# An Experiment Environment for Definition, Training and Evaluation of Electrocardiogram-Based AI Models

N. Gumpfer[1], J. Prim[1], D. Grün[2], J. Hannig[1], T. Keller[2], and M. Guckert[1,(✉)]

[1]Cognitive Information Systems, Kompetenzzentrum für Informationstechnologie, Technische Hochschule Mittelhessen, 61169 Friedberg, Germany, `michael.guckert@kite.thm.de`

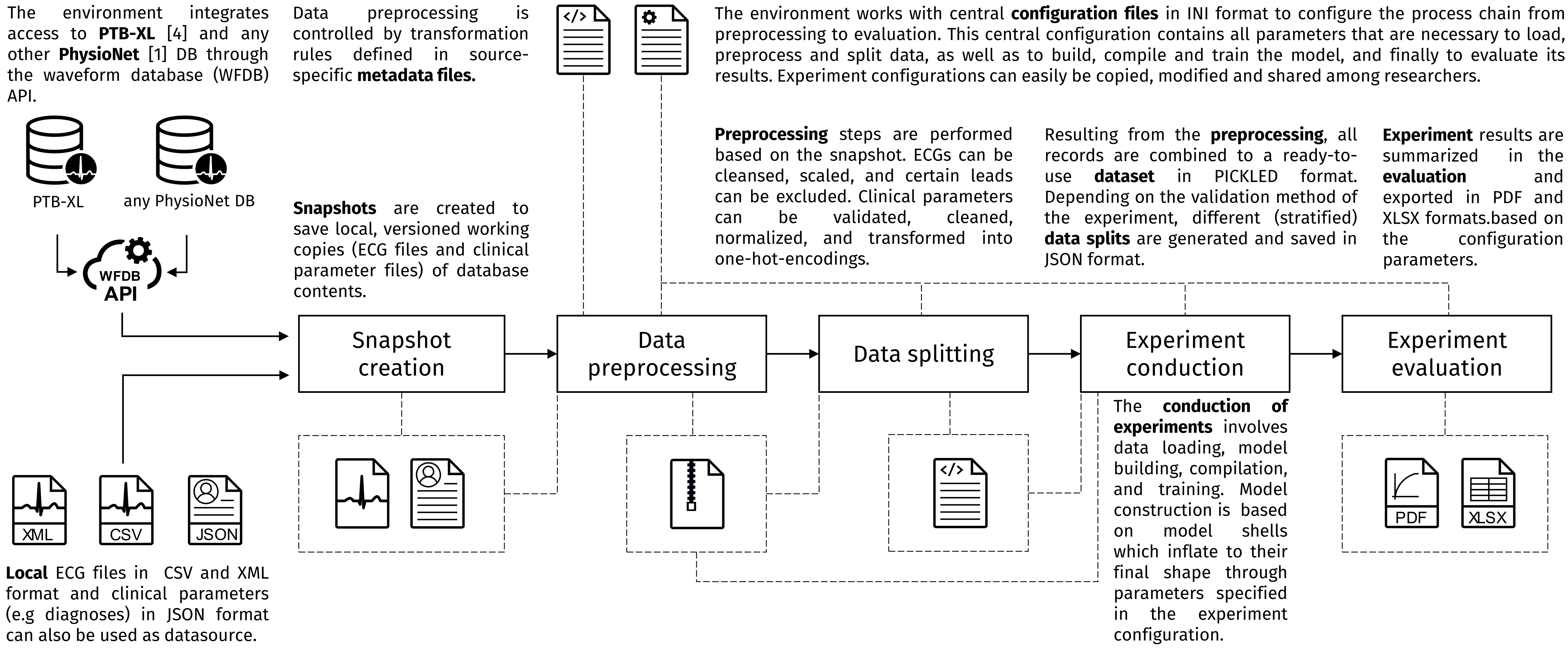
[2]Department of Internal Medicine I, Cardiology, Justus-Liebig-University Gießen, 35390 Gießen, Germany

*Financial support:* Research Campus of Central Hesse (FCMH) via Flexi Funds

## Abstract

The use of artificial intelligence (AI) for analysis of electrocardiogram (ECG) data has recently gained much interest in the AI and medical communities. The discussed models have shown to be able to deliver high diagnostic sensitivity and specificity for detection of various cardiac diseases including rhythm disorders and ischemic events [2, 3]. However, the experiments leading to these results are often difficult to reproduce outside of the original experimental setup and researchers who want to externally validate such results or use them as starting points for new experiments are forced to develop their own models from scratch. We therefore propose a software environment that enables to build, train and evaluate AI models for ECG classification in a reproducible manner and offers sharing of experiment configurations among researchers. The environment further provides simple connection of publicly available data sources of validated ECG recordings. It offers various validation techniques such as bootstrapping and cross-validation. A proof of concept is given for a deep learning model consisting of a convolutional neural network for the classification of acute myocardial infarction based on ECG data.

## Components




## Proof of Concept

As proof of concept (POC), we use a convolutional architecture to classify ECGs with myocardial infarction and healthy controls from PTB-XL [4] and evaluate the performance based on bootstrapping (n=100). The script `download_ptbxl.sh` creates a snapshot from PTB-XL. A configuration file `experiments/ptbxl_poc.ini` contains necessary general hyperparameters (e.g. optimizer, learning rate, loss function) and model-specific parameters (e.g. number of layers, number of neurons). To link all process steps, snapshot and dataset name, metadata file, and splits are defined as well. Further, required metrics and thresholds for evaluation are listed. We only use a subset (I-AVF) of 12 leads. In preparation for the experiment, the data is preprocessed via `python3 preprocessing_runner.py -e ptbxl_poc`. Based on the resulting dataset, 100 different bootstrapping splits are generated via `python3 split_runner.py -e ptbxl_poc`. The training is then started via `python3 experiment_runner.py -e ptbxl_poc`. After training, the evaluation is performed automatically. The achieved performance metrics for AUC, sensitivity and specificity are $0.96 \pm 0.01$, $79.49 \pm 3.35$ %, $98.54 \pm 0.42$ %, respectively.

## Conclusions and Future Work

**Conclusions:** The proposed environment enables researchers to share experiment configurations as well as intermediate results such as datasets. This enables research groups to exchange their data and configurations more easily and to mutually reproduce results. With this environment, we aim to support faster and more densely connected research. The presented concept is currently limited to AI models for ECG classification, but can be extended to other time-series based settings, and with some effort also to image-based concepts.

**Future Work:** Beside general improvements, we plan to extend the environment by methods for explainability and a graphical user-interface.

## References

[1] Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E.: PhysioBank, PhysioToolkit, and PhysioNet. Circulation **101**(23), e215–e220 (2000). https://doi.org/10.1161/01.CIR.101.23.e215

[2] Grün, D., Rudolph, F., Gumpfer, N., Hannig, J., Elsner, L.K., Von Jeinsen, B., Hamm, C.W., Rieth, A., Guckert, M., Keller, T.: Identifying Heart Failure in ECG Data with Artificial Intelligence-a Meta-Analysis. Frontiers in Digital Health **2**, 67 (2020). https://doi.org/10.3389/fdgth.2020.584555

[3] Gumpfer, N., Grün, D., Hannig, J., Keller, T., Guckert, M.: Detecting myocardial scar using electrocardiogram data and deep neural networks. Biological Chemistry (0), 000010151520200169 (03 Oct 2020). https://doi.org/10.1515/hsz-2020-0169

[4] Wagner, P., Strodthoff, N., Bousseljot, R.D., Kreiseler, D., Lunze, F.I., Samek, W., Schaeffter, T.: PTB-XL, a large publicly available electrocardiography dataset. Scientific Data **7**(1), 154 (May 2020). https://doi.org/10.1038/s41597-020-0495-6

## Code Availability

Code for the environment and reproduction of POC results is available from GitHub at https://github.com/nilsgumpfer/experiment-environment-ecg-ai.

THM TECHNISCHE HOCHSCHULE MITTELHESSEN

KITE Kompetenzzentrum für Informationstechnologie

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

biomarker research

FORSCHUNGSCAMPUS MITTELHESSEN