Final Publishable Summary for 18HLT07 MedalCare
Metrology of automated data analysis for cardiac arrhythmia management

Overview

The central goal of the project was to develop a novel validation strategy of cardiac arrhythmia classification algorithms based on multiparametric data analysis of electrocardiography (ECG) data through metrological research. A novel synthetic reference database was developed and made publicly available that will enable the assessment of the performance modern data analysis approaches, such as machine learning (ML) in medicine and will contribute to standardising ML methods and improving their quality measures in health applications, specifically by establishing a novel metrological validation platform of such algorithms. The project has supported the vision of improvement in the treatment and diagnosis of cardiovascular disease (CVD) through metrologically-validated diagnosis by automated analysis of multiparametric ECG abnormalities. The project developed multiple consistent large-scale 12 lead ECG databases that can be used as reference data and used with new statistical approaches, dimensionality reduction, and ML techniques. Further to this, the project developed and applied benchmarking protocols for automated algorithms for ECG diagnosis that provide quantitative information on their performance, associated uncertainties, and robustness. The project found that the uncertainties in ECG classification depend on many different factors such as size and composition of the training set, noise in the data or accuracy of labelling.

Need

This project supported the vision of personalised medicine in cardiovascular disease (CVD) through the development of metrologically validated automated analysis of ECG data. CVD is the most relevant and epidemiologically significant non-communicable disease in Europe. ECG is a non-invasive and cost-effective tool for the initial examination and monitoring of patients presenting with cardiac complaints. Prior to the start of this project the American Heart Association (AHA) established a list of 83 cardiovascular abnormalities requiring further research and expertise. As this lack of information resulted in an increased rate of misinterpretation among non-specialised physicians. ECG is essential in particular in the diagnosis of cardiac ischemia and arrhythmias. Arrhythmias remain one of the major causes of sudden cardiac death and stroke with a rising number of patients worldwide and especially in Europe. Guidelines of the European Society of Cardiology (ESC) stress the importance of ECG monitoring for atrial fibrillation, a cardiac arrhythmia with high prevalence. However, automated analysis tools are needed for the detection and classification of episodes of ischemia and arrhythmia. Future telemedicine and home monitoring systems will also boost the need for automated and validated ECG analysis.

Computer assisted diagnosis techniques have been used for the analysis of large volumes of measurement data. Recently, ML techniques have been applied as they have the advantage of and the ability to examine multivariate features not obvious to the human eye. However, key challenges of ML are the investigation of the influence of data uncertainty and the assessment of the techniques' uncertainty itself. Therefore, it is difficult to convince health professionals and patients to trust in algorithms that are so complex.

In addition, from a regulatory point of view, there is an urgent need for a metrological validation of ECG analysis algorithms using reference data with a traceable ground truth. Ground truth in medicine is a particular challenge and usually addressed either by consensus of multiple experts or use of synthetic data. The latter has the advantage that the uncertainty of data can be modified from "pure true data" to noisy and faulty annotated data to investigate its influence on analysis algorithms. However, the large databases of synthetic data required for ML, do not exist. Such databases would allow the direct comparison of algorithms with defined metrics, so called "benchmarking" and therefore are urgently needed.

Objectives

The overall goal of this project was to develop a novel metrological validation strategy of medical analysis algorithms that allowed traceability to digital reference values. The focus of the project was on multiparametric data analysis of ECGs and the specific objectives of the project were:

1. To develop synthetic ECG reference data of a virtual population. This would involve existing biophysical modelling frameworks to develop a synthetic ECG reference dataset allowing the assessment of uncertainty of automated data analysis methods such as ML. An ECG-database of a representative virtual population including healthy variations and selected pathologies will be generated.
2. To carry out the uncertainty analysis of reference data by assessing the sensitivity of different parameters on results of the biophysical modelling resulting in an uncertainty evaluation of the synthetic ECG data. For this, the influence of the model input parameters, such as anatomies, conduction blocks, tissue conductivity, infarction and fibrotic tissues, will be assessed.
3. To assess and compare the effect of different classification approaches focusing on uncertainty analysis along two directions: the influence of uncertainty of features of ECG data on the output of the classification algorithm and the influence of wrongly labelled training data on the output of the classification. The project would investigate whether hidden features can be detected by modern ML-approaches for “quantitative classification” of ECG.
4. To carry out thorough investigation of clinical application of multi-parametric data analysis that includes detection and classification of cardiac ischemia and arrhythmia. A comparison of performance of experienced physicians with multiparametric data analysis methods will be performed in the project - "Clinical Turing Test”.
5. To facilitate the take up of the technology and measurement infrastructure developed in the project by end users and by supporting the modification of the ANSI-A-AMI (EC38, EC57) standard to address the challenge of ML-approaches. This should include the publication of a guidance document on software validation to support the new EU medical device regulation (MDR 2017/745 and 2017/746) and clinical guidelines of the European Society of Cardiology (ESC).

Progress beyond the state of the art

Establishing synthetic ECG reference data of a virtual population.

One of the most important challenges of computer aided diagnosis of electrophysiological abnormalities is the lack of ground truth reference data. Prior to the start of this project, although a number of clinical datasets have been made available on physionet.org, such as MIT-BIH or the PTB-database, they all rely on correct labelling by clinical experts. Thus, any wrongly labelled data can result in the unpredictable behaviour of algorithms. Over the last decade biophysical modelling of cardiac electrophysiology has allowed the simulation of the electrophysiology of the heart and the realisation of the resulting electrical potentials on the body surface (i.e. the ECG). The project went beyond the stat of the art by using fast modelling frameworks to generate synthetic ECG data for different types of conduction blocks and tissue alterations. From the data the project has established for the first time ever a “virtual population” of ECG-data with digital ground truth reference values.

Enabling traceability of medical data analysis algorithms.

Over the last decade, the number of ML classification tools, like support vector machines (SVM) and “deep neural network” approaches, has significantly increased for ECG analysis. But unfortunately, the state of the art for uncertainty evaluation in metrology as provided by the GUM cannot be applied to these approaches. Recently, the concept of layerwise relevance propagation (LRP) has been introduced in the ML community. LRP determines which features in the input data contribute most strongly to a neural network's output by running backward in the neural network process. However, traceability of LRP algorithms requires a ground truth, which is very seldom available, e.g. from large clinical outcome studies. This project has gone beyond the state of the art by developing synthetic reference data that can provide a digital ground truth and allow the investigation of traceability and uncertainties for both the data and of the non-linear algorithm.

Better understanding of ECGs features for an improved clinical classification of arrhythmia.

Some cardiac diseases are known to lead to characteristic changes in ECGs that can be identified by visual inspection, whilst others lead to a significant number of false negatives. Some diseases also lead to very small changes in the ECG that can only be identified using advanced algorithms of bio-signal analysis. This project went beyond the state of the art by quantitatively determining whether selected cardiac diseases lead to significant changes in the ECG that can be detected reliably or whether they are hidden and need further examination beyond ECG. This will enable a better characterisation of heart diseases through improved classification.

Improved regulatory approval pathway and higher clinical acceptance of modern software approaches.

The project created synthetic and clinical reference data sets and exemplary benchmark protocols for modern analysis software as first steps towards a regulatory approval process in medical ECG-devices.

Results

Objective 1: To develop synthetic ECG reference data of a virtual population.

The project has generated a reference database of synthetic ECG signals (MedalCare-XL) obtained from simulations in a cohort of models representing the variability in humans. The synthetic database contains a total of 16,900 lead ECGs based on multi-scale mechanistic electrophysiological simulations equally distributed into 1 healthy control and 7 pathology classes. The pathological class of myocardial infarction has 6 sub-classes with varying location and extent of the infarcted zone. Clinically important cases of conduction block were included and in one version of the data, realistic noise was added to the synthetic ECGs. Besides the noise-free original signals, filtered noisy signals were also included. The project's reference database of synthetic ECG signals was made publicly available on Zenodo (https://doi.org/10.5281/zenodo.7293655 and in a technical report https://doi.org/10.48550/arXiv.2211.15997). All modelling parameters as well as the data structure of the synthetic ECGs are described in detail to achieve full traceability.

To generate the synthetic 12 lead ECGs, existing biophysical modelling tools were used to simulate the electrical activity of the heart through a range of parameter variations. The influence of the positioning of the ventricles, the atrial fibre orientation and the realisation of the Purkinje system on the ECG was investigated and the results showed that all these aspects contribute to important features of the ECG signal and are necessary to produce synthetic ECG signals that resemble real clinical ones.

The project has assessed parameter sensitivities and the uncertainty of different model formulations with different methodologies. More specifically, two independent approaches were used for modelling the atrial and ventricular contribution to the ECG respectively. (1) For simulating ventricular ECGs, a virtual cohort of heart geometries and torsos was created from MRI data of healthy volunteers. (2) The simulation of atrial signals was carried out with statistical shape models humans. Advanced methods were then developed to combine the atrial P-wave with the ventricular QRS-complex and T-wave.

In addition, methods were developed to create a series of 10 successive heartbeats that show a natural variation (e.g. heart rate variability). The two modelling approaches were also used and tested in a number of relevant cardiological applications such as (i) the estimation of the degree of atrial fibrosis, (ii) the localisation of atrial flutter substrates, (iii) generation of personalised atrial models, and (iv) the effect of electrode placement on ECG signals.

The simulated ECG data in the MedalCare-XL data base was evaluated in a validation process, that compared the distribution of features for healthy simulated 12 lead ECGs and several classes of pathological 12 lead ECGs with the variability of ECG features in clinical data extracted from the PTB-XL database (Objective 3). This was then summarised in the new PTB-XL+ data base (Objective 3) containing electrocardiographic features for more than 20,000 ECGs. To be able to do this, a software tool was developed to extract features from ECGs (ECGDeli) and comparison of the extracted timing and amplitude features demonstrated similar characteristics. Thus, the project's novel dataset of simulated ECG signals can be successfully used in the future for development and testing of ML algorithms for CVD.

Objective 2: To carry out the uncertainty analysis of reference data by assessing the sensitivity of different parameters on results of the biophysical modelling resulting in an uncertainty evaluation of the synthetic ECG data.

Two methods for sensitivity analysis were selected and implemented by the project. One method is based on feature extraction and relates changes in input parameters of biophysical models to changes in the ECG features. The second method is a combined analysis of sensitivity to input parameters and uncertainty quantification for the complete ECG time series and uses a polynomial chaos expansion (PCE) based surrogate model. Due to the high computational demand of a single simulation and the large parameter space of the models, the surrogate-based sensitivity analysis was used to analyse different aspects of ECGs with respect to different sets of input parameters. With a PCE surrogate, the uncertainty quantification is improved due to the ability for very quick sampling. By drawing samples from appropriate parameter distributions, it then became possible to propagate known parameter uncertainties through the PCE surrogate to assess the associated signal variations.

Of clinical importance are intervals and amplitudes derived from the ECGs, (so-called features), to interpret the signal and diagnose conditions. These 'so-called feature' quantities can be obtained from ECG signals automatically using software. The sensitivity of these quantities to the inputs in the underlying models has also been explored by the project. In this work, two methods of numerical calculation of Sobol indices were compared and yielded comparable results.

Real ECG signals are noisy due to various known sources of uncertainty during the measurement process, that cannot be accounted for in the numerical models of the heart. The project has successfully identified three of the main sources of noise in ECGs. The project then developed methods to simulate the effects of these 3 sources and have used them to produce noisy versions of a set of synthetic ECGs.

Objective 3: To assess and compare the effect of different classification approaches focusing on uncertainty analysis along two directions: the influence of uncertainty of features of ECG data on the output of the classification algorithm and the influence of wrongly labelled training data on the output of the classification.

The PTB-XL database, consisting of nearly 22,000 12-lead 10 second ECG signals, was made publicly available on PhysioNet (see https://doi.org/10.13026/x4td-x982 and the associated paper https://doi.org/10.1038/s41597-020-0495-6). The project also generated a second dataset, PTB-XL+, where widely used clinical ECG features were extracted with three different ECG analysis software packages, two commercial ones and the open-source software package ECGDeli developed within Objective 1. Non-standard ECG features have also been considered using the PTB-XL+ and used for benchmarking. The PTB-XL+ database is available on Zenodo (https://doi.org/10.5281/zenodo.7817567).

Several datasets of 12 lead ECGs were constructed for use with ML from publicly available datasets, including a medium dataset (2,863 records) and a large dataset (10,406 records). ML was used on these features for classification of the signals. In addition, images were generated from the signals in the form of scalograms and Symmetric Projection Attractor Reconstruction attractors and transfer learning was used on these images for classification. Finally, the raw signals were used as input for a 1D convolutional neural network (CNN) model.

The project found that the 1D CNN model generally gave the best classification results. Further improvements were obtained by using self-supervised representation learning and structured state-space models.

In addition to this, the project investigated how noise on the electrocardiogram signal can result in deep learning misclassifications and how to subsequently to address this. The work showed that if the ECG signals to be classified are all clean, then the network should be trained on clean data. However, if the ECG signals are either clean or noisy, then the network should be trained on noisy data in order to give a more consistent performance.

Feature rankings were obtained using a selection of feature importance methods and were compared with the features used by cardiologists to diagnose three particular conditions; (i) AV block, (ii) left bundle branch block (LBBB) and (iii) right bundle branch block (RBBB). From the results it became clear, that random forest methods performed consistently well whereas logistic regression methods performed consistently poorly

Further work included studying the feature importance rankings for the multiclass classification of Normal, AV block, LBBB and RBBB. The effect of variations in the training data was also studied and showed that a 1D CNN model is particularly sensitive to the sampling frequency of the signals and gave significantly better results with a lower sampling frequency.

Further to this, the reproducibility of ML was investigated by training a ML model using three different training datasets but testing each model on the same test set. The project found that the training data had a big impact on the test results and that the training and test data should be qualitatively similar for good ML results to be obtained. The effect of label noise on model performance was also considered and shown that prediction of some labels is quite sensitive to noise and that other labels are quite insensitive to label noise. This was partly due to the number of samples per label in the test set, with the error generally decreasing with increasing number of samples for a label. As expected, model performance decreases with increasing label noise.

Objective 4: To carry out thorough investigation of clinical application of multi-parametric data analysis that includes detection and classification of cardiac ischemia and arrhythmia. A comparison of performance of experienced physicians with multiparametric data analysis methods will be performed in the project - "Clinical Turing Test".

The work aimed to develop a simulation protocol to produce realistic ECGs and evaluate the clinical capabilities of the generated signals contained within the synthetic ECG database in objective 1, as well as to validate the integrity of the underlying model of electrophysiology. As part of this work clinical Turing tests were conducted to check if the synthetic 12 lead ECGs within the synthetic ECG database would pass as clinical signals under diagnostic conditions. In the clinical Turing tests, medically trained cardiologists were asked to determine the origin of both measured and simulated 12 lead ECGs (type classification), as well as to make diagnostic predictions for clinical and synthetic ECGs representing pathological conditions (pathology classification). These clinical Turing tests provided valuable feedback to the project for improving modelling of electrophysiology. In particular, feedback on specific morphological features within the 12 lead ECG were shown to relate to model parameters that can be altered to improve diagnostic capabilities.

This work also provided an assessment of how realistic the synthetic ECG data from the database in Objective 1 appears to medically trained cardiologists. The overall validation results for the synthetic ECG database were successful and satisfactory and indicated possibilities for further refinements.

Further to this, the project conducted a study of a personalised ECG to ascertain whether the underlying model of cardiac electrophysiology is capable of representing the intrinsic nature of the heart, not only in terms of the 12 lead ECGs, but in all forms of non-invasive clinical data such as electro-anatomical maps and body surface potential maps. Overall, we successfully demonstrated that a validation to ensure model integrity and performance can be made through comparison of different synthetic and measured ECG signals or signal-derived features such as activation maps.

Impact

Results of the project have been disseminated through 30 open access publications in peer-reviewed journals, 31 presentations at conferences and 5 open-source datasets.

In connection with the publication of the PTB-XL database (Objective 3 and validated in Objective 4) and the related ML benchmark paper on ECG classification, PTB issued a press release.

Partner KIT has also developed an open-access software tool for the extraction of ECG features (ECGDeli from Objective 1) from clinical and virtual datasets, that is available to end users.

Impact on industrial and other user communities

This project has established a synthetic ECG reference database and assessed the performance of algorithms using metrologically sound methods. The project has ensured open access of its synthetic ECG database MedalCare-XL (Objective 1) and two extended data sets of clinically recorded ECGs (PTB-XL and PTB-XL+ from Objective 3) that were used for validation of synthetic data and for the benchmarking of algorithms. This will allow free access to end users in the medical and other user communities to such important reference data, its sensitivity analysis and uncertainty analysis.

Further to this partner KIT's software tool for the extraction of ECG features (ECGDeli) from clinical and virtual datasets, is also available open access at https://github.com/KIT-IBT/ECGdeli.

Clinicians were part of the project team and were used together with external medically trained cardiologists for the clinical Turing tests (Objective 4). They were also used as advisors for the design of the MedalCare-XL database and PTB-XL, PTB-XL+ datasets (Objectives 1 & 3). This important input supports the usability and clinical application of the project's results. In particular, in an assessment of how realistic the synthetic ECG data from the database in appears to medically trained cardiologists – the overall results were successful and satisfactory and indicated possibilities for further refinements.

This project has provided important insights in devising reference data as well as benchmark tests for automated AI algorithms for ECG diagnosis. Significant findings (Objectives 1, 2,3 and 4) include:

- The project has established a modelling pipeline ("digital twin") to simulate human 12-lead ECGs in a realistic manner by building modules for the ventricular and atrial signals separately and by devising a procedure of stitching the respective contribution together to yield a complete ECG recording. In addition, the models allow for varying anatomic and electrophysiological features in order to map a whole population (virtual cohort) rather than a single "personalized" ECG.

- Based on this pipeline, the project has created a new MedalCare-XL database of more than 16,000 simulated ECG signals and developed a number of validation strategies such as a direct comparison to a lager clinical database via considering distribution of ECG features, cross validation by applying a neural network-based ML algorithm and validation by experts within the framework of a clinical Turing test.
- The project has identified three of the main sources of noise in ECGs, which were elect rode movement, motion artefacts, and baseline wander. The identification of these sources is important as the noise in real ECG signals cannot be accounted for in the numerical models of the heart and has to be added to make the synthetic signals look comparable to the clinical ones.
- 1D CNN models (i) gave the best classification results, (ii) are particularly sensitive to the sampling frequency of signals (during feature analysis) and (iii) gave significantly better results with a lower sampling frequency
- to give a more consistent performance networks should be trained on (i) clean data. if ECG signals to be classified are clean and (ii) trained on noisy data if ECG signals are either clean or noisy.
- For the reproducibility of ML, - the training data has a big impact on test results and thus the training and test data should be qualitatively similar for good ML results to be obtained.

In January 2023 the EU started the Testing and Experimentation Facilities (TEF) Health project https://www.tefhealth.eu/. TEF Health is part of the EU's Digital Europe Programme, which involves this project's partners PTB, LNE, and FhG who are involved in establishing agile approval processes for trustworthy AI. A particular focus is on the definition of data quality and the TEF health project plans to use the three open access databases from this project (Objectives 1 & 3) as use cases.

Impact on the metrology and scientific communities

The project's outputs have supported the expansion of European metrology in the growing field of digital health. The project has demonstrated the impact of metrologically sound approaches for the development of novel ECG analysis approaches (Objectives 3 & 4). In addition, the synthetic ECG database, it's uncertainty and sensitivity analysis and extended datasets produced in this project (Objectives 1, 2 & 3) should enable key comparisons using traceable digital reference values. The external project partners (i.e. A-A, KCL, KIT, MUG and TUB) are leading experts in their fields and are active in European framework projects. In particular, this project has benefitted from collaborations with the H2020 multidisciplinary training network myAtria which is focussed on the diagnosis of atrial fibrillation.

Dissemination of the project's results o metrological and scientific communities has also been through open access publications in scientific journals and presentations and proceedings of conferences in computational cardiology, biomedical engineering, ML and mathematics for metrology. In addition, several young scientists involved in the project have won awards for their work: at KIT a young scientist received the Best Oral Presentation Award at 12th Workshop on Statistical Atlases and Computational Modelling of the Heart in 2021; at MUG young scientists were honoured with the Best Collaborative BioTechMed-Graz Paper Award 2021 and an award from the Austrian Society for Biomedical Engineering; and at NPL a researcher was awarded NPL's Rayleigh Award for an outstanding contribution by an early career scientist.

The project's PTB-XL data base has already been used in the the PhysioNet/Computing in Cardiology Challenge 2020 (https://moody-challenge.physionet.org/2020/) for the classification of 12-lead ECGs as well as in the PhysioNet/Computing in Cardiology Challenge 2021 (https://moody-challenge.physionet.org/2021/) on varying dimensions in electrocardiography. Therefore, PTB-XL has become the most widely used ECG-dataset with well-defined data for training and testing of new AI-based ECG-analysis tools. Recently, the PTB-XL dataset has been extended by extracting features using different public-domain and commercial software. The extended version has been published as PTB-XL+ open dataset allowing the benchmarking of new feature extraction software.

The project also had strong links to EURAMET's European Metrology Network (EMN) on Mathematics and Statistics in Metrology (MathMet) and its clinical PTB-XL dataset will be used as an example of a reference dataset in the 18NET05 project associated with the EMN MathMet.

Impact on relevant standards

In Europe, the medical device regulations (MDR 2017/745, 2017/746) are traditionally focussed on physical devices and unfortunately, current trends in software are only indirectly addressed. In the US, the Food and Drug Administration (FDA) has been reviewing the rising importance of software for medical devices. In the area of automated ECG analysers, the FDA has recognised the ANSI-A-AMI (EC38, EC57) standard for the approval of new devices. In response to this, this project has drafted software guidance for testing ECG analysis software, which emphasises the importance of clinical reference data along with the potential advantages of synthetic reference data. The document also gives recommendations and an outline for benchmark tests of AI algorithms. In the document the project defined protocols for evaluating the performance of automated ECG analysers, which are recognised by the FDA as consensus standards. ANSI-A-AMI EC38 specifies the use of the MIT, AHA and ESC databases., however, these databases only contain a limited amount of ECG signals (<200) hampering their application to ML approaches. Furthermore, the new ISO 80601-2-86 standard requires automated ECG analysis and interpretation and states the limitations of current CSE and CTS test data sets (i.e. small numbers of signals). Although the concepts can also be applied to ML approaches, current standards stress the importance of reference datasets for development, testing and comparison of ML algorithms.

The results of the project have also been presented to workshops of DIN (on safety of algorithms), to the CIPM's Task Group on the Digital SI (CIPM-TG-DSI), to JCTLM WG1 Uncertainty and to the WHO (World Health Organisation) focus group AI4Health.

Longer-term economic, social and environmental impacts

Each year CVD causes 3.9 million deaths in Europe. CVD manifests itself in diseases such as coronary artery disease, congestive heart failure, and cardiac arrhythmias. Overall CVD is estimated to cost the EU economy € 210 billion a year. Therefore, early detection, reliable diagnosis and cost-effective management of CVD are key for improving patient care and for reducing healthcare costs. In the longer-term this project will support more effective diagnosis of cardiac arrhythmias by assessing the performance of new medical devices. Especially as the application of automated analysis software is now an integral part of many medical devices.

It is expected that over the next decade, more biosensors-based products (e.g. smart ECG devices) for use outside the hospital will be developed for monitoring of patients. This project should have a wider impact on this rising home-care market. In particular, home monitoring devices are expected to be used more extensively in chronic patients and they will require new methods for their regulatory approval.

Outcomes from this project that are vital for future developments of automated algorithms for medical diagnostics are:

1. reference data sets need to be defined and provided. These need to be in the form of validated synthetic databases based on simulation with a "ground truth" and clinical databases based on validation by experts. For ML and AI algorithms there is a need for large databases (e.g. > 10000 ECGs as per this project).
2. Benchmark protocols for AI algorithms need to be developed and agreed. They must be fit-for-purpose for the given medical task (here ECG analysis) and should focus on the overall performance of the algorithm for given tasks (e.g. classifying ECGs as healthy and normal). Other necessary features that must be tested are (i) robustness (insensitivity of performance against poor data quality/ low signal-to-noise ratio) and (ii) explainability (identifying the aspects of the signal that prompted a classification by the algorithm).

A final key issue for the future development of the field is the availability of open data sets for the training of algorithms such as the ones developed in this project.

List of publications

1. Aston, P., Lyle J. V., Bonet-Luz E., Huang C. L. H., Zhang Y. Jeevaratnam K., Nandi, M. (2019), Deep learning applied to attractor images derived from ECG signals for detection of genetic mutation, **Computing in Cardiology**, https://doi.org/10.22489/CinC.2019.097

2. Wagner, P., Strodthoff, N., Bousseljot, R.-D., Kreiseler, D., Lunze, F.I., Samek, W., and Schaeffter, T. (2020), PTB-XL: A Large Publicly Available ECG Dataset. **Scientific Data**, Volume 7 (1), 154.https://doi.org/10.1038/s41597-020-0495-6.

3. Nagel, C., Pilia, N., Loewe, A., and Dössel, O. (2020), Quantification of Interpatient 12-lead ECG Variabilities within a Healthy Cohort. **Current Directions in Biomedical Engineering**, Volume 6(3): 20203127, https://doi.org/10.1515/cdbme-2020-3127

4. Strodthoff, N., Wagner, P., Schaeffter, T., and Samek, W. (2020). Deep Learning for ECG Analysis: Benchmarks and Insights from PTB-XL. **IEEE Journal of Biomedical and.Health Informatics**, Volume 25 (5) 1519. https://doi.org/10.1109/JBHI.2020.3022989

5. Pilia, N., Severi, S., Raimann, J. G., Genovesi, S., Dössel, O., Kotanko, P., and Loewe, A. (2020). Quantification and classification of potassium and calcium disorders with the electrocardiogram: What do clinical studies, modeling, and reconstruction tell us? **APL Bioengineering** 4(4), 041501. https://doi.org/10.1063/5.0018504

6. Venton J., Harris P. M., Sundar A., Smith N. A. S., Aston P. J. (2021). Robustness of convolutional neural networks to physiological electrocardiogram noise. **Philosophical Transactions of the Royal Society A**, Volume 379 Article ID:20200262. https://doi.org/10.1098/rsta.2020.0262.

7. Pilia, N., Nagel, C., Lenis, G., Becker, S., Dössel, O., Loewe, A. (2021). ECGdeli - An open source ECG delineation toolbox for MATLAB. **SoftwareX**. Volume 13, 100639; https://doi.org/10.1016/j.softx.2020.100639

8. Nagel, C., Schuler, S., Dössel, O., and Loewe, A. (2021). A bi-atrial statistical shape model for large-scale in silico studies of human atria: Model development and application to ECG simulations, **Medical Image Analysis**. Volume 74, 102210. https://doi.org/10.1016/j.media.2021.102210

9. Nagel C, Luongo G, Azzolin L, Schuler S, Dössel O, Loewe A. (2021). Non-Invasive and Quantitative Estimation of Left Atrial Fibrosis Based on P Waves of the 12-Lead ECG—A Large-Scale Computational Study Covering Anatomical Variability. **Journal of Clinical Medicine**, 10(8):1797. https://doi.org/10.3390/jcm10081797

10. Nagel, C., Dössel, O. & Loewe, A. (2021). Sensitivity and Generalization of a Neural Network for Estimating Left Atrial Fibrotic Volume Fractions from the 12-lead ECG. **Current Directions in Biomedical Engineering**, *7*(2), 307-310. https://doi.org/10.1515/cdbme-2021-2078

11. Welle, H., Nagel, C., Loewe, A., Mikut, R., Dössel, O. (2021). Classification of Bundle Branch Blocks with QRS Templates Extracted from 12-lead ECGs. **Current Directions in Biomedical Engineering,** 7(2), 582-585. https://doi.org/10.1515/cdbme-2021-2148

12. Dössel O., Luongo G., Nagel C., Loewe. A. (2021). Computer Modeling of the Heart for ECG Interpretation—A Review. **Hearts**.; 2(3):350-368. https://doi.org/10.3390/hearts2030028

13. Venton J.and Aston P. J. (2021) Investigating the Robustness of Deep Learning to Electrocardiogram Noise*,* **Computing in Cardiology***,* https://openresearch.surrey.ac.uk/esploro/outputs/conferencePaper/Investigating-the-Robustness-of-Deep-Learning/99634465502346?institution=44SUR_INST

14. Gillette, K., Gsell, M. A. F., Prassl, A. J., Karabelas, E., Reiter, U., Reiter, G., Grandits, T., Payer, C., Štern, D., Urschler, M., Bayer, J. D., Augustin, C. M., Neic, A., Pock, T., Vigmond, E. J., & Plank, G. (2021). A Framework for the generation of digital twins of cardiac electrophysiology from clinical 12-leads ECGs. **Medical Image Analysis**, Volume71, 102080. https://doi.org/10.5281/zenodo.7944373

15. Gillette, K., Gsell, M. A. F., Strocchi, M., Grandits, T., Neic, A., Manninger, M., Scherr, D., Roney, C. H., Prassl, A. J., Augustin, C. M., Vigmond, E. J., & Plank, G. (2022). A personalized real-time virtual model of whole heart electrophysiology. **Frontiers in Physiology,** Volume 13, 907190. https://doi.org/10.3389/fphys.2022.907190

16. Gillette, K., Gsell, M. A. F., Bouyssier, J., Prassl, A. J., Neic, A., Vigmond, E. J., Plank, G. (2021). Automated Framework for the Inclusion of a His-Purkinje System in Cardiac Digital Twins of Ventricular Electrophysiology. **Annals of Biomedical Engineering,** Volume 49(12), 3143–3153. https://doi.org/10.5281/zenodo.7944267

17. Bender, J., Nagel, C., Fröhlich, J., Wieners, C., Dössel, O. Loewe, A. (2022). A Large-scale Virtual Patient Cohort to Study ECG Features of Interatrial Conduction Block. **Current Directions in Biomedical Engineering**, Volume8(2), 97-100. https://doi.org/10.1515/cdbme-2022-1026

18. Karli Gillette, Matthias AF Gsell, Stefan Kurath-Koller, Anton J. Prassl, Gernot Plank (2022) Exploring Role of Accessory Pathway Location in Wolff-Parkinson-White Syndrome in a Model of Whole Heart Electrophysiology. **Computing in Cardiology** https://doi.org/10.22489/CinC.2022.057

19. Nagel, C., Osypka, J., Unger, L., Nairn, D., Luik, A., Wakili, R., Dössel, O. Loewe, A. (2022). Improving Clinical ECG-based Atrial Fibrosis Quantification With Neural Networks Through in silico P waves From an Extensive Virtual Patient Cohort. **Computing in Cardiology (CinC).** https://doi.org/10.22489/CinC.2022.124

20. Winkler B, Nagel C, Farchmin N, Heidenreich S, Loewe A, Dössel O, Bär M. (2023). Global Sensitivity Analysis and Uncertainty Quantification for Simulated Atrial Electrocardiograms. **Metrology**. Volume 3(1), 1-28. https://doi.org/10.3390/metrology3010001

21. Luongo, G., Vacanti, G., Nitzke, V., Nairn, D., Nagel, C., Kabiri, D., Almeida, T. P., Soriano, D. C., Rivolta, M. W., Ng, G. A., Dössel, O., Luik, A., Sassi, R., Schmitt, C., Loewe, A. (2022). Hybrid machine learning to localize atrial flutter substrates using the surface 12-lead electrocardiogram. **Europace**, Volume 24(7), 1186–1194. https://doi.org/10.1093/europace/euab322

22. Nagel, C., Espinosa, B. C.,Gillette, K., Gsell, M., Sánchez Arciniegas, J., Plank, G., Dössel, O., Loewe, A. (2023). Comparison of Propagation Models and Forward Calculation Methods on Cellular, Tissue and Organ Scale Atrial Electrophysiology. **IEEE Transactions on Biomedical Engineering**, Volume 70 (2), 511-522. https://dx.doi.org/10.1109/TBME.2022.3196144

23. Azzolin, L., Nagel, C., Nairn, D,, Sánchez Arciniegas, J,, Zheng, T., Eichenlaub, M.,adidi, A., Dössel, O., Loewe, A. (2021). Automated Framework for the Augmentation of Missing Anatomical Structures and Generation of Personalized Atrial Models from Clinical Data. **Computing in Cardiology.** http://www.cinc.org/archives/2021/pdf/CinC2021-242.pdf

24. Aston, P. J., Mehari, T., Bosnjakovic, A., Harris, P. M., Sundar, A., Williams, S. E., Dössel, O., Loewe, A., Nagel, C., & Strodthoff, N. (2022). Multi-Class ECG Feature Importance Rankings: Cardiologists vs Algorithms. **Computing in Cardiology.** https://doi.org/10.22489/CinC.2022.087

25. Venton, J., Gillette, K., Gsell, M., Loewe, A., Nagel, C., Winkler, B., & Wright, L. (2022) Sensitivity Analysis of Electrocardiogram Features to Computational Model Input Parameters. **Computing in Cardiology.** https://doi.org/10.22489/CinC.2022.024

26. Mehari, T., and Strodthoff, N., (2022). Self-supervised representation learning from 12-lead ECG data. **Computers in Biology and Medicine**. Volume141. https://doi.org/10.1016/j.compbiomed.2021.105114

27. Strodthoff, N., Mehari, T., Nagel, C., Aston, P. J., Sundar, A., Graff C., Kanters, J. K., Haverkamp, W., Dössel, O., Loewe, A., Bär, M., Schäffter, T. (2023). PTB-XL+, a comprehensive electrocardiographic feature dataset. **Scientific Data**. Volume 10, Article Number 279. https://doi.org/10.1038/s41597-020-0495-6

28. Mehari, T, and Strodthoff, N. (2022). Advancing the State-of-the-Art for ECG Analysis through Structured State Space Models. **ML4H (Machine Learning for Health),** https://arxiv.org/abs/2211.07579

29. Mehari, T, Sundar, A, Bosnjakovic, A, Harris, P, Williams, SE, Loewe, A, Dössel, O, Nagel, C, Strodthoff, N, and Aston, PJ (2023). ECG Feature Importance Rankings: Cardiologists vs. Algorithms, **Medical Physics.** https://doi.org/10.48550/arXiv.2304.02577

30. K. Gillette, M. A. Gsell, A. J. Prassl, G. Plank (2021). Influence of Electrode Placement on the Morphology of In Silico 12 Lead Electrocardiograms. **Computing in Cardiology**. https://doi.org/10.5281/zenodo.7941047

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

<table>
<tr><td colspan="2">Project start date and duration:</td><td>July 1st, 2019, 39 months</td></tr>
<tr><td colspan="3">Coordinator: Markus Bär, PTB Tel: +49 30 3481 7687 E-mail: markus.baer@ptb.de
Project website address: https://www.ptb.de/empir2019/medalcare</td></tr>
<tr><td>Internal Funded Partners:
1. PTB, Germany
2. IMBiH, Bosnia and Herzegovina
3. LNE, France
4. NPL, United Kingdom</td><td>External Funded Partners:
5. A-A, United Kingdom
6. KCL, United Kingdom
7. KIT, Germany
8. MUG, Austria
9. TUB, Germany</td><td>Unfunded Partners:
10. FhG, Germany</td></tr>
<tr><td colspan="3">RMG: -</td></tr>
</table>