FINAL PUBLISHABLE REPORT

Grant Agreement number	18HLT07
Project short name	MedalCare
Project full title	Metrology of automated data analysis for cardiac arrhythmia management

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

TABLE OF CONTENTS

1 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.

2 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.

3 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).

4 Results

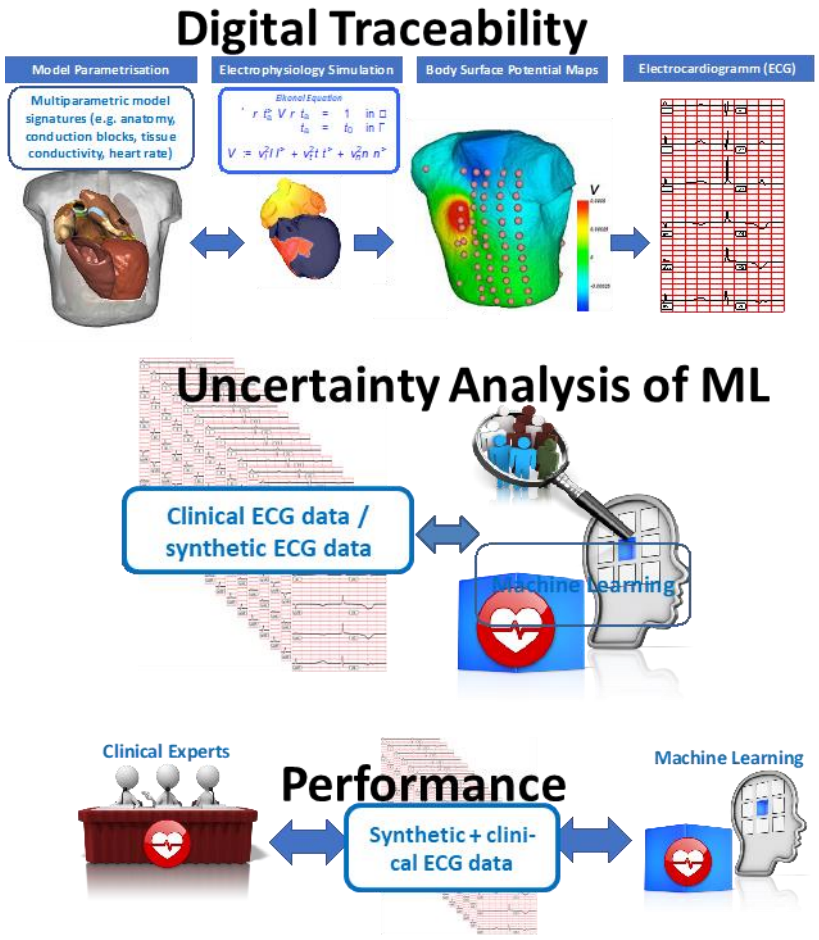


Figure 1: *MedalCare project outline: WP1 - Model pipeline that combines representative anatomical information with biophysical modelling and uncertainty quantification to obtain a database with synthetic simulated ECGs derived from a virtual cohort of models (top panel); WP2 – Benchmarking and quantitative evaluation of ML algorithm for automated ECG diagnosis (middle panel) based on clinical ECG database(s); WP3 – Validation of synthetic ECGs from WP1 by clinical experts, “clinical Turing” test by comparison of appearance of synthetic and clinical ECGs and assessment of performance of ML algorithm in comparison to clinical expert (bottom panel).*

4.1 Objective 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.

KIT and MUG supported by A-A, KCL, PTB and TUB have generated a reference database of synthetic ECG signals (called 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 [11] and in one version of the data, realistic noise was added to the synthetic ECGs [6]. 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 is documented in technical report https://doi.org/10.48550/arXiv.2211.15997). All modeling parameters as well as the data structure of the synthetic ECGs were 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.

A sketch of the modelling pipeline, anatomies used in the simulations and examples of simulated ECGs can be see Figures 2, 3, and 4. Figure 2 shows that the pipeline exists of two modules for the simulation of the atrial and the ventricular part of the ECG signal, i.e. two independent pipelines were used for the modelling of the atrial and ventricular contribution, respectively.

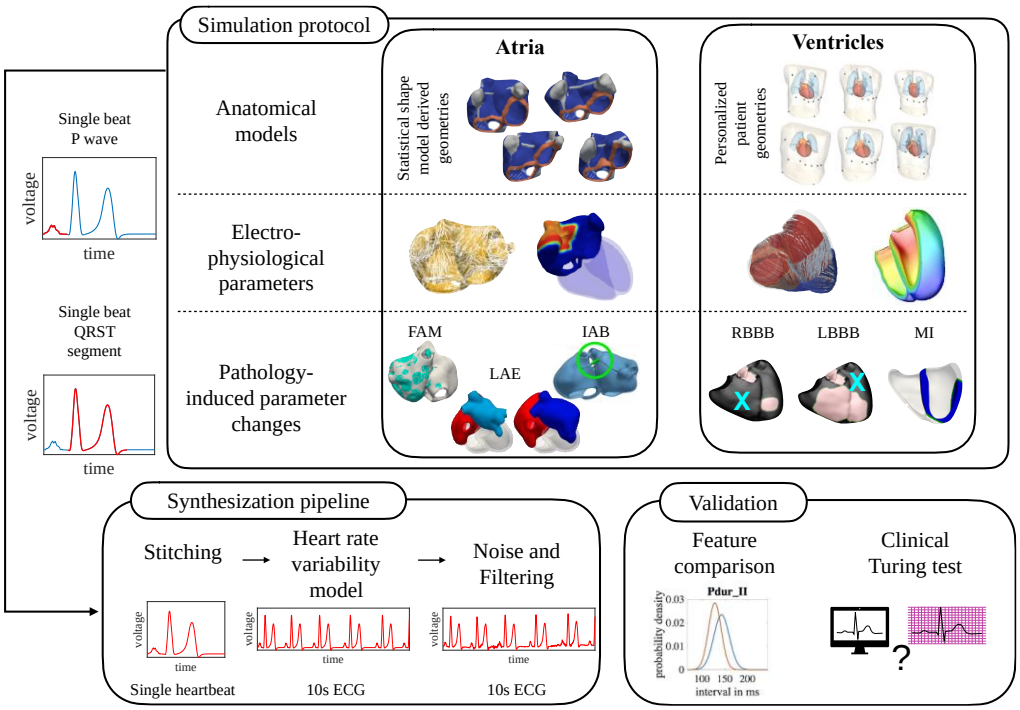


Figure 2: *Sketch of modelling pipeline to generate 12 lead ECGs with two separate atrial and ventricular models.*

The two resulting separate signals were then stitched together to obtain the clean simulated ECG signal. The pipeline also allows superposition of the signals by noise that is typical for clinical ECG recordings.

Figure 3 shows the anatomical and physiological elements underlying the modelling of the atria and ventricles. For both cases the anatomy of the heart was varied with respect to shape, size and position in order to achieve a representative sample for the virtual ECG “cohort”. For simulating ventricular ECGs, a virtual cohort of heart geometries and torsos was created from MRI data of 13 healthy volunteers [14]. The simulation of atrial signals was carried out with statistical shape models of the atria and the torso of humans [3,9]. Advanced methods were also 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). Pathologies of the atria and ventricles such as conduction blocks or fibrosis were implemented by using local changes of the physiological parameters of the used biophysical models.

Finally, Figure 4 displays examples of simulated ECGs ranging from a healthy control to pathologies like conduction blocks and myocardial infarctions. The underlying computational ECG generation included a realistic representation of variability in anatomy and physiology and a detailed model of the Purkinje system [14, 16]. The project has also assessed parameter sensitivities and the uncertainty of different model formulations with varying degrees of complexity used for simulation with different methodologies (conduction velocities [20] and full bidomain, monodomain and Eikonal simulations, [22]).

The project has assessed parameter sensitivities and the uncertainty of different model formulations with different methodologies, namely variability of ECG features [10], computation of Sobol coefficients based on ECG features [25] or on an approximate model of the full ECG signal based on polynomial chaos [20]. More specifically, two independent approaches were used for the modelling of 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 [14].
2. The simulation of atrial signals was carried out with statistical shape models humans [8].

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) estimation of the degree of atrial fibrosis [9,19], (ii) the localisation of atrial flutter substrates [21], (iii) generation of personalised atrial models [23], and (iv) the effect of electrode placement on ECG signals [17].

The simulated data 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 [2] and collected in the PTB-XL+ database ([31], see Figure 5). To be able to do this, a software tool was developed to extract features from ECGs (ECGDeli) [7]. It was then used to specify the variation of ECGs features in clinical data sets such as PTB- XL as well as in the simulated database. Comparison of extracted timing and amplitude features demonstrated similar characteristics [3,14]. These outcomes demonstrate that the project’s novel dataset of simulated signals can be used in the future to extend clinical input data for ML algorithms thus leading to more reliable analyses and improved classification outcomes.

Overall, qualitative agreement was found for the most important ECG features. For a substantial subset, a quantitative match of the synthetic MedalCare-XL and the clinical PTB- XL database was also observed and hence this data can be used in the future for the development and testing of ML algorithms for CVD similar to what was done e. g. in [26].

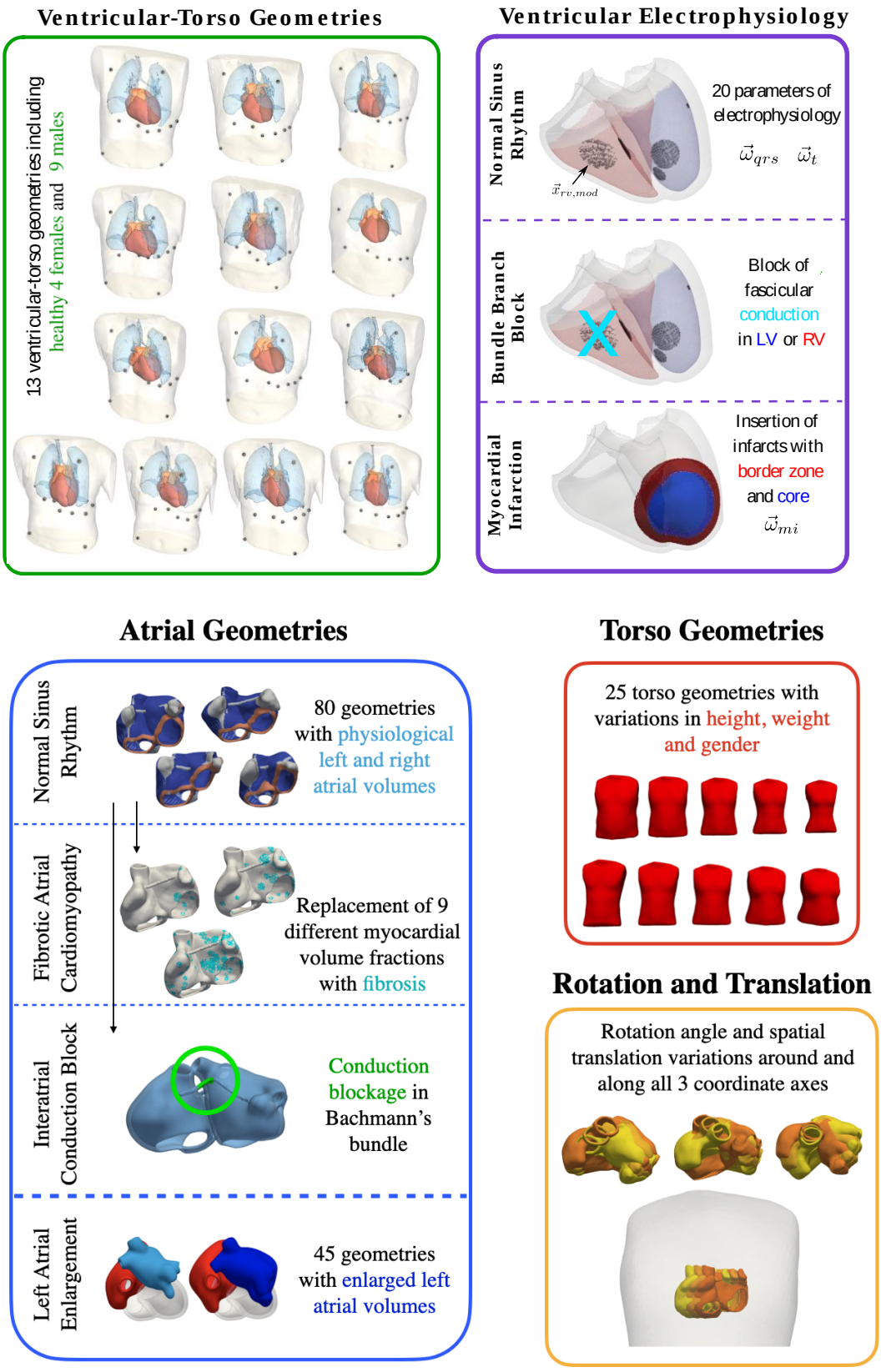


Figure 3: *Ventricular (upper) and atrial (lower) models used to generate the virtual database of 12 lead ECGs.*

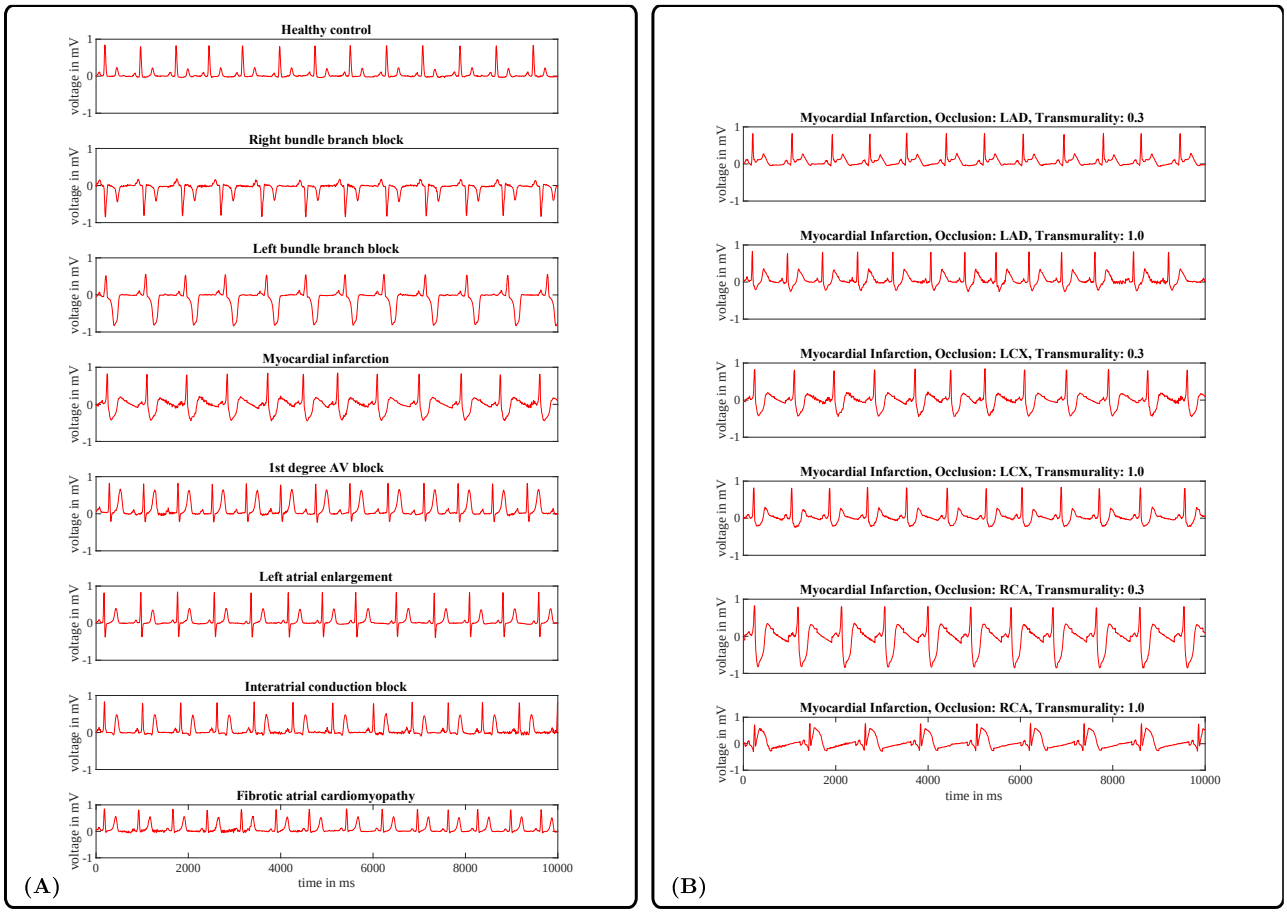


Figure 4: *Examples for one lead of the 12 lead ECG of various simulated ECGs for healthy control and selected pathologies (A). Examples for the 6 pathological subclasses of myocardial infarction are also shown (B).*

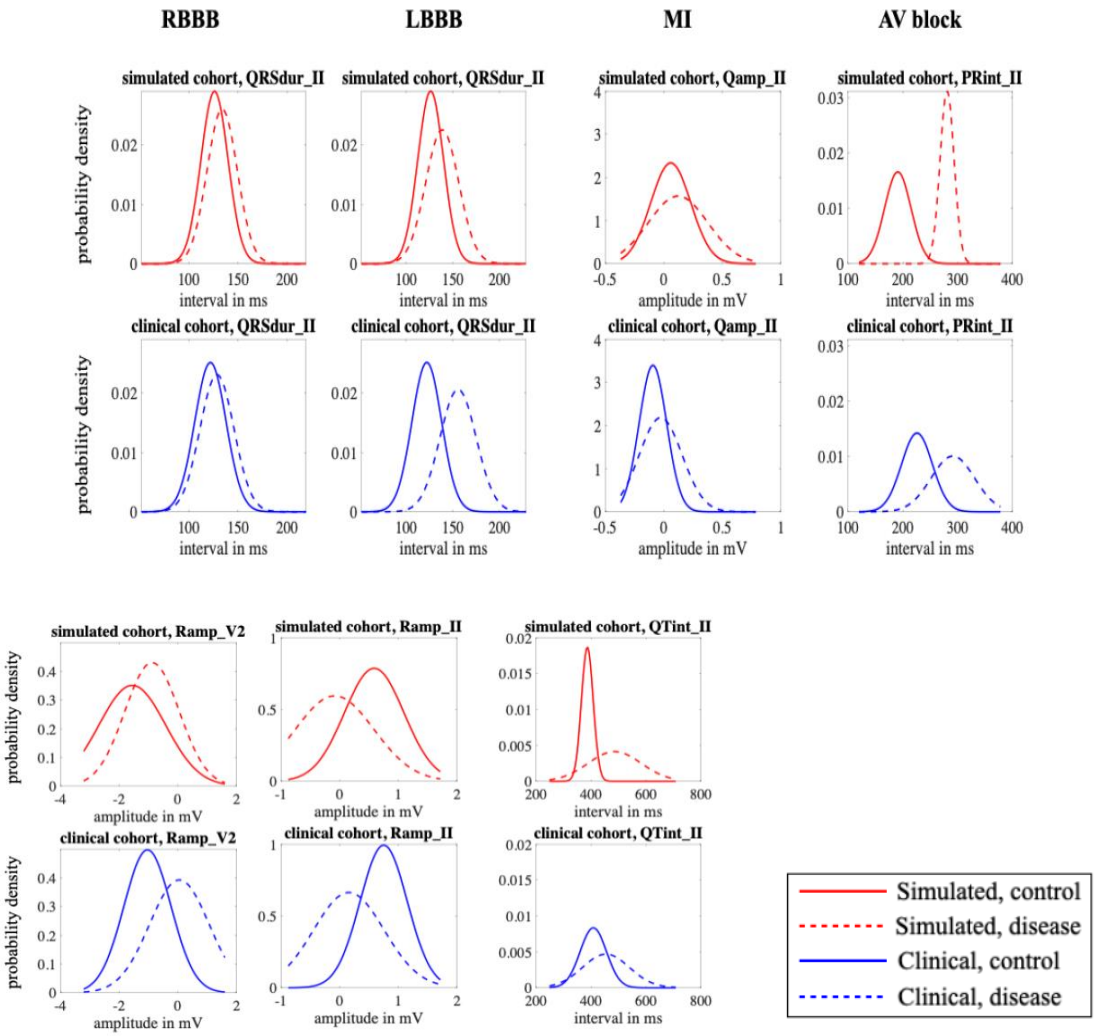


Figure 5: *Validation results by comparison of simulated and clinically recorded ECGs for selected pathologies and ECG features as indicated. Distribution of features for the respective pathologies are given as dashed lines. Full lines indicate the healthy control case as comparison.*

4.2 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.

NPL and PTB supported by LNE, KIT and MUG selected two methods for sensitivity analysis and then implementation as part of the project. One method is based on feature extraction and relates changes in input parameters of biophysical models to changes in the ECG features [3,28]. 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 [20]. 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 becomes 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 (see Figure 6). The sensitivity of these quantities to the inputs in the underlying models has been explored by the project. In this work, two methods of numerical calculation of Sobol indices were compared and yielded comparable results, see Figure 6 and [28]. The interpretation of sensitivity analysis results together with modeling validated the project’s obtained results and improved the understanding of the underlying model structure. For example, in order to create realistic ECG signals, (e.g. with parameter and signal distributions comparable to a cohort of healthy patients), understanding how the input parameters of the models affect the overall shape of the signal provided valuable information to the modellers and enabled them to select appropriate parameter ranges.

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, which were (i) electrode movement, (ii) motion artefact and (iii) baseline wander. 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. [6,13].

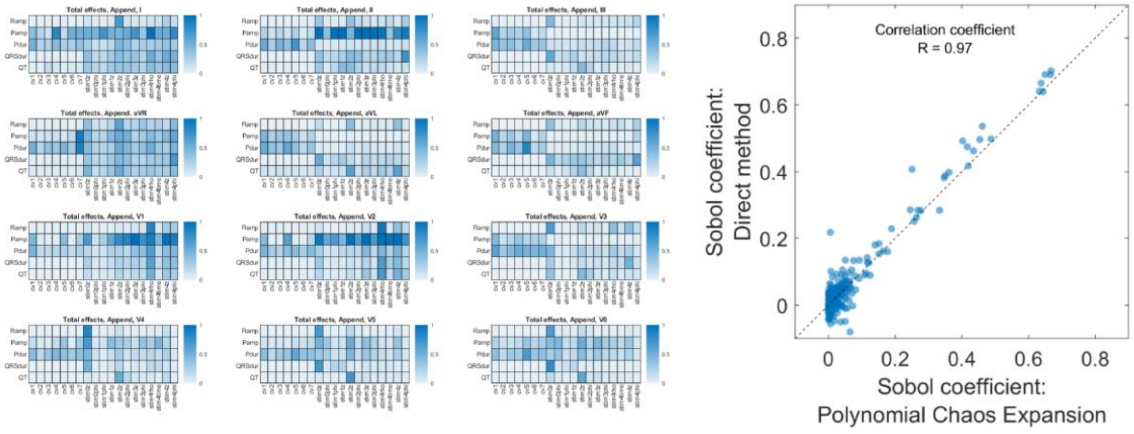


Figure 6: *Heatmaps of Sobol coefficients associated with different model parameter evaluated directly from ECG features (left) and comparison of Sobol coefficients obtained from direct feature-based method with Sobol coefficient obtained from polynomial chaos expansion based on full ECG data*

The activities to explore the sensitivity of synthetic ECGs to input parameters were performed on computational models, of detailed simulations of the creation and propagation of electrical signals during a heartbeat. Variance-based sensitivity analysis decomposes the observed variations of the signal into parts caused by the different model parameters via Sobol indices, which can be calculated numerically from stochastic sampling

schemes on defined parameter distributions. 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.

Figure 7 shows an example of sensitivity analysis and uncertainty quantification performed on virtual ECG data of the atria signal under the variation of the relative position and orientation of the atria inside the respective torso model. Panels a) and b) show the signal variations in leads I and V1, with lead V1 exhibiting the largest overall signal variance. Panels c) and d) show the respective sensitivities of the signal at each point in time with respect to shifts (S3-S5) and orientation angles (S6-S8). The depicted uncertainties correspond to an upper bound estimation from the surrogate error. In panels e) and f), the quick convergence of PCE surrogates with respect to the statistical properties of the data is shown. Panel e): very basic PCEs (green) reconstruct the signal distribution very well (lead V1, at t=55 ms in black), but more involved surrogates increase the accuracy further (red).

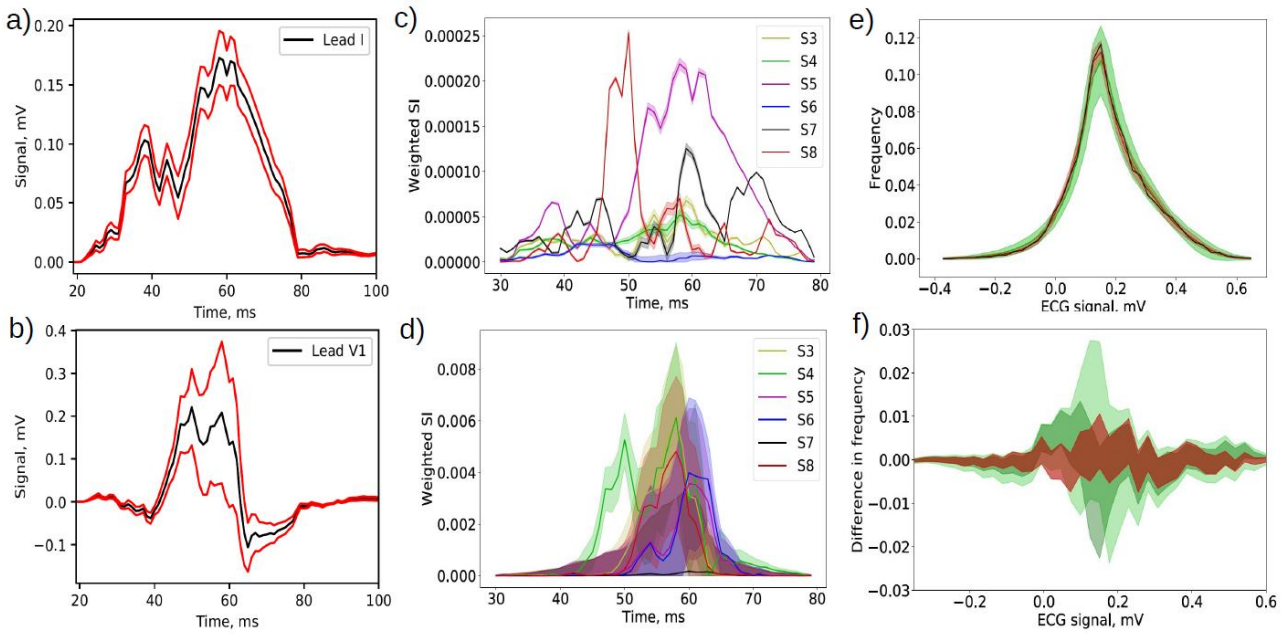


Figure 7: *Sensitivity Analysis and Uncertainty quantification via PCE surrogate model. The virtual ECG signal exhibits variations due to defined input parameter variations during the sampling process, cf. a) and b) for the mean signal in black and the range of one standard deviation in red. c) and d): The sensitivity of the signal from the different input parameters, corresponding to Sobol indices Si(t) as obtained via PCE surrogate modeling. The error ranges are upper bounds obtained from the surrogate model error. Parameters 1 and 2 do not influence the signal in this case. e) Uncertainty quantification via PCE model is improved significantly due to the fast surrogate evaluations. Basic PCE models reflect the statistical distributions (here the V1 signal at t=55 ms) already quite well. The sampled distribution (black) is compared to a PCE with 100 samples and 2nd order reconstruction (green) and the surrogate from 9900 samples in 4th order of reconstruction (red). f) The respective differences of the signal value distributions by subtracting the mean in panel e).*

4.3 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 aim of this work was to explore aspects of ML and the uncertainty in ML predictions arising from many different sources. Several different types of ML were considered.

PTB together with FhG has created and the PTB- XL Database (see https://doi.org/10.13026/x4td-x982 and associated paper https://doi.org/10.1038/s41597-020-0495-6) by formatting and transferring existing ECG records into a user-friendly open source format. This database consists of 21,801 12 lead 10 second clinical ECG signals [2] which is widely used for ML studies, see e. g. the benchmarking of algorithms in FhG, TUB and PTB´s joint study [4]. Inside the project also generated a second dataset, PTB- XL+, where widely used clinical ECG features were extracted with three different conventional ECG analysis software packages (see https://doi.org/10.5281/zenodo.7817567 and associated paper in [28]):

(1) General Electric's Marquette 12SL, which is widely used in hospitals,

(2) the University of Glasgow ECG analysis program, that is also well accepted and used in clinical and research environments,

(3) and the open source ECGdeli software that was developed by partner KIT in Objective 1.

Non-standard ECG features have also been considered using PTB-XL+ and used for benchmarking. The PTB-XL+ database is also available on Zenodo (see https://doi.org/10.5281/zenodo.7817567 and associated paper in [28]).

NPL together with IMBiH, LNE and TUB constructed several datasets of 12 lead ECGs for use with ML from publicly available datasets. These included a small dataset (100 records, 2 classes), a medium-sized dataset (2,863 records, 4 classes), a large dataset (16,272 records, 5 classes) and a more balanced large dataset (10,406 records, 5 classes) where the large Normal class in the large dataset was reduced to a similar size to the other classes. In addition, further datasets were created for specific activities e.g. investigating the scalability of the performance of ML algorithms considering small, medium-size and large datasets.

ML was used for classification of ECG signals using a variety of features or deep learning. In different benchmarking studies, FhG and PTB as well as NPL with FhG, IMBiH, KCL, LNE, KIT, PTB and TUB. found that a 1D convolutional neural network (CNN) which used the raw signals as input generally gave the best results. A support vector machine (SVM) on ECG interval and amplitude features also gave good results. The results for the medium dataset are shown in Figure 8.

The synthetic ECG signals generated in Objective 1 and summarised in the MedalCare-XL data set were used for ML. NPL created two identical datasets, one consisting of selected real signals from PTB-XL and the other one made of the corresponding synthetic signals MedalCare-XL. The results obtained where the training and the test data were synthetic were better than where both training and test data were real, which is probably due to the noisy nature of real signals. Training on synthetic data and testing on real data gave very poor results while training on real data and testing on synthetic data gave slightly better results. Finally, training on the real and synthetic data combined gave very similar results for a real data test set as training on real data only. This showed that the synthetic data is still sufficiently different from the real data so that classification of real signals does not currently benefit from the synthetic data. A summary of results is shown in Figure 9.

This procedure provided an alternative approach to the validation of the synthetic database from Objective 1 with established clinical databases. Notably it also showed that the benchmark results for the different algorithms relative to each other are similar for synthetic and clinical data sets. The performance indicator F1 score gave a better result for the synthetic data than for the measured clinical data pointing towards better quality of synthetic data.

However, algorithms trained with synthetic data did not perform reliable anymore in test with clinical data and vice versa. If synthetic data were added to the real training data, the performance also became worse.

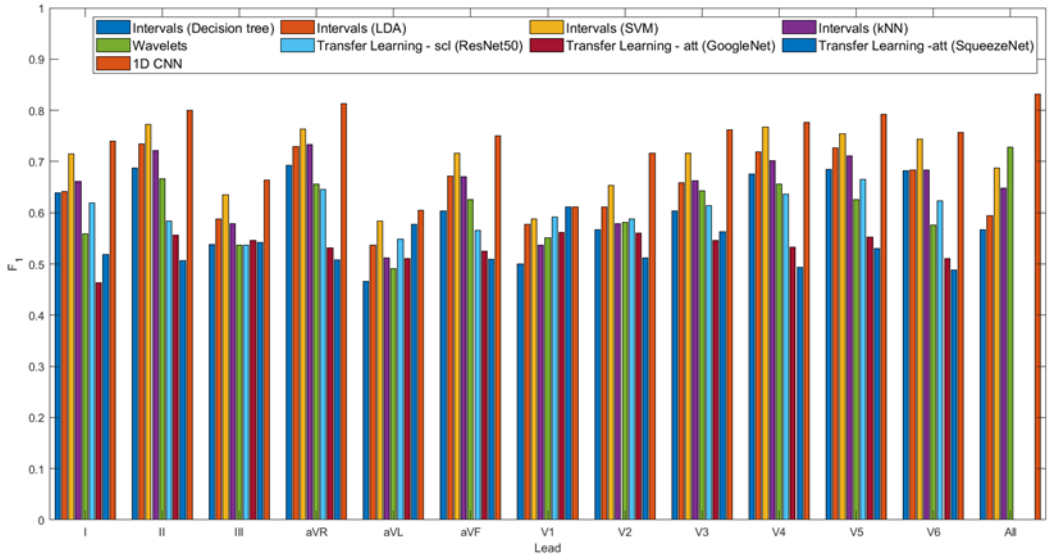


Figure 8: *Typical benchmark result for nine different ECG classification algorithms for a medium-size data set Summary of results for different methods on the medium dataset for each of the 12 leads*

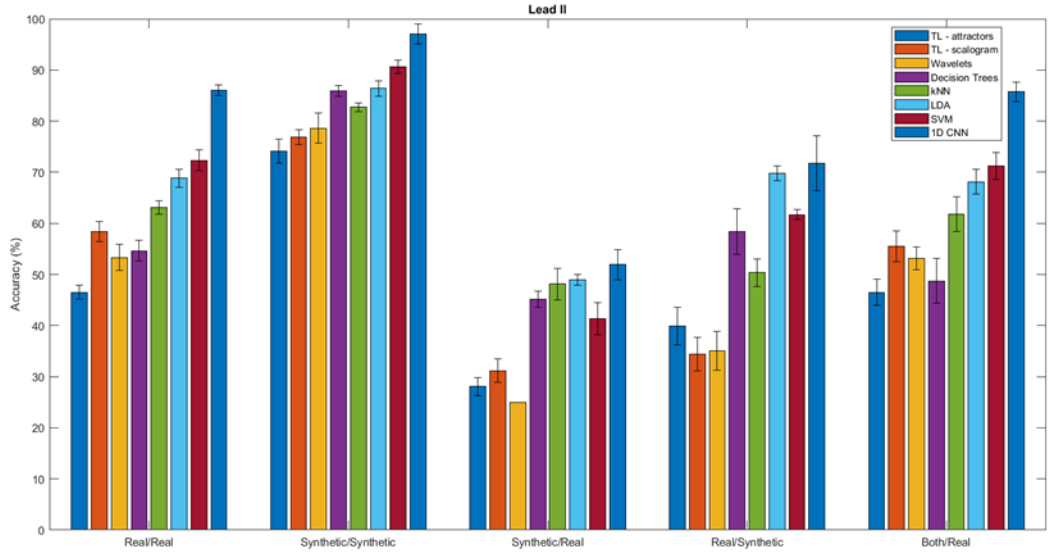


Figure 9: *Results on accuracies for various ML methods applied to different combinations of the real and synthetic data of a single lead (Lead II). Benchmarking and comparison of a clinical and synthetic medium-size dataset. The two left-most columns compare benchmark results (F1 scores) obtained for seven selected algorithms with clinical (real) data (first block from left) and (second block from left). The third and fourth block from left show cross-validation where the algorithm were trained with synthetic data and evaluated with real data (and vice versa). The rightmost block shows results were a combination of synthetic and real clinical data was used for training and testing was performed with clinical data only.*

In addition to this, FhG, NPL, PTB and TUB investigated how noise on the ECG 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. This is illustrated in Figure 10 where F1 scores for a neural network trained on either clean or noisy is tested on data with varying levels of noise [6].

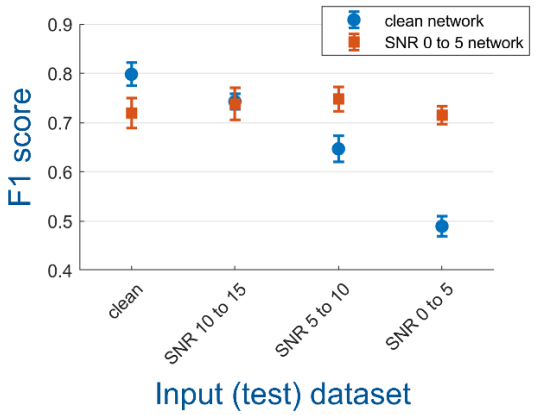


Figure 10: *Results for a neural network trained on either clean or noisy signals and tested on signals with different noise levels.*

Feature importance methods give a ranking of features according to importance for a given classification task. There are many such methods, but they often give very different feature rankings which are inherently difficult to evaluate due to a lack of ground truth. 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) byNPL together with A-A, FhG, IMBiH, KIT, KCL, PTB and TUB. 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 and showed how different methods worked for different ECG classes, see [27,30].

Several interpretability methods were studied by NPL as well as FhG, PTB and TUB to determine what was important for a deep learning classification. LIME and Layerwise Relevance Propagation were applied to Symmetric Projection Attractor Reconstruction (SPAR) images to show particular regions of interest. Integrated Gradients were used with the 1D CNN to show which regions of the signal were important for the classification, as illustrated in Figure 11. Integrated Gradients were also used for the classification of signals as either real or synthetic and to highlight the regions which were important for this classification. This then provided useful input to the computational modellers in Objective 1 as it showed where the synthetic ECGs were different from real ECGs.

The effect of variations in the training data was also studied. The project showed that the Inception1D CNN model is particularly sensitive to the sampling frequency of the signals and gave significantly better results with a lower sampling frequency (see Figure 12). A wavelet model also had slightly better performance with a lower sampling frequency. Both methods performed better when there were more records per patient. 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. It was found that the training data has 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 both uniform and structured label noise on model performance was considered, where structured label noise assumes that human annotators are more likely to confuse labels corresponding to closely related conditions. The project showed that prediction of some labels can be sensitive to noise whereas other labels can be insensitive to noise. This effect was shown to be 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 (see Figure 13) but uniform noise is more harmful than structured noise at all levels.

Further to this the project investigated transfer learning on SPAR attractor and scalogram images. The effect on performance was considered in terms of (i) balancing the dataset in different ways, (ii) Bayesian optimisation of network hyperparameters, (iii) finetuning the networks with Bayesian optimisation and (iv) combining automatically extracted features from the images with additional manually derived features.

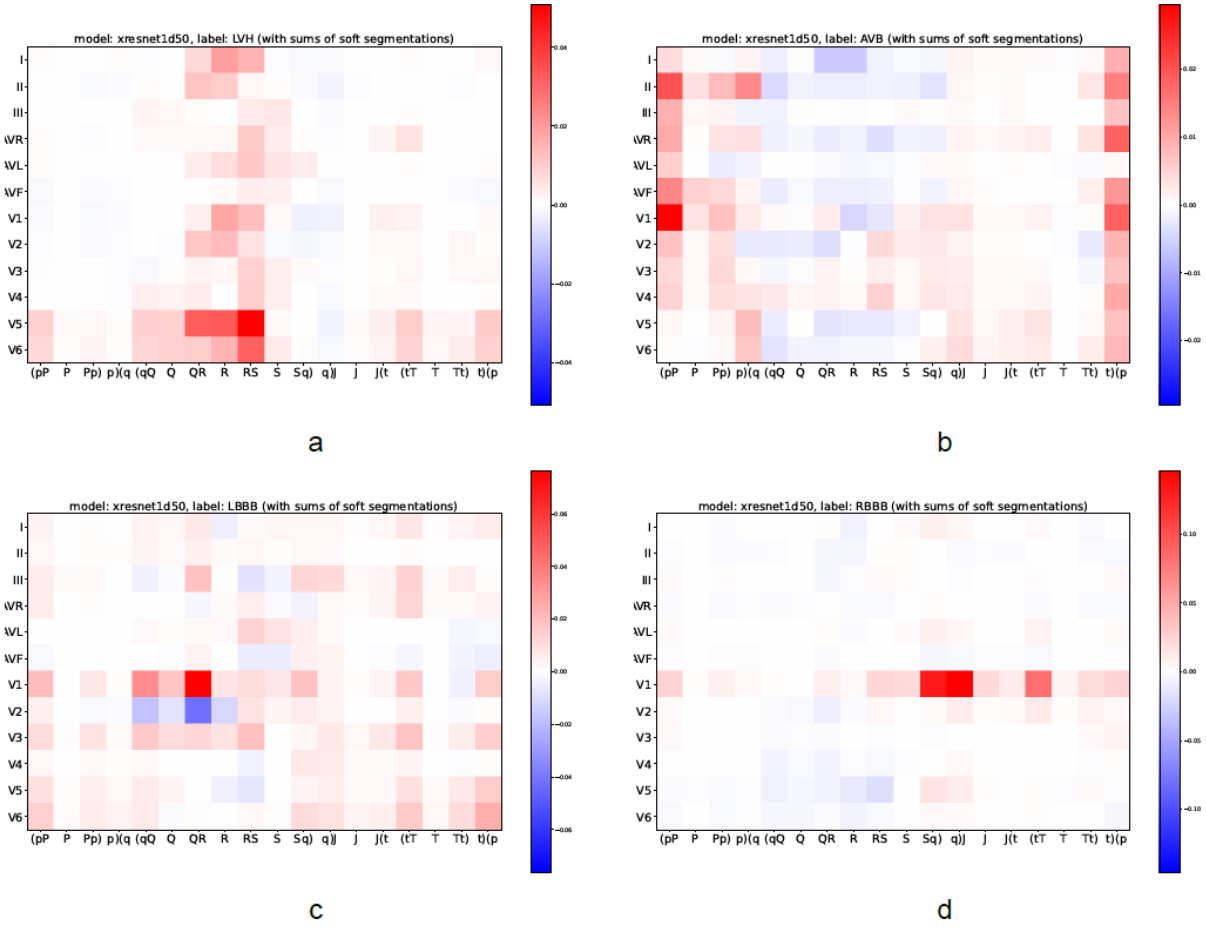


Figure 11: *Heatmaps obtained using an xresnetd150 model with ECG features against lead for (a) left ventricular hypertrophy (b) atrioventricular block (c) left bundle branch block (d) right bundle branch block.*

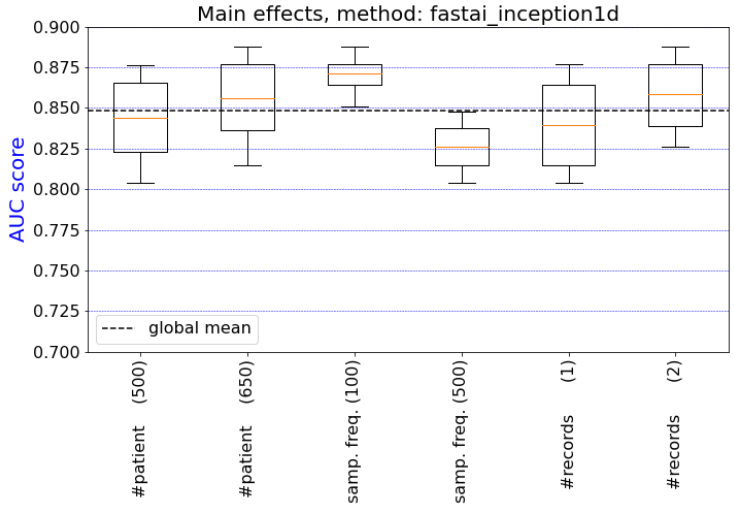


Figure 12: *Boxplot of the main effects resulting from variations in the training data for the Inception1D CNN model.*

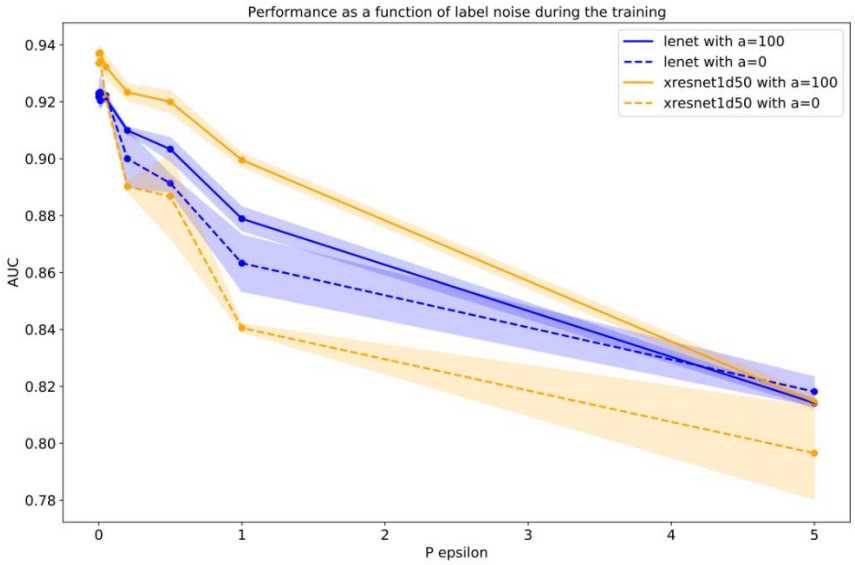


Figure 13: *The performance of the xresnet1d50 and lenet models both decrease with increasing label noise.*

4.4 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".

This objective aimed to evaluate the capability of the generated signals in the synthetic ECG database in Objective 1 to be useful for clinical applications, as well as to improve automated ECG classification and detection of cardiac arrhythmias using ML approaches as described in Objective 3. Towards this end, MUG in cooperation with KCL, KIT and MUG carried out an analysis on clinical misdiagnosis by further comparing the actual pathology of ECG signals (both clinical and simulated) with the reported pathology from clinicians through means of clinical Turing tests. This analysis aimed to better understand which pathologies and pathological subclasses could be benefitted by ML algorithms. We also demonstrated correspondence between the measured 12 lead ECG and activation maps in an invasive experimental setup with a simulated replica to ensure model integrity and performance for assisting in clinical understanding and EP procedures. Finally, we have compared the outcome of the pathology classification done by cardiologists with the predictive capabilities of machine learning models.

These clinical Turing tests provided valuable feedback to the project for improving the underlying biophysical models (i.e. modelling of electrophysiology). In particular, iterative improvements in model parameters were used to ensure that specific morphological features that are diagnostically relevant in 12 lead ECG were accounted for.

Figure 14: *Final Initial prototype of the online survey platform used for performing clinical Turing tests. The platform was developed in coordination with the Know Center in Graz, Austria and supports annotation and remote feedback of clinical ECG signals by cardiologists. Three buttons (zoomed area) are displayed in the top row to label the signal as measured or generated, or to give no statement.*

This work also provided an assessment of how realistic the synthetic ECG data from the database in Objective 1 appears to medically trained cardiologists. To enable this partner MUG generated an online survey platform through which clinical experts could conveniently look at synthetic and clinical ECGs and provide their judgment on origin of the ECG or their diagnosis, for an illustration of the survey platform, see Figure 14. Within

the clinical Turing test performed for healthy control, it can be observed that accuracy in identifying whether a signal was simulated or clinical was 77 % accurate. Within the pathological clinical Turing test this increased to 83 %. Pathological ECGs were diagnosed correctly by the two expert cardiologists in 51 % of cases for both simulated and measured signals. The overall validation results for the synthetic database in Objective 1 were hence successful and satisfactory, but also indicated possibilities for further refinements.

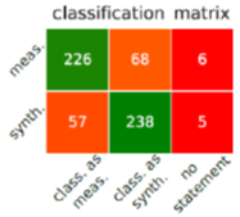
Further to this, the project conducted a study of a personalised ECG [15] to ascertain whether the underlying model of cardiac electrophysiology is capable of representing the intrinsic nature of the heart, not only in terms of the 12 lead ECGs, but in all forms of non-invasive clinical data such as electro-anatomical maps (EAMs) and body surface potential maps (BSPMs). Another crucial part of this objective was the validation of synthetic ECGs with measured clinical data. For this, a simulation replica was made of an experimental study 1 (see Figure 16) using the same modelling pipeline used to generate the data for the ventricles in Objective 1..To do this, the cardiac sources in the form of activation maps and the 12 lead ECG were compared to assess the capabilities of the simulated model to match experimental reality (see Figure 17). . Activation maps were computed within the experimental data collected from the sock and needle electrodes using an open-source toolbox for ECG feature extraction. A 12 lead ECG was then constructed for both the simulated and measured data from the torso electrodes taken from clinical guidelines and comparison in QRS morphology of the 12 lead ECG was made revealing good agreement but also notable differences (see Figure 16). Primary differences in activation occurred in close proximity to the pacing site and at the base of the ventricles.

Overall, the project 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.

A summary of the most important aspects can be seen in Figure 15. Overall, the simulated normal healthy reference ECGs were more difficult to distinguish from real clinical ECGs than the pathological ones. The diagnostic success rate dropped from around 60 % for clinical ECG data from PTB- XL (Objective 3) to around 40 % for synthetic ECG database MedalCare-XL (Objective 1). Both values were much lower than the best performance of ML algorithm (see Objective 3), which could be between 80 % and 95 % for the best forming algorithms (usually neural networks).

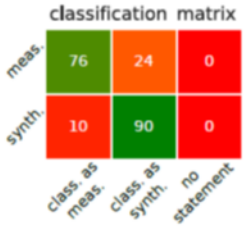
(a)

	meas. class. as meas.	meas. class. as synth.	synth. class. as meas.	synth. class. as synth.	no statement	% correct
Clin.1	44	6	0	50	0	94
Clin.2	26	24	13	37	0	63
Clin.3	43	6	15	31	5	74
Clin.4	47	3	16	34	0	81
Clin.5	22	23	6	43	6	65
Clin.6	44	6	7	43	0	87



(b)

	meas. class. as meas.	meas. class. as synth.	synth. class. as meas.	synth. class. as synth.	no statement	% correct
Clin.1	37	13	5	45	0	82
Clin.2	39	11	5	45	0	84



(c)

	meas. with corr. path.	meas. with wrong. path.	synth. with corr. path.	synth. with wrong. path.	no statement	% corre
Clin.1	34	16	18	32	0	52
Clin.2	28	22	21	29	0	48

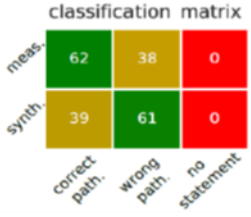


Figure 15: *Results of assessment of synthetic and clinical ECG data taken from the MedalCare-XL and PTB- XL databases, respectively. (a) Classification of unknown ECGs corresponding to healthy (normal) cardiac function as synthetic or clinically recorded ("clinical Turing test"), shown are scores of six cardiologists; (b) classification of unknown ECGs corresponding to selected pathologies as synthetic or clinically recorded ("clinical Turing test"), shown are scores of two cardiologists; (c) outcome of diagnostic classification of a selection of pathologies. Both validation approaches allow definitions of quantitative measures of the quality of the synthetic database.*

Another crucial part of Objective 4 was the validation of synthetic ECGs with measured clinical data. For this, a simulation replica was made of an experimental study 1 (see Figure 16) using the same modeling pipeline used to generate the data for the ventricles in Objective 1. To do this, the cardiac sources in the form of activation maps and the 12 lead ECG were compared to assess the capabilities of the simulated model to match experimental reality (see Figure 17). Activation maps were computed within the experimental data collected from the sock and tank electrodes using an open-source toolbox for ECG feature extraction. A 12 lead ECG was then constructed for both the simulated and measured data from the torso electrodes (taken following clinical guidelines) and subsequently comparison in QRS morphology of the 12 lead ECG was made revealing often a good agreement but also notable differences (see Figure 17). Primary differences in activation occur in close proximity to the pacing site and at the base of the ventricles. This comparison nevertheless generally revealled good agreement although notable differences could be seen (see Figure 17).

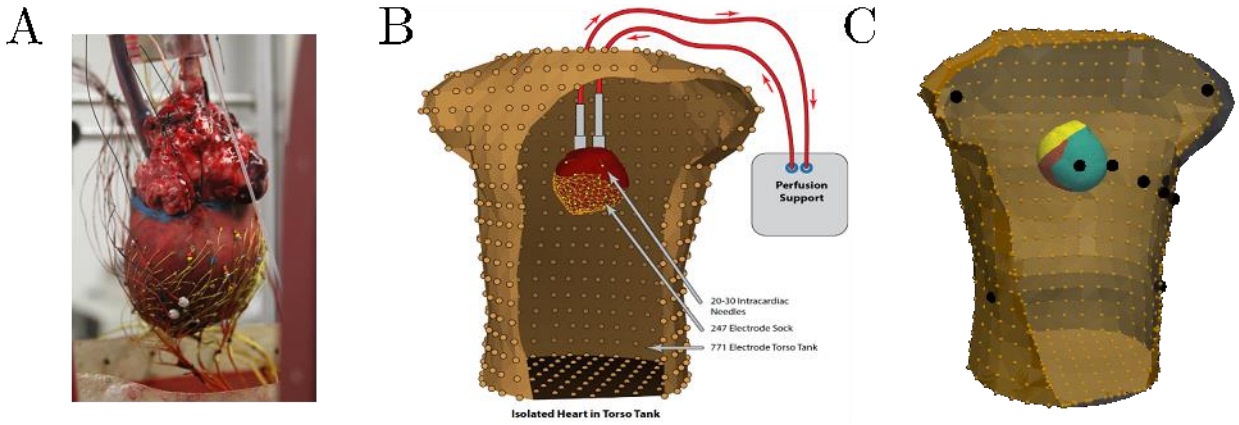


Figure 16: *(A) A canine heart with inserted needles perfused using a Langendorf setup is placed in a sock of electrodes. (B) Illustrative setup of the heart placed within the torso tank with electrodes. (C) Replication of the experimental setup as a cardiac model of EP using model generation pipeline. Manually selected electrodes for 12 lead ECG construction can be visualized in black. All experimental images are provided and generated by the University of Utah.*

Three pathologies (LBBB, RBBB, and 1AVB) were selected based on the outcome of the pathological clinical Turing test to test the capabilities of ML-based approaches to improve clinical diagnosis and classification. ML-based classification was then performed on the compiled ECG dataset consisting of LBBB, RBBB, healthy control, and 1AVB. All methods for performing classification on both clinical and synthetic signals were developed and reported in the [D2] report. Methods were tested using a single lead (lead II) (see Figure 8 in Objective 2), as well as for the entire 12 lead ECG (see Figure 19). All ML methods were also trained and tested on all mixture combinations of simulated and real clinical data to explore performance. In general, the 1D convolutional neural network (CNN) gave the best performance. The highest accuracy of 100 % was achieved using the 1D CNN on all leads when the network was trained and tested on synthetic data. A similar performance of 95 % is attained for the same 1D CNN tested and trained on real clinical data. When the 1D CNN was trained on a mixture of clinical and synthetic data, but tested on pure clinical data, a similar accuracy of 95 % was attained, exceeding clinical performance.

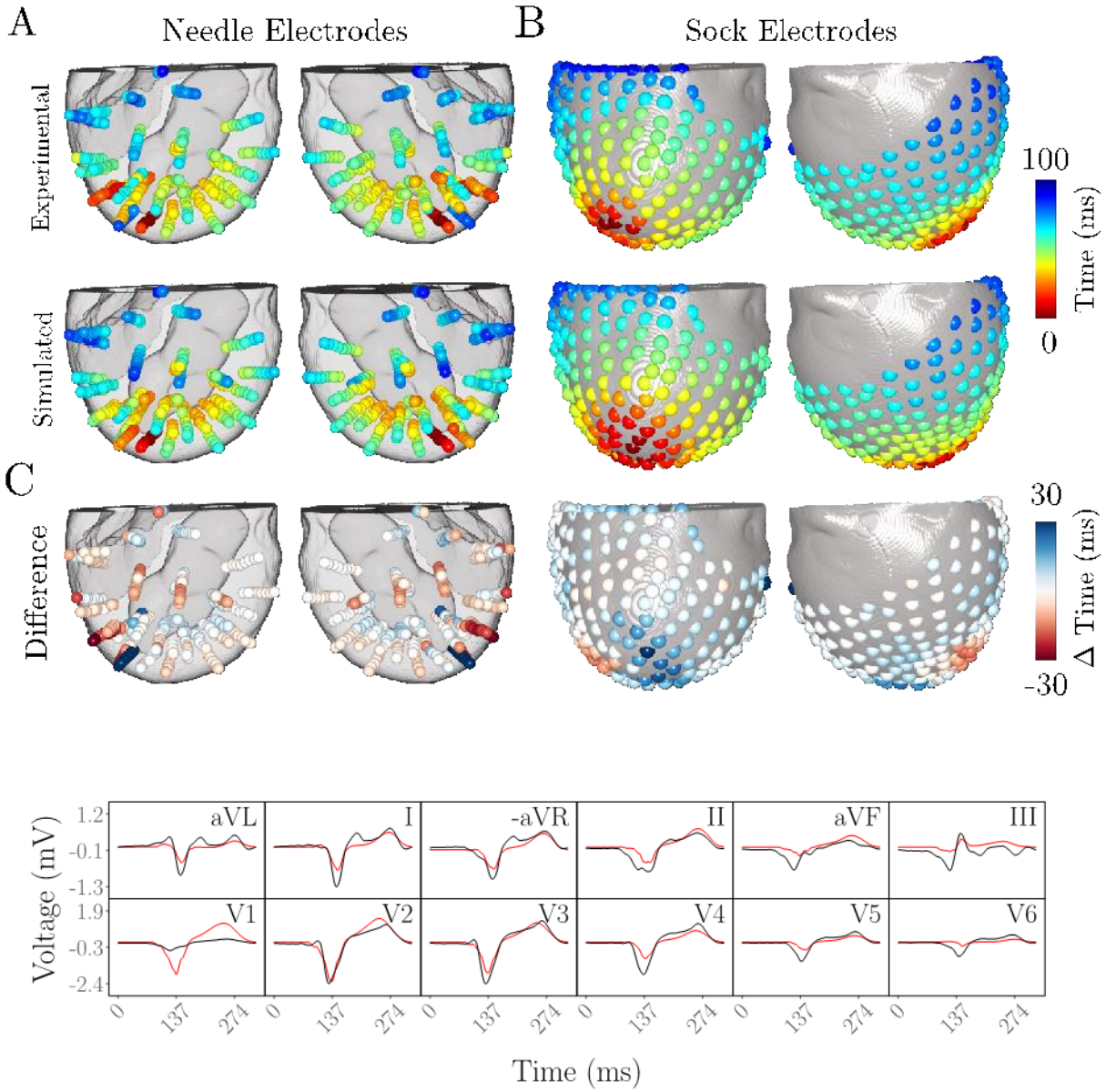


Figure 17: *Comparison of activation maps between experimental and simulated data on both the needle electrodes (A) and sock electrodes (B). Largest differences in activation timings (C) in both the needle (left) and sock (right) electrodes is primarily noticed near the pacing site. For all visualization, front and back orientations are observed. (D) Comparison of 12 lead ECG from the experimental (black) and simulated (red) show correspondence.*

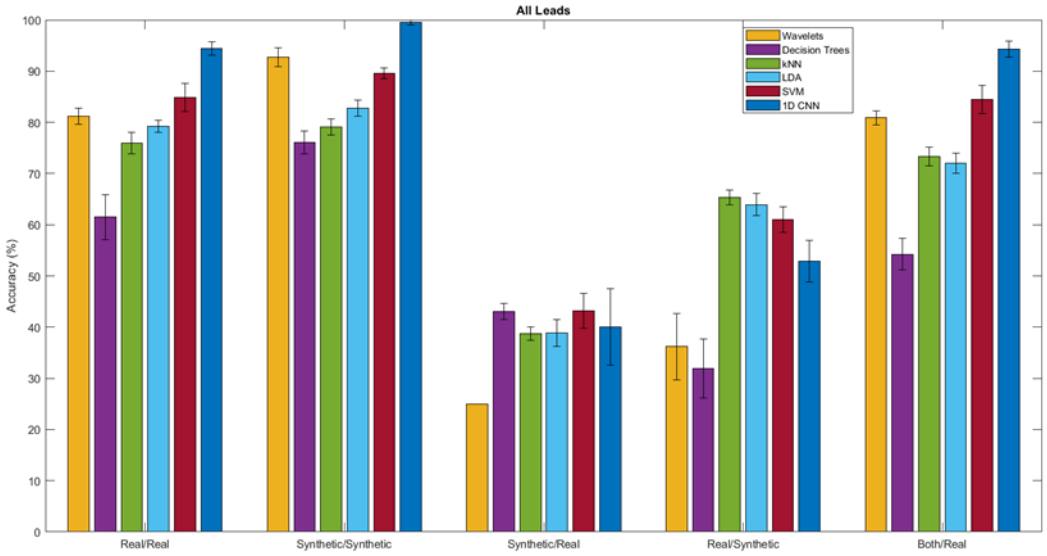


Figure 18: *Pathology classification: ML-based: Accuracy of the various ML methods performed on different mixtures of clinical or synthetic 12 lead ECGs during training or testing. For each histogram partition, the label indicates the training data set on the left and testing data on the right. For example, Real/Real indicates the ML methods were both trained and tested on real data.*

4.5 Summary of Results

The project successfully created a synthetic ECG database and successfully evaluated automated algorithms for ECG analysis and diagnosis with clinical datasets. The project's main findings were:

- reference data sets in the form of validated synthetic databases based on simulation with a "ground truth" and clinical databases based on validation by experts need to be defined and provided. Since ML and AI algorithms were found to be most promising, there is a specific need for large databases (e.g. > 10,000 ECGs for the applications studied in this project).
- Benchmark protocols for AI algorithms that are fit-for-purpose for the given medical task (here ECG analysis) should be developed and agreed. As a first step, this should focus on the overall performance of the algorithm for given tasks, (e.g. classifying ECGs as healthy and normal or determining the specific pathology or arrhythmia for ECGs that are judged as anomalous). Other necessary features that need to be tested are robustness (insensitivity of performance against poor data quality/ low signal-to-noise ratio) and explainability (i.e. identification of the signal aspects that produce a classification by the algorithm).

The criteria for reference data sets and benchmarking protocols are outlined below. A key issue for their development is the availability of open data sets for the training of ML algorithms, such as those developed in this project. Clinical data sets were initially difficult to obtain from public, open sources, thus, this project used previously collect data from PTB.

Reference data sets:

Synthetic databases should contain:

- Clean smooth data from deterministic models
- Noisy data where typical measurement noise is added to the clean data
- Information on the scientific background in the models used and how classes of pathologies were modelled
- Information on the modelling pipeline in the numerical methods, software implementation, and experimental data (e. g. imaging data for anatomies) incorporated in the models
- Information on validation by comparisons to accepted clinical databases
- Information on cross-validation with other accepted synthetic or clinical databases

Clinical databases should contain:

- Sufficiently large sets of identical records, ideally at least several thousand records in a suitable format, for an example see PTB- XL
- Labels for classes and sub-classes including diagnoses for each entry
- Information on measurement parameters e.g length of record, recording devices
- Information on labelling methodology
- Information on cross-validation with other existing clinical data sets e. g. by comparing benchmark results for different algorithms

The validation for both types of databases is crucial. Measures and criteria for the databases must also be defined and agreed.

Important aspects for benchmarking and uncertainty quantification for ML algorithms:

- Definition of a benchmarking protocol including the relevant scores (e. g. AUC, accuracy, F1 etc.) and the amount and types of training and testing data required
- Selection of reference data set(s) (i.e. clinical and/ or synthetic)
- Assignment of error bars to the uncertainty of benchmark results
- Tests for robustness against error sources such as measurement noise, label noise, initialisation of algorithm parameters etc.
- Definition of the requirements for algorithms to be considered acceptable for use for their given purpose
- For algorithms that are designed to support human experts, aspects of explainability are needed

5 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.

6 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/

7 Contact details

Markus Bär, PTB

Tel: +49 30 3481 7687

E-mail: markus.baer@ptb.de