Publishable Summary for 22HLT01 QUMPHY
Uncertainty quantification for machine learning models applied to photoplethysmography signals

Overview

Photoplethysmography (PPG) signals are rich in information and easy to measure passively without any physical or mental limitations of the subject. As it is impossible for physicians to infer physiological parameters from the PPG signal by themselves, they need to rely on algorithms based on machine learning (ML) techniques for diagnosis. As of today, no regulations specifying how these ML algorithms have to be applied, how their performance has to be measured or how their associated uncertainties have to be specified exist. At the core of this project stands the development of measures to quantify the uncertainties associated with ML algorithms applied to medical problems, in particular the analysis and processing of PPG signals. To achieve this the following tasks will be addressed: (i) benchmark datasets will be generated using publicly available in vivo, and synthetic data (ii) different ML models and uncertainty quantification (UQ) methods will be used to analyse the processing of the PPG signals and specify the associated uncertainty and (iii) a good practice guide with accompanying software repository showcasing the used models, methods and benchmarks will be developed and made publicly available.

Need

A Photoplethysmograph measures the intensity of reflected or transmitted light through body tissue. Measured at almost any part of the human body (e.g., finger, wrist, arm, ankle, ear, neck), PPG signals change throughout the cardiac cycle as the changing volume of blood present in the tissue influences the intensity of the transmitted/reflected light. PPG signals are used both in a clinical setting and are collected by many wearable devices. They contain valuable information on the cardiovascular, respiratory, and autonomic nervous systems which is not yet routinely exploited. They are popular as they are easy to obtain non-invasively and PPG devices are cheap and widely available. Moreover, using smart devices or even contactless measurements by external cameras yields the possibility for long-term monitoring without restriction of patient comfortability. Until today, an algorithmic evaluation of PPG signals to infer physiological parameters or detect diseases is crucial for saving patients' lives but almost never used in clinical environments. One of the major reasons for this is the lack of trust in the output of any such algorithms.

Due to the vast amount of collected data, machine learning methodologies are essential for the extraction and evaluation of key features used for diagnosis. When applying machine learning in a medical context, however, confidence in the performance and predictions of the algorithms is particularly crucial since diagnostic mistakes can be fatal (false negative) or result in unnecessary anxiety and detrimental overtreatment (false positive). This is supported by the Executive Vice-President for a Europe fit for the Digital Age, Margrethe Vestager, who said "On Artificial Intelligence, trust is a must, not a nice to have". Hence an analysis of the uncertainty associated to ML algorithms and their predictions is indispensable to provide both clinicians and users of wearable devices with critical information about the quality and trustworthiness of the results produced. According to the US Center for Disease Control and Prevention (CDC) over 400 million people worldwide suffer from diabetes with predictions reaching 0.5 billion people worldwide by 2040. Similarly, the World Health Organisation (WHO) estimates that hypertension (elevated blood pressure) affects more than 1.2 billion adults between ages 30 and 79 worldwide with approximately half of the people being unaware that they have this condition or need treatment. A recent EC report states that: "By using digital solutions, such as wearables (…)

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EURAMET. Neither the European Union nor the granting authority can be held responsible for them.

European Partnership Co-funded by the European Union

The project has received funding from the European Partnership on Metrology, co-financed from the European Union's Horizon Europe Research and Innovation Programme and by the Participating States.

citizens can actively engage in health promotion and self-management of chronic conditions. This in turn can help control the rising demand for health and care". Wearable devices such as smart watches allow early detection, monitoring and counteracting such conditions through lifestyle and diet changes or early medical treatment potentially saving millions of lives and billions in medical treatment costs. One of the major obstacles of this approach originates from the low signal quality obtained from PPG measurements through smart wearables at e.g., the upper side of the wrist. This again emphasises the need for both ML techniques to process the data, as well as reliable and trustworthy methods to quantify the uncertainties associated to the algorithms' predictions.

The goal of this project is to satisfy these needs by developing an environment, i.e., a good practice guide including a software framework for independent assessment of accuracy and uncertainty of ML algorithms (Objectives 1 and 3) and benchmark cases to test and compare ML algorithms against (Objective 2), to increase trust in ML applications for PPG signals and lay a foundation towards standardisation of ML in healthcare (Objective 4).

Objectives

The overall objective is to provide trustworthy machine learning models for analysing photoplethysmography signals in a medical context, by developing methods for the quantification of uncertainty in supervised machine learning and deep learning models applied to photoplethysmography signals and generating reference datasets to benchmark those models, supported by software being developed that will be publicly available for independent review of the models.

The specific objectives are:

1. To develop methods for quantifying the uncertainty for at least 3 existing classification and 3 existing regression supervised machine learning and/or deep learning models using photoplethysmography (PPG) data, considering the effects of both aleatoric (data) uncertainty and epistemic (model) uncertainty on model predictions.
2. To generate at least 5 measurement problems and their corresponding 5 datasets, using real and/or synthetic photoplethysmography data, that can be used to benchmark accuracy and uncertainty of supervised machine learning and deep learning models. In addition, to make those reference problems and datasets available to the medical device and digital health communities via an online repository.
3. To validate the uncertainties obtained for existing machine learning and deep learning models of Objective 1 and to compare the accuracy and uncertainty of at least 3 classification and 3 regression machine learning and/or deep learning models in order to identify models and methods which have high accuracy and low uncertainty for a wide range of tasks.
4. To engage with the medical device, digital and health communities to (a) promote the use of the good practice guide and the accompanying software repository through conference contributions, peer-reviewed journal articles and stakeholder workshops, (b) support the adoption of the benchmarking problems and datasets by providing guidelines for their use, and (c) develop a framework for independently reviewing machine learning models proposed by industry to assist them in getting regulatory approval.
5. To facilitate the take up of the technology and measurement infrastructure developed in the project by the measurement supply chain (NMIs, DIs, medical device calibration services), standards developing organisations (IEC, ISO), and end users (clinical practitioners, digital experts within the health communities, manufacturers of medical and healthcare products).

Progress beyond the state of the art and results

Generate benchmark measurement problem datasets:

Accompanied by leading stakeholders the consortium will identify relevant benchmark diagnostic tasks for PPG signals, such as detection of atrial fibrillation and hypertension or blood glucose monitoring and will collect the required datasets from openly available databases and measurements provided by members of the consortium. These datasets may include in vivo (human), in vitro (phantom) and synthetic (simulated) measurements of PPG signals at different locations of the human body. The project will apply uncertainty quantification methods to machine learning/deep learning algorithms using PPG data. This will include model agnostic methods and non-model-agnostic methods which will require the algorithms themselves to be updated to generate uncertainty output. A classification of PPG signals by biological sex will be performed in order to determine whether there is a discernible difference between PPG signals for males and females, considering both performance and uncertainty. This database may include benchmarks focussing on e.g., sex and skin tone, which will foster the investigation of heterogeneous and more diverse populations.

Develop methods for quantifying the uncertainty of supervised ML/DL models:

The detection of quantities of interest from PPG signals cannot be done via direct observation but relies on the solution of inverse problems. While for direct measurements the determination of the measurement uncertainties is clearly covered by the GUM, for indirect measurements this is very challenging and a topic of current metrological research, which requires substantial modelling and goes far beyond the framework of the GUM. Additionally, as no closed-form analytical models to describe the human body with arbitrary accuracy exist, the derivation of surrogate models through modern ML algorithms to approximate any functional relation between measurements and hidden quantities are necessary. These surrogates, however, introduce an unavoidable and often unobservable approximation error caused by network architecture and training on noisy data. Especially for neural networks, due to their excellent performance on many cutting edge industrial and societal problems, generalisation and explainability are at the centre of modern research. In other words, methods to quantify the trustworthiness and confidence of ML algorithms are essential for applications, in particular, in high-stakes areas such as diagnosis and monitoring of diseases. The consortium will investigate which UQ methods are suitable for different ML algorithms applied to PPG signals for the developed benchmark problems. Therefore, the UQ methods employed will be comprised of model dependent and independent approaches encompassing both aleatoric (data) and epistemic (model) uncertainties. The ML models and UQ methods investigated in this project will cover a wide spectrum of algorithms providing a basis to establish confidence in applications even beyond what is covered by the benchmarks defined in this project. Further, these UQ methods can be seen as a first step towards standardisation and certification of machine learning in medical tasks using PPG signals. This is the first time such a large collection of different ML/DL models and UQ methods are developed and investigated to jointly be used to infer hidden parameters or detect diseases from PPG signals.

Validation of uncertainties of ML/DL models for the benchmark problems:

Using the UQ methods investigated, and the benchmark problems defined in this project, the consortium will be able to accompany ML models trained on PPG signals with an uncertainty budget, which goes far beyond any state-of-the-art application. The uncertainty budget will distinguish between aleatoric contributions caused by measurements with varying levels of signal complexity and quality, and epistemic contributions due to the specific choice of ML model and optimisation routine during the training phase. Moreover, as the benchmark problems are not of academic nature, but include real measurements for clinically relevant scenarios, the validation of the investigated algorithms by several different UQ methods will yield the first results on the credibility associated to diagnostic routines already used in consumer devices. Further, the validation procedure itself will create a precedent for the evaluation of an uncertainty budget in scenarios related to those of the benchmark problems possibly serving as a basis to inform future certification standards.

Engage with medical device, digital and healthcare communities:

The consortium and the stakeholder committee that will be formed after the start of the project will include members from the medical device, digital and healthcare communities, which will ensure that the benchmark scenarios considered will be relevant for and needed by medical device manufacturers as well as clinicians. This is the first-time benchmark datasets will be defined to achieve comparability of ML algorithms trained on PPG signals within the EU. The main output of the project will develop a good practice guide describing the defined benchmark datasets, the ML algorithms employed and the UQ concepts considered. This guide will additionally focus on explaining how the different methods should be applied to the respective benchmark problems functioning as a stand-alone guideline for the targeted end-user communities, thus building the foundation towards guidelines for regulatory approval of ML applications in healthcare areas involving PPG signals. To improve the impact of the good practice guide, the benchmark datasets will be made publicly available, including meta data on their origin, their proposed use and validation results acting as target baselines for user communities. Additionally, a software repository containing the ML algorithms and UQ methods used for the validation as well as required information on setup and training procedures will be made available to the medical device, digital and healthcare communities. The repository will include tutorials showing how to apply the implemented algorithms and methods to the benchmark problems, specifying all necessary software interfaces of the framework to allow easy adaptation to similar scenarios by potential users. These tools will be disseminated throughout the target user communities with a very high expected engagement rate, and the good practice guide, benchmark datasets and software repository will enable end users to establish trust in both medical device production and usage in an unprecedented fashion.

Outcomes and impact

Outcomes for industrial and other user communities

PPG signals are collected by many wearable devices, such as smart watches which are now widely available. In 2021, sales of smart watches worldwide were estimated to be 142 million, and this figure is projected to almost double over the next 4 years. By making digital health apps based on machine learning available on smart watches, individuals will be able to monitor different aspects of their own health. This would be possible both for the general population, for example monitoring blood pressure, or for a specific health need, such as the monitoring of blood glucose levels for diabetics. Combining uncertainty quantification with the machine learning predictions will ensure that only good quality predictions are presented to users which will in turn mean that they learn to trust the predictions and so will continue to use the app. Such continuous monitoring will result in early detection of health conditions, such as hypertension (high blood pressure), and early detection and treatment invariably results in better health outcomes and avoidance of more serious and costly health conditions and hospital admissions which can arise from undetected problems. Similarly, continuous monitoring of chronic conditions, such as blood glucose monitoring for diabetics, will enable patients to manage these conditions more effectively, and thus avoid the complications and costs associated with poor health management. This in turn will result in a lower of demand on the health system, which will result in economic benefit.

This was summarised in a recent European Commission report which states that “By using digital solutions, such as wearables (…) citizens can actively engage in health promotion and self-management of chronic conditions. This in turn can help control the rising demand for health and care”.

In a hospital setting, pulse oximeters currently monitor patients’ heart rate and blood oxygen continuously. By incorporating additional machine learning algorithms into these monitors, many more aspects of a patient’s health and well-being could be monitored, such as atrial fibrillation or detection of the onset of sepsis, which is potentially life threatening. Alarms could be triggered if the algorithm detects an adverse change in a patient and information provided on which condition the algorithm has detected, thus aiding clinical staff. Automatic detection systems that give unreliable alarms are often ignored, so the uncertainty quantification will be essential to provide some confidence in the alarms. This will provide continuous monitoring of all patients, which nursing staff cannot provide, and will enable early detection of health deterioration which will result in better health outcomes and will also translate into economic benefit.

Outcomes for the metrology and scientific communities

This project will develop new methods for quantifying the uncertainty of machine learning predictions that are based on the use of features, image transformations of the signal, and the raw signal. Machine learning is applied to many problems including autonomous vehicles, medical imaging, and industrial sensor networks for which quantification of uncertainties is equally important and so the methods developed in this project are widely applicable in other application domains. The benchmarking datasets will be of benefit to the metrological and scientific communities who may want to use these datasets with their own machine learning models or for other studies. Research papers will be submitted for publication in high impact peer-reviewed journals and the work in the project will be presented at relevant international conferences.

Outcomes for relevant standards

PTB, NPL, LNE, IPQ and IMBiH will contribute to national and international standards and guidelines throughout the project, especially for AI in medicine. This includes dissemination of the project’s results to standard committees to propagate the results and make them available to the user community. Special attention will be given to developing a good practise guide for uncertainty quantification of ML algorithms applied to PPG signals, which can act as a foundation for standardisation of PPG based medical applications in the future. The consortium anticipates high impact of the mathematical tools and advanced uncertainty quantification and propagation methods through international committees such as IMEKO TC 6, ISO/TC 69, ITU/WHO FG-AI4H and JCGM Working Group 1.

Longer-term economic, social and environmental impacts

Hypertension, diabetes, and myocardial infarction rank among the most common causes of death in human populations worldwide. Often, especially in the early stages of the diseases, a change in lifestyle and diet can be sufficient to mitigate these diseases, eradicating the need for expensive and possibly harmful treatment or medication.

PPG signals are collected by smart watches which are now widely available worldwide. By making digital health apps based on machine learning available on smart watches, individuals will be able to monitor different aspects of their own health. This would be possible both for the general population, for example monitoring blood pressure, or for a specific health need, such as the monitoring of blood glucose levels for diabetics. Combining uncertainty quantification with the machine learning predictions will result in more trustworthy predictions. Such continuous monitoring will result in early detection of health conditions and better management of chronic conditions, both of which will result in better health outcomes leading to reduced demand on health systems with corresponding economic benefits. Similarly in a hospital setting, patients could be monitored continuously which will enable early detection of health deterioration resulting in better health outcomes which will translate into economic benefit. This project will contribute to the rapidly growing digital health industry which is providing wearable devices that enable individuals to continuously monitor their own health and well-being resulting in early detection of health issues or better management of chronic conditions, as mentioned above. This will result in better health for users and, in some cases, to a longer life. In hospitals, continuous monitoring of all patients will result in trustworthy early alerts of health deterioration enabling early treatment which will improve health outcomes and may even save lives.

As a wider impact, this project will provide a boost to Europe’s rapidly growing digital health industry, leading to higher skilled employment and wealth for society, by providing digital healthcare companies as well as clinicians with an understandable and deployable good practice guide to assess the accuracy and uncertainty of ML algorithms in healthcare applications. Additionally, this guide can be seen as a foundation for standardisation of ML in health, making it easier to get regulatory approval of healthcare applications based on ML in the EU in the future.

List of publications

-

This list is also available here: https://www.euramet.org/repository/research-publications-repository-link/

Project start date and duration:	1 July 2023, 36 months	
Coordinator: Nando Hegemann, PTB Tel: +49 30 3481 7773 E-mail: qumphy@ptb.de Project website address: www.qumphy.ptb.de		
Internal Beneficiaries: 1. PTB, Germany 2. CMI, Czechia 3. IMBiH, Bosnia and Herzegovina 4. IPQ, Portugal 5. LNE, France	External Beneficiaries: 6. FC, Belgium 7. FVB, Germany 8. KTU, Lithuania 9. THM, Germany 10. UGent, Belgium 11. Uni-Oldenburg, Germany	Unfunded Beneficiaries: -
Associated Partners: 12. KCL, United Kingdom 13. NPL, United Kingdom 14. SectorHealth, United Kingdom 15. SURREY, United Kingdom 16. UCAM, United Kingdom		