

## Großer maschinenlesbarer EKG-Datensatz für Entwicklung und Test maschineller Lernverfahren veröffentlicht

**(Gemeinsame Pressemitteilung des Fraunhofer Heinrich-Hertz-Instituts, HHI, und der Physikalisch-Technischen Bundesanstalt, PTB) PTB-XL ist der bisher größte öffentlich zugängliche EKG-Datensatz mit mehr als 20 000 Aufnahmen**

Wissenschaftlerinnen und Wissenschaftler des Fraunhofer Heinrich-Hertz-Instituts (HHI) und der Physikalisch-Technischen Bundesanstalt (PTB) haben in einem Artikel in der aktuellen-Ausgabe von Scientific Data, dem Datenjournal von Nature, die Beschreibung eines Datensatzes von 21 837 10s-EKG-Signalen von 18 885 Patienten veröffentlicht. Er ist damit der größte frei zugängliche Datensatz seiner Art und beispielsweise etwa 40-mal größer als die PTB Diagnostic Database, die bis jetzt häufig verwendet wurde. Er liefert maschinenlesbare Befunde und über 70 verschiedene EKG-Annotationen von bis zu zwei Kardiologinnen und Kardiologen. Der Datensatz ist öffentlich verfügbar bei PhysioNet (<https://www.physionet.org/content/ptb-xl>) und im Scientific-Data-Artikel detailliert beschrieben. Der Datensatz wurde mit besonderem Hinblick auf die Verwendung zur Entwicklung von maschinellen Lernverfahren aufbereitet und zu deren besseren Vergleichbarkeit in Trainings- und Testabschnitte organisiert.

Künstliche Intelligenz ist in der Medizin ein großer Trend. Gerade in Bereichen wie bei der EKG-Auswertung, bei denen sehr viel Erfahrung gefordert ist, kann Deep Learning potenziell seine großen Vorteile ausspielen. Diese Algorithmen erkennen Muster in großen Datenmengen, wie es bisher nur erfahrene Kardiologinnen und Kardiologen können, und können den Arzt oder die Ärztin bei der zeitaufwendigen Überprüfung der vielen EKG-Signale unterstützen. Existierende Algorithmen mit exzellenter Performance wurden dazu typischerweise auf nicht-öffentlichen Datensätzen trainiert und entziehen sich damit der Nutzung der breiteren wissenschaftlichen Community, während öffentliche Datensätze bis dato zu klein zum Training und insbesondere für eine verlässliche Evaluation maschineller Lernalgorithmen waren. Zudem ist die Evaluierungsmethodik nicht standardisiert, was für eine mangelnde Vergleichbarkeit der Ergebnisse sorgt. Im Rahmen des EU-EMPIR-Projektes Medalcare arbeiten die Wissenschaftlerinnen und Wissenschaftler des Fraunhofer Heinrich-Hertz-Institut (HHI) zusammen mit Kollegen von der Physikalisch-Technischen Bundesanstalt (PTB) daran, verschiedene maschinelle Lernalgorithmen anhand dieses großen Datensatzes zu vergleichen. Eine erste Benchmark-Studie zu diesem Thema ist als Preprint erschienen und vergleicht gängige Klassifikationsalgorithmen anhand verschiedener Aufgaben und klar definierten Evaluationsprozeduren.

Die publizierten Ergebnisse sollen als Anregung für weitere Wissenschaftlerinnen und Wissenschaftler dienen, die mit der Datenbank weiterarbeiten wollen. Denn bevor die Deep-Learning-Algorithmen breite Anwendung in der klinische Praxis Anwendung finden können, werden noch einige Untersuchungen folgen müssen – beispielsweise zur Frage, welchen Einfluss Begleiterkrankungen auf die EKG-Daten haben.

### **Originalpublikation:**

Patrick Wagner, Nils Strodthoff, Ralf-Dieter Bousseljot, Dieter Kreiseler, Fatima I. Lunze, Wojciech Samek, Tobias Schaeffter: PTB-XL, a large publicly available electrocardiography dataset. Scientific

Data 7, 154 (2020),

<https://doi.org/10.1038/s41597-020-0495-6>

Nils Strodthoff, Patrick Wagner, Tobias Schaeffter, Wojciech Samek. Deep Learning for ECG Analysis: Benchmarks and Insights from PTB-XL. arXiv preprint 2004.13701. 2020

<https://arxiv.org/abs/2004.13701>