European Partnership on Metrology
Call 2025 – Health, Integrated European Metrology, Regulation and Research Potential

Selected Research Topic number: **SRT-h11**
Version: 1.0

Title: Combining machine learning and physiological modelling to generate cardiovascular reference data

Abstract

The two most clinically relevant medical time series for the cardiovascular system are non-invasive Photoplethysmograms (PPG) and Electrocardiograms (ECG), which can infer a range of physiological parameters or detect certain heart diseases. However, for PPG the signal quality can be poor, and the outputs can be biased towards certain skin complexions. Medical data in general is laborious and sometimes expensive to collect and is also subject to several ethical and privacy concerns as even anonymised data is considered personal data under the EU's General Data Protection Regulation. Artificially generated synthetic data is an alternative but most simulated medical signals are too distinctive from real measurements to be used to train or evaluate modern machine learning algorithms. Proposals are sought to advance metrological standards in healthcare by providing high-fidelity reference data for ECG and PPG cardiovascular signals, improving physiological models, and developing machine learning techniques to generate synthetic data that is indistinguishable from patient measurements. This approach offers access to large amounts of unbiased, high-quality data at significantly reduced costs, free from privacy concerns. The reference data produced will provide a robust foundation for the development and certification of European machine learning models, enhancing the reliability and trustworthiness of AI-driven healthcare applications.

Keywords

synthetic reference data, machine learning, physiological modelling, cardiovascular system, ECG, PPG, quality control, healthcare,

Background to the Metrological Challenges

The cardiovascular system's two clinically most relevant medical time series are Photoplethysmograms (PPG) and Electrocardiograms (ECG) which contain an abundance of hidden information about the cardiovascular system and the condition of the arterial network. These are often used to infer physiological parameters such as blood pressure, arterial stiffness, and heart rate or to detect certain diseases such as atrial fibrillation or stenosis in a non-invasive manner. The ability to process these datasets using machine learning (ML) approaches will accelerate diagnosis and monitoring and improve patient outcomes.

For PPG depending on the measurement device, the signal quality might be comparatively poor and biased towards certain skin complexions. However, for both PPG and ECG systematic distortions in data can result from a multitude of factors, which makes correcting for these biases challenging in clinical settings. For this reason, the large amounts of high-quality data that are required for machine learning (ML) applications are difficult to obtain.

Moreover, managing these data in accordance with European privacy and data security standards poses another challenge because even anonymised data are considered personal data under the EU's General Data Protection Regulation [1]. Similarly, the uptake of AI in the medical sector is limited by the EU AI Act's [2] requirement of high-quality data to "promote the uptake of human-centric and trustworthy artificial intelligence."

Artificially generated synthetic data is an alternative, but most simulated medical signals are currently too distinctive from real measurements to be used to train or evaluate modern ML algorithms. One way to overcome this problem is to use two primary strategies to enhance PPG and ECG cardiovascular simulations to improve their usability for ML. The first approach focuses on enhancing the underlying physiological models and refining existing simulation software. Among the most promising features to be added are the integration

EURAMET MSU,
Email: part.msu@euramet.org
Metpart.eu

European Partnership Co-funded by the European Union
METROLOGY PARTNERSHIP EURAMET

of (i) human behaviour, e.g., respiration or movement, (ii) measurement device influences, e.g., baseline wander or varying signal quality, (iii) physiological or medical dispositions, e.g., gender, age or skin tone, and (iv) diseases, e.g., stenosis, arrhythmia or sleep apnoea. The second strategy employs ML techniques to refine existing data, generating more realistic simulations. Some of the most promising approaches include (i) generative modelling, i.e., to transform existing or generate new data that are hard to distinguish from real measurements, (ii) feature selection, i.e., to extract relevant features from the time series to facilitate generative modelling, and (iii) optimal experimental design, i.e., to identify data points with maximum relevance for training.

Synthetic ECG and PPG data sets suitable for ML represent a significant advancement in European metrology as they offer easy access to standardised data sources for research and industry without privacy concerns and can be used for testing, certification, and calibration. In standard multi-stage testing, an algorithm's performance is typically evaluated using data sets of varying difficulty, such as clean simulated data, noisy simulated data, and real-world data. At the moment there is a lack of access to the first two types of reference data for ECG and PPG signals. Furthermore, as generating synthetic data is comparatively inexpensive and does not require human intervention, the use of synthetic data means it becomes virtually impossible to cheat tests by tuning algorithms to specific test data. A number of recent and current European projects have addressed related topics but have not led to large-scale synthetic datasets. The QUMPHY project could not incorporate simulated PPG signals due to the lack of suitable available synthetic data sets, the MedalCare project demonstrated the challenges in combining simulated and real ECG signals to improve disease detection, and VascAgeNet recently published a roadmap article highlighting the importance of synthetic data for assessing vascular properties.

Objectives

Proposers should address the objectives stated below, which are based on the PRT submissions. Proposers may identify amendments to the objectives or choose to address a subset of them in order to maximise the overall impact, or address budgetary or scientific / technical constraints, but the reasons for this should be clearly stated in the protocol.

The proposal shall focus on enhancing metrological standards in healthcare by providing synthetic reference data for machine learning applications.

The specific objectives are:

1. To gather state-of-the-art models and measurements to support the selection of clinical applications. This will include at least 1 clinically relevant application involving Electrocardiogram (ECG) and at least 1 clinically relevant application involving Photoplethysmogram (PPG) signals. The models may include long-term signals with arrhythmias, obstructive and central sleep apnoea, hypertension disorder, stenosis or aneurysms, and will be compared to real measurements to assess their usability.
2. To advance physiological models by improving at least 2 aspects of each of the 2 models for clinical applications from objective 1. This may include improved modelling of the physical process (e.g., breathing artefacts, movement, minority groups), the introduction of typical device biases (e.g., baseline wander) or modelling of medical conditions (e.g., stenosis, arrhythmias, sleep apnoea).
3. To improve the quality of the simulated data by applying at least 3 different machine learning methods for each of the 2 clinical applications. This may include methods involving data generation, data transformation, optimal sampling of data or feature extraction approaches.
4. To generate reference data sets by combining the developed physiological modelling and machine learning methods. At least 1 improved synthetic reference data set will be created and published through an online repository for each of the 2 clinical applications. In addition, to create and publish a comprehensive software guide that will include recreation of the generated data sets, investigation of their authenticity and evaluation of their suitability for machine learning applications.
5. To facilitate the take-up of the developed evaluation framework and reference data sets by the measurement supply chain (NMIs, DIs, metrology networks, medical device calibration services), standards developing organisations, and end users (medical machine learning communities, manufacturers of medical and healthcare products).

These objectives will require large-scale approaches that are beyond the capabilities of single National Metrology Institutes and Designated Institutes, and it is expected that multidisciplinary teams will be required. To enhance the impact of the research, the involvement of the appropriate user community such as medical

practitioners, medical (academic) hospitals and industry is strongly recommended, both prior to and during methodology development. Where relevant, proposals are encouraged to build on, or seek collaboration with, existing projects and develop synergies with other relevant European, national or regional initiatives and funding programmes. In particular, links are encouraged with (i) the projects funded under earlier relevant topics of the Horizon Europe programme; or (ii) other relevant European Partnerships.

Proposers should establish the current state of the art and explain how their proposed project goes beyond this. In particular, proposers should outline the achievements of the EMPIR project 18HLT07 MedalCare or Metrology Partnership project 22HLT01 QUMPHY and how their proposal will build on those.

Proposers should note that the programme funds the activity of researchers to develop the capability, not the required infrastructure and capital equipment, which must be provided from other sources.

EURAMET expects the average EU Contribution for the selected JRPs in this TP to be 2.1 M€ and has defined an upper limit of 2.6 M€ for this proposal.

EURAMET also expects the EU Contribution to the external funded beneficiaries to not exceed 35 % of the total EU Contribution across all selected projects in this TP.

Any industrial beneficiaries that will receive significant benefit from the results of the proposed project are expected to be beneficiaries without receiving funding or associated partners.

Potential Impact

Proposals must demonstrate adequate and appropriate participation/links to the 'end user' community, describing how the project partners will engage with relevant communities during the project to facilitate knowledge transfer and accelerate the uptake of project outputs. Evidence of support from the "end user" community (e.g. letters of support) is also encouraged.

You should detail how your proposal's results are going to:

- Address the SRT objectives and deliver solutions to the documented needs,
- Feed into the development of urgent documentary standards through appropriate standards bodies,
- Facilitate improved industrial capability, or improved quality of life for European citizens in terms of personal health, protection of the environment and the climate, or energy security,
- Transfer knowledge to the healthcare sector.

You should detail other impacts of your proposed JRP as specified in the document "Guide 4: Writing Joint Research Projects (JRPs)"

You should also detail how your approach to realising the objectives will further the aim of the Metrology Partnership to develop a coherent approach at the European level in the field of metrology and include the best available contributions from across the metrology community. Specifically, the opportunities for:

- improvement of the efficiency of use of available resources to better meet metrological needs and to assure the traceability of national standards
- the metrology capacity of EURAMET Member States whose metrology programmes are at an early stage of development to be increased
- organisations other than NMIs and DIs to be involved in the work.

Timescale

The project should be of up to 3 years duration.

Additional information

The links provided in this section are only correct at the time of publication up until the end of the Call year.

The references below were provided by PRT submitters; proposers should therefore establish the relevance of any references.

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation) https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

[2] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (Artificial Intelligence Act) https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng