

Third Stakeholder Workshop

**Report on selected Benchmark Problems and Datasets in
work package 2 and Deliverable D4**

Urs Hackstein (THM)

10 April 2026



WP1: Uncertainty quantification

Focus:

- Develop UQ methods for ML algorithms applied to classification/regression tasks involving PPG signals.
- Validate uncertainties of ML models on benchmark problems.

WP2: Data and Promotion

Focus:

- Generate at least 5 benchmark problems containing datasets of clinical interest and make them publicly available
- Engage with medical device, digital and healthcare communities
- Develop tools for end-users to employ UQ for ML applied to tasks involving PPG-data



This session: Benchmark Problems and Datasets

Criteria:

- Problems of interest for clinicians and hospitals, healthcare communities, digital healthcare companies, and metrological and scientific communities
- Suitable datasets are publicly or upon request available + suitable for Machine Learning and Uncertainty quantification, e.g. clear ground truth, less label noise, enough information on measurement protocol etc.

available at: www.qumphy.ptb.de and <https://arxiv.org/abs/2604.01398>

- ➔ Selected six Benchmark Problems together with Benchmark Datasets with respect to these criteria
- ➔ Publication of our selection in D4 report together with information on
 - relevance of the problem
 - proposed datasets
 - information how to evaluate these datasets
 - etc.

Software repository: <https://gitlab.com/qumphy/d4-code>



Selected Benchmark Problems

- 1. Blood pressure regression**
- 2. Detection of Atrial Fibrillation**
- 3. Classification of Blood pressure levels**
- 4. Determination of vascular age**
- 5. Detection of sleep apnoea**
- 6. Regression of respiratory rate**



1. Blood pressure regression

Determine systolic and diastolic blood pressure

Relevant:

- BP is one of the most commonly taken physiology measurements
- BP regression from PPG enables non-invasive, continuous and cuffless BP estimation

Datasets

- 1) **AuroraBP**: available on request from Microsoft: 1,125 participants, measurements in a lab over a 24 hour period, some subjects also had ambulatory measurements
- 2) **VitalDB**: subset of PulseDB; 2,938 non-cardiac surgery patients (=ICU) from South Korea, publicly available



2. Detection of Atrial fibrillation

Relevant:

- More than 33 million people worldwide are already diagnosed with AF
- AF detection can be challenging due to paroxysmal and sometimes brief episodes
- PPG could be an alternative to 12-lead ECG, maybe enables early detection in asymptomatic cases

Datasets

A) Wrist-based datasets:

- 1) **DeepBeat**: publicly available from Stanford University, restructured in Qumphy D1 report; advantage: large scale dataset (more than 47,000 AF samples and more than 84,000 non-AF samples) + existing code (**TRAINING**)
ATTENTION: SOME LABEL NOISE, not a common SOP for all subjects
- 2) **TriggersAF**: publicly available on zenodo (KTU + Vilnius University Hospital), 133 subjects (**TEST**)



2. Detection of Atrial fibrillation

Datasets

B) Finger-tip datasets:

- 1) **MIMIC-III-EXT-PPG**: NEW DATASET generated from MIMIC III Waveform Database by University of Oldenburg within Qumphy, available on Physionet (**TRAINING**)
- 2) **Liu2022 dataset**: not derived from MIMIC, 91 subjects (**external TEST set**)



3. Classification of blood pressure levels

Relevance:

- **Elevated blood pressure and Hypertension** have a very high prevalence (30% in Europe)
- One of the major risk factors in many diseases (stroke, AF, diabetes, heart failure etc.)

Dataset:

AuroraBP: cf. BP regression problem

Usage: see explanation in D4 report



4. Determination of Vascular Age

Relevance:

- Early detection of premature Vascular age is critical for timely identification and treatment of cardiovascular diseases
- Actual gold standard: carotid - femoral pulse wave velocity
- Comparison of signal-based estimation of vascular age to a person's chronological age

Datasets:

1) AuroraBP: cf. BP regression problem and BP classification problem

Usage: see explanation in D4 report

2) Pulse Wave Database: available on zenodo, simulated PPG-signals, perhaps idealised



5. Detection of Sleep apnoea

Relevance:

- Obstructive Sleep apnoea (OSA): prevalent sleep disorder: repeated episodes of partial or complete upper airway obstruction during sleep
- Associated with a range of comorbidities

Datasets:

- 1) **OSASUD (Obstructive Sleep Apnea Stroke Unit Dataset):** Udine University Hospital (Italy), 30 patients, open-access
- 2) **MESA:** available upon request, 2055 patients

Usage: see D4 report



6. Regression of respiratory rate

Relevance:

- RR gives much information of the clinical state
- Very sensitive vital sign marker that indicates clinical deterioration

Datasets:

- 1) **OSASUD:** cf. Sleep apnoea detection
- 2) **MIMIC III-Ext-PPG:** cf. AF classification
- 3) **MIMIC Perform Large:** **Mimic Perform** dataset was enlarged during Qumphy, more than 6500 recordings from more than 800 subjects

Usage: see D4 report



www.qumphy.ptb.de



gitlab.com/qumphy



qumphy-software.rtfid.io

