Paper Overview: PTB-XL

# Background & Summary

* **PTB-XL** is the largest freely available clinical **12-lead ECG dataset**, containing **21,837 records** from **18,885 patients**, each **10 seconds long**.
* It aims to address two key challenges in machine learning-based ECG analysis:
  1. **Lack of large public datasets** for training and validation.
  2. **Lack of standardized benchmarking procedures** for algorithm comparisons.
* The dataset is multi-label, annotated by up to two cardiologists, and follows a hierarchical structure with **5 diagnostic superclasses** and **24 subclasses**.

A circular chart with different colored circles

AI-generated content may be incorrect.

A table of medical information

AI-generated content may be incorrect.

* It includes **metadata** such as demographic details, diagnostic statements, likelihood estimates, and signal quality annotations.

A bar chart with text

AI-generated content may be incorrect.

* The **diverse pathology coverage** and inclusion of a large fraction of healthy records make it a **valuable resource for ECG algorithm development**.

# Methods

## Data Acquisition

* Data was collected using **Schiller AG devices** between **October 1989 and June 1996**.
* ECG waveforms were **trimmed to 10-second segments**, stored in a **compressed proprietary format**, and sampled at **400 Hz**.
* ECG metadata was **entered by nurses** and included:
  + **Initial report:** 67.13% manually interpreted by a **cardiologist**, 31.2% **automatically generated**, with some manually validated.
  + **Report conversion:** Standardized into **SCP-ECG statements** with likelihood values.
  + **Additional metadata:** Includes heart axis, infarction stadium, and second validation (if available).

## Preprocessing

* ECG data was **converted into a binary format** with **16-bit precision**.
* **Minor noise removal** (e.g., switch-on/off artifacts).
* **Upsampling to 500 Hz** and **downsampled versions at 100 Hz** were also made available.
* Patient identifiers were anonymized to comply with **HIPAA regulations**.

# Data Records

## Waveform Data

* Available in **Waveform Database (WFDB) format**, a standard used by **PhysioNet**.
* Provided in **two versions**: **500 Hz (original)** and **100 Hz (downsampled)**.
* Each ECG is stored as **two files**: a **binary data file** and a **header file**.

## Metadata

* **Stored in ptbxl\_database.csv** and consists of **28 columns** with:
  + **Identifiers:** Unique IDs for ECGs and patients.
  + **General Metadata:** Age, sex, height, weight, recording device, location.
  + **ECG Statements:** Diagnostic labels, likelihood values, heart axis, infarction stadium.
  + **Signal Metadata:** Noise levels, baseline drift, electrode issues, presence of pacemakers.
  + **Cross-validation folds:** Predefined training and test sets to prevent **data leakage**.

## ECG Statements

* ECG statements follow **SCP-ECG standard**.
* Includes **71 different labels**, categorized as:
  + **Diagnostic statements (44)**
  + **Form statements (19)**
  + **Rhythm statements (12)**
* **Likelihood estimates** were extracted from ECG reports.

## Signal Metadata

* Signal quality is annotated with **baseline drift, noise levels, electrode issues, and extra beats**.
* 77.01% of records are of **high signal quality**.

**Cross-validation Folds**

* **10 stratified folds** were created for **machine learning model evaluation**.
* **Folds 9 & 10** contain only **human-validated ECGs** and are suggested for **validation and testing**.

# Technical Validation

## Quality Assessment for Waveform Data

* ECG signals contain **natural variations** found in real-world clinical settings.
* **Noise artifacts were labeled**, including:
  + **Baseline drift (7.36%)**
  + **Static noise (14.94%)**
  + **Burst noise (2.81%)**
  + **Electrode problems (0.14%)**
* 77.01% of data is **high-quality**, but the dataset maintains **real-world variability** to improve algorithm robustness.

## Quality Assessment for Annotation Data (ECG Statements)

* **73.7% of records were verified by human cardiologists**.
* Some records **lack clear documentation** on second validation.
* **To ensure high-quality test data**, **fold 10 only contains human-validated ECGs**.

# Usage Notes

**Conversion to Other Annotation Standards**

* ECG labels can be mapped to **AHA, CDISC, and DICOM** standards.
* scp\_statements.csv provides **cross-references** for different annotation systems.

**Prediction Tasks and Train-Test Splits for ML Algorithms**

* **Main tasks:** ECG classification (multi-label) using diagnostic, form, and rhythm statements.
* **Recommended metrics:** Macro-averaged **AUROC** (Area Under the Receiver Operating Characteristic Curve).
* **Temporal patient data** allows for **longitudinal studies**.
* **Folds 9 and 10** are suggested as **validation and test sets** due to their **high label quality**.

**Example Code**

* Python code is provided to:
  + **Load waveform and metadata** using wfdb and pandas.
  + **Aggregate diagnostic labels** into **superclasses**.
  + **Split data into training and test sets** based on stratified folds.

**CODE-15**

**Abstract**

* The study presents a DNN model trained on 2+ million ECG exams from the Telehealth Network of Minas Gerais (TNMG).
* The DNN outperformed cardiology residents in detecting six ECG abnormalities, achieving:
  + F1 scores above 80%
  + Specificity over 99%
* The findings suggest that DNNs generalize well to 12-lead ECGs, bringing the technology closer to clinical application.

**Introduction**

* 12-lead ECGs (S12L-ECG) provide more comprehensive heart assessments than single-lead ECGs.
* DNN-based interpretation of S12L-ECGs could be especially useful in:
  + Primary care and emergency settings (where specialists are absent)
  + Low and middle-income countries (which account for 75% of global CVD deaths)
* A major barrier to DNN adoption is the lack of large, labeled 12-lead ECG datasets.
* This paper addresses this gap by:
  + Building a large dataset (CODE study)
  + Training a DNN to classify 6 ECG abnormalities

**Results**

**Model Specification and Training**

* Dataset:
  + 2,322,513 ECGs from 1,676,384 patients
  + Collected from 811 counties in Minas Gerais, Brazil
* ECG abnormalities classified:
  + 1st-degree AV block (1dAVb)
  + Right bundle branch block (RBBB)
  + Left bundle branch block (LBBB)
  + Sinus bradycardia (SB)
  + Atrial fibrillation (AF)
  + Sinus tachycardia (ST)

**Testing and Performance Evaluation**

* Test dataset:
  + 827 ECGs annotated by 3 cardiologists
* Results:
  + F1 scores > 80% for all abnormalities
  + Specificity > 99%
* Comparison with medical professionals:
  + DNN matched or outperformed:
    - 4th-year cardiology residents
    - 3rd-year emergency residents
    - 5th-year medical students
* Error analysis:
  + Most DNN mistakes were related to borderline cases (e.g., HR just above/below thresholds).
  + DNN made fewer errors due to noise than human doctors.

**2. Discussion**

* DNN-based ECG analysis is a paradigm shift from traditional methods.
* Classic ECG analysis extracts features before classification.
* End-to-end learning (DNN approach) directly processes raw signals, learning feature extraction automatically.
* Advantages of the DNN model:
  + Trained on the largest ECG dataset ever used
  + Handles both rhythmic & morphological abnormalities
  + More robust to noise than human doctors
* Limitations:
  + The test set was small (827 ECGs), limiting statistical confidence.
  + The model was not tested on complex conditions like myocardial infarction or hypertrophy.
  + Real-world clinical settings are more complex than the study environment.
* Future research:
  + Extend the model to detect more abnormalities.
  + Test the system in real-time clinical environments.
  + Integrate DNNs into large telehealth systems (e.g., TNMG, which handles 2,000+ ECGs daily).

**3. Methods**

3.1 Dataset Acquisition

* ECGs collected by TNMG, a telehealth service in Brazil.
* Performed in primary care facilities using digital electrocardiographs.
* Stored in a database with patient history and cardiologist reports.

3.2 Labeling Training Data from Text Reports

* Challenge: Reports were unstructured text.
* Solution: Used natural language processing (NLP) and a semi-supervised classifier to extract labels.

3.3 Training and Validation Set Annotation

* Combining:
  1. Automated diagnoses (from ECG software)
  2. Manual cardiologist annotations
* Filtering process:
  1. Accepted diagnoses if both sources agreed.
  2. Rejected if only one source identified an abnormality.
  3. Manually reviewed 34,512 ambiguous cases.

3.4 Test Set Annotation

* Test set (827 ECGs) was independently annotated by two expert cardiologists.
* Disagreements resolved by a third senior cardiologist.

3.5 Neural Network Architecture and Training

* DNN Design:
  + Convolutional ResNet adapted for 1D ECG signals.
  + Input: 12-lead ECGs, resampled at 400 Hz.
  + Output: Multi-label classification (since multiple abnormalities can exist in one ECG).
* Optimization:
  + Used Adam optimizer.
  + Trained for 50 epochs.
  + Learning rate adjusted dynamically.

3.6 Hyperparameter Tuning

* Manually optimized over ~30 iterations.
* Experimented with:
  + Different architectures (ResNet, VGG, LSTM)
  + Various kernel sizes, batch sizes, dropout rates, activation functions.

3.7 Statistical and Empirical Analysis

* Performance metrics:
  + Precision, recall, specificity, F1 score.
* Bootstrap analysis:
  + 1,000 samples used to estimate metric distributions.
* McNemar test:
  + Compared misclassification rates between DNN and medical residents.
* Error classification:
  + DNN errors: Measurement, noise, or unexplained.
  + Doctor errors: Conceptual or attention related.

**Conclusion**

* Developed an end-to-end DNN that accurately detects 6 ECG abnormalities.
* Matched or outperformed medical residents and students.
* Potential future impact:
  + Improve ECG analysis accuracy
  + Expand access to automatic ECG interpretation
  + Reduce diagnostic errors in remote & under-resourced areas
* Next steps:
  + Expand classification to more abnormalities.
  + Implement real-time clinical trials.