	Minirocket	12.65 (0.85)
	iTransformer	13.15 (0.88)
	TimesNet	13.09 (0.88)
	PPNet	12.56 (0.84)
	TCN + MLP	12.72 (0.85)
F	WAVELET + MLP	14.21 (0.95)
	CIF + GPR	12.90 (0.87)
	CIF + MLP	14.02 (0.94)
I	CWT	13.49 (0.90)
	Spectrogram-ResNet-18	13.01 (0.87)
	Spectrogram-ResNet-50	13.53 (0.90)

Outlook

- benchmarking of model performance was just the first step
- D2 considers the same datasets/task but evaluates uncertainty quantification (c.f. talk of Vivek Desai)
- In D1, we only considered ID performance, out-of-distribution (OOD) performance is typically worse
  - investigated in follow-up work for performance (Moulaeifard et al 2025) and uncertainty estimation (Moulaeifard et al 2026)

Thank you for your attention!