Cuffless Blood Pressure Estimation from Photoplethysmography Using Time-Series Feature Engineering and Ensemble Learning

Vignan Kamarthi   Ariv Ahuja

DS4400: Machine Learning 1 – Spring 2026

1 Problem Description

Cardiovascular diseases (CVDs) are the leading cause of mortality worldwide. Blood pressure (BP) is a critical CVD risk indicator, yet conventional measurement requires an inflatable cuff, a method that is intrusive and only captures isolated snapshots rather than continuous readings. This makes it unsuitable for ongoing monitoring throughout daily life. Cuffless BP monitoring from wearable sensors would enable early hypertension detection and real-time cardiovascular tracking.

Photoplethysmography (PPG), the optical pulse-sensing technology found in consumer smartwatches, captures blood volume changes that correlate with arterial pressure dynamics. This project frames cuffless BP estimation as a **supervised regression task**: given time-series features extracted from a 10-second PPG waveform segment, predict the corresponding systolic blood pressure (SBP, the peak pressure when the heart contracts) and diastolic blood pressure (DBP, the lowest pressure when the heart relaxes), both measured in mmHg.

The core methodological contribution is a **dual feature extraction pipeline** combining Catch22 canonical time-series features with entropy-based complexity measures, followed by ensemble learning for BP regression. The Catch22 approach adapts methods from prior work on physiological signal classification [Boda et al., 2025], while the entropy-based features build on established information-theoretic measures including permutation entropy [Bandt and Pompe, 2002] and complexity-entropy analysis [Rosso et al., 2007], applied in ongoing biosignal complexity research. Together, these capture complementary aspects of PPG morphology relevant to arterial pressure.

2 Dataset

PulseDB v2.0 [Wang et al., 2023]
`https://github.com/pulselabteam/PulseDB`

- **Source:** MIMIC-III Waveform Database Matched Subset (Beth Israel Deaconess Medical Center, Boston) + VitalDB (Seoul National University Hospital, South Korea), providing geographic and demographic diversity across two independent hospital systems
- **Size:** 5,245,454 ten-second segments from 5,361 subjects (approximately 14,570 hours of recording)

- **Signals:** Raw PPG waveform sampled at 125 Hz (1,250 data points per 10-second segment), with ground-truth SBP and DBP labels derived beat-by-beat from invasive arterial blood pressure (ABP) waveform recordings
- **Splits:** Training (2,506 subjects), calibration-based test (same subjects as training but with held-out segments), calibration-free test (279 completely disjoint subjects never seen during training, with ground-truth labels retained for evaluation)

The calibration-free split is the more clinically meaningful evaluation: it tests whether the model generalizes to entirely new individuals without any patient-specific calibration data.

3 Approach and Methodology

3.1 Feature Extraction

Each 10-second raw PPG segment (1,250 samples at 125 Hz) is transformed into a fixed-length feature vector through three complementary extraction methods:

1. **Catch22** (22 features): Canonical time-series features capturing autocorrelation structure, distributional properties, successive differences, and fluctuation scaling [Lubba et al., 2019]. Applied following prior work on physiological signal classification [Boda et al., 2025].
2. **Statistical** ($\sim$8 features): Mean, median, standard deviation, skewness, kurtosis, root mean square, min, max.
3. **Entropy-based** ($\sim$5 features): Sample entropy, permutation entropy [Bandt and Pompe, 2002], approximate entropy, spectral entropy, and Hjorth complexity. These quantify signal regularity and complexity, capturing aspects of PPG morphology not represented by Catch22 or statistical summaries.

The resulting feature vector ($\sim$35 features per segment) is compact enough for efficient training over the full 5.2 million segments while capturing distinct dimensions of the PPG waveform relevant to blood pressure.

3.2 Feature Normalization

Standard scaling (z-score normalization) fitted on training data only, then applied to test data to prevent data leakage. Normalization is critical for Ridge regression, which is sensitive to feature scale. Tree-based models (Random Forest, XGBoost, LightGBM) are scale-invariant since they split on thresholds rather than feature magnitudes, but we normalize uniformly across all models for pipeline consistency and will run an ablation to confirm the impact.

3.3 Models

Separate models trained for SBP and DBP. Hyperparameters tuned via cross-validation.

Model	**Type**	**Role**
Ridge Regression	Regularized linear (L2 penalty)	Interpretable baseline
Random Forest	Bagging ensemble	Variance reduction; feature importance
XGBoost	Gradient boosting	Sequential error correction
LightGBM	Histogram-based boosting	Efficient large-scale training

3.4 Language and Packages

Python: `pycatch22`, `antropy`, `EntropyHub`, `scikit-learn`, `xgboost`, `lightgbm`. Feature extraction parallelized on NEU Explorer cluster (SLURM, 56 CPUs).

3.5 Evaluation Metrics

- **MAE**: Primary accuracy metric. Target < 5 mmHg per AAMI standard.
- **SD of Error**: Prediction consistency. Target < 8 mmHg per AAMI. Low MAE with high SD indicates unreliable individual predictions.
- **RMSE**: Penalizes large errors more heavily than MAE, surfacing dangerous outlier predictions.
- R^2: Proportion of BP variance explained by the model.
- **Bland-Altman analysis**: Clinical agreement plot (mean vs. difference) showing bias and limits of agreement. Standard in medical device validation.

Results reported on both calibration-based and calibration-free test sets to quantify the generalization gap.

4 Outcome

- Predict SBP and DBP from PPG with MAE approaching the AAMI clinical standard (< 5 mmHg)
- Demonstrate that Catch22 + entropy features capture BP-relevant PPG physiology
- Quantify calibration-free vs. calibration-based performance gap
- Identify top predictive PPG features via importance analysis and ablation

5 Plan (Rough, Subject to Change)

Task	**Owner**
Dataset acquisition and EDA	Vignan
Feature extraction pipeline	Vignan
Baseline models (Ridge, RF)	Ariv
Boosted models + tuning	Vignan
Milestone report	Both
Feature ablation study	Ariv
Bland-Altman + AAMI evaluation	Vignan
Final report + presentation	Both

References

Christoph Bandt and Bernd Pompe. Permutation entropy: A natural complexity measure for time series. *Physical Review Letters*, 88(17):174102, 2002.

Sai Revanth Reddy Boda, Vignan S. Kamarthi, Burcu Ozek, Zhenyuan Lu, and Srinivasan Radhakrishnan. Canonical time series features for pain classification. In *Companion Proceedings of the 27th International Conference on Multimodal Interaction*, ICMI Companion '25, page 6, New York, NY, USA, October 2025. ACM. doi: 10.1145/3747327.3764784.

Carl H Lubba, Sarab S Sethi, Philip Knaute, Simon R Schultz, Ben D Fulcher, and Nick S Jones. catch22: Canonical time-series characteristics. *Data Mining and Knowledge Discovery*, 33(6): 1821–1852, 2019.

Osvaldo A Rosso, Hilda A Larrondo, Maria T Martin, A Plastino, and Miguel A Fuentes. Distinguishing noise from chaos. *Physical Review Letters*, 99(15):154102, 2007.

Weinan Wang, Pedram Mohseni, Kevin L Kildar, Jonathan E Sutherland, et al. Pulsedb: A large, cleaned dataset based on mimic-iii and vitaldb for benchmarking cuff-less blood pressure estimation methods. *Frontiers in Digital Health*, 4:1090854, 2023.