

Dynamic Time Warping with Gradient Boosting Tree Ensemble for 12-Lead Electrocardiogram Multilabel Classification

Alexander W Wong¹, Weijie Sun², Sunil V Kalmady², Padma Kaul², Abram Hindle¹

¹ University of Alberta, Edmonton, Canada

² Canadian VIGOUR Center, Edmonton, Canada

Abstract

Standard 12-lead electrocardiograms (ECGs) are commonly used to detect cardiac irregularities such as atrial fibrillation, blocks and irregular complexes. For the Physionet/CinC 2020 Challenge, we built an algorithm using gradient boosted tree ensembles fitted on morphology and signal processing features.

We used the ecgpuwave implementation of the Pan Tompkins method to detect the P-wave, QRS complex, and T-wave. We selected templates that exhibited maximum similarity with the rest of the ECG records in the same class using minimum distance criteria to isolate candidate templates for each cardiac abnormality. We leveraged these templates using Dynamic Time Warping (DTW), an algorithm used for measuring similarity between temporal sequences of varying speeds. From the annotated signals we derived additional features related to their durations, amplitudes, and intervals.

We concatenated the features for all 12 leads to fit an ensemble of gradient boosting trees and predicted probabilities of ECG instances belonging to each class. We evaluated using a 5-fold cross validation split of the provided dataset. TODO: new scoring metric

1. Introduction

Cardiovascular diseases are a non-communicable leading cause of death, responsible for an estimated 17.8 million fatalities world wide in 2017 [1]. The electrocardiogram (ECG) is the most effective tool and current best practice strategy for detecting cardiac diseases, outperforming screening history and physical examinations in accuracy and sensitivity [2]. However, ECG interpretation is a complex and highly skilled task undertaken by multiple health care professionals ranging from physicians, nurses, paramedics, and technicians. Multiple studies have been conducted highlighting disagreements between non-cardiologists and cardiologist reference ECG interpretations, with up to a 33% interpretation error rate [3]. De-

spite active research in computerized interpretations of ECGs, trained human over-reading and confirmation is required and emphasized in published reports [4].

This work classifies standard 12-lead ECGs to their clinical diagnosis as part of the *PhysioNet/Computing in Cardiology Challenge* [5]. We developed a multi-label classification algorithm using natural language and signal processing inspired features and a gradient boosting tree ensemble.

1.1. Dataset & Scoring Criteria

The official phase dataset contained a total of 43,101 ECG records. After dropping 6 records that had less half or more of their leads zero'd out, the remaining 43,095 records were used for further analysis. Each record was labeled with one or more SNOMED CT codes, with 123 different labels observed in the provided data.

Table 1. Top 7 labels within the provided dataset.

| Label | Percent | Count |
|------------------------------|---------|-------|
| sinus rhythm | 48.4% | 20846 |
| left axis deviation | 14.1% | 6086 |
| myocardial infarction | 14.0% | 6019 |
| t wave abnormal | 10.8% | 4673 |
| left ventricular hypertrophy | 8.7% | 3760 |
| nonspecific st t abnormality | 8.3% | 3557 |
| atrial fibrillation | 8.1% | 3473 |

Table 1 highlights the label imbalance, with over 87% of the dataset diagnosed as one of the four most common labels. Approximately 55% of the dataset is classified with more than one label, as shown in Table 2.

TODO: Scoring criteria, when released

2. Methodology

Our approach is inspired by existing methods which use feature engineering and shallow learning classifiers [6, 7]. An overview our methodology is presented in Figure 1.

Table 2. Dataset multi-label records proportion.

| Labels Per Record | Percent | Count |
|-------------------|---------|-------|
| 1 | 45.2% | 19461 |
| 2 | 21.8% | 9378 |
| 3 | 14.9% | 6426 |
| 4 | 9.9% | 4280 |
| 5 | 4.9% | 2113 |
| 6 | 2.2% | 935 |
| 7 | 0.8% | 359 |
| > 7 | 0.3% | 143 |

$$\text{DTW}(X, Y) = \min_W \left(\sum_{h=1}^P (w_k) \right) \quad (1)$$

Dynamic Time Warping (DTW) measures the similarity of two different sequences [8]. Given two time-series $X := (x_1, x_2, \dots, x_N)$ and $Y := (y_1, y_2, \dots, y_M)$ where $N, M \in \mathbb{N}$, an $N \times M$ matrix is constructed containing the distance between each element of the two series, $d(x_i, y_j)$. A warping path $W := (w_1, w_2, \dots, w_k)$ aligns X and Y such that the distance between them is minimized. Equation 1 defines DTW as the minimization over all warping paths based on the cumulative distance for each path.

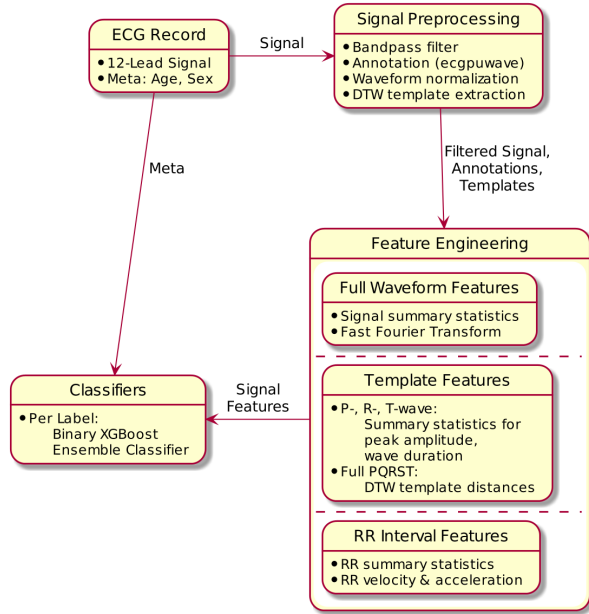


Figure 1. Methodology overview. Feature engineering is performed concurrently for each lead then concatenated.

2.1. Signal Pre-processing

We first resampled all of our signals to 300 Hz. We then filtered the raw signal using a finite impulse response filter, limited between 3 Hz and 45 Hz. Next, we used the *ecg-puwave* QRS detector to annotate each filtered signal lead, resulting in waveform start, end, and peaks for P-waves, T-waves, and R-waves [9]. ECG PQRST templates were extracted by segmenting out continuous P-wave waveform starts to T-wave waveform ends.

For DTW template matching candidates, the extracted PQRST signals were first pooled by lead and label. The BioSPPy implementation of the minimum distance template selection and clustering method was then used to isolate relevant templates per diagnosis [10, 11].

2.2. Feature Engineering

Our engineered features can be categorized as one of three categories. Full waveform features are derived using the end-to-end ECG signal. Template features are constructed from the extracted PQRST segments during pre-processing. RR interval features rely on the underlying heart rate-the relative distances between each R-peak. Summary statistics refer to the min, max, median, mean, standard deviation, and variance. Each extraction technique was performed independently per lead.

For full waveform features, we used summary statistics for the filtered signal amplitude and the discrete fourier transform of the sequence. Applying the Hann window to our signal, we binned the fourier transform output to 50 discrete bins containing the range 0-50 Hz.

Template features included summary statistics for the P-wave, T-wave, and R-wave peak amplitudes and wave durations. Additionally, we applied DTW to compare the PQRST segments to each of the candidate segments per diagnosis to derive distance measurements.

We used summary statistics obtained from the distances between consecutive R-peaks. To capture change in heart rate, we also derived summary statistics from RR interval acceleration and velocity.

2.3. Classification

We trained an XGBoost binary classifier to for each clinical diagnosis, taking as input all constructed features, the patient age and sex as input [12]. A 5-fold cross validation split was used to tune for hyperparameters.

3. Results

TODO

4. Discussion

TODO

Acknowledgments

We would like to thank Eric Ly and Leiah Luoma for their help and guidance during our research journey.

References

- [1] Roth GA, Abate D, Abate KH, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* November 2018;392(10159):1736–1788. ISSN 0140-6736.
- [2] Harmon KG, Zigman M, Drezner JA. The effectiveness of screening history, physical exam, and ECG to detect potentially lethal cardiac disorders in athletes: A systematic review/meta-analysis. *Journal of Electrocardiology* May 2015;48(3):329–338. ISSN 0022-0736.
- [3] Mele P. Improving electrocardiogram interpretation in the clinical setting. *Journal of Electrocardiology* September 2008;41(5):438–439. ISSN 0022-0736.
- [4] Schläpfer J, Wellens HJ. Computer-Interpreted Electrocardiograms: Benefits and Limitations. *Journal of the American College of Cardiology* August 2017;70(9):1183–1192. ISSN 0735-1097, 1558-3597. Publisher: Journal of the American College of Cardiology Section: The Present and Future.
- [5] Reyna M, Alday EAP, Gu A, Liu C, Seyedi S, Rad AB, Elola A, Li Q, Sharma A, Clifford G. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. *PhysioNet*, 2020. DOI: 10.13026/f4ab-0814.
- [6] Tziakouri M, Pitris C, Orphanidou C. Identification of AF and Other Cardiac Arrhythmias from a Single-lead ECG Using Dynamic Time Warping. September 2017; .
- [7] Goodwin A, Goodfellow S, Eytan D, Greer R, Mazwi M, Laussen P, Goodfellow S. Classification of Atrial Fibrillation Using Multidisciplinary Features and Gradient Boosting. September 2017; .
- [8] Berndt DJ, Clifford J. Using Dynamic Time Warping to Find Patterns in Time Series. In *KDD workshop*, volume 10. Seattle, WA, April 1994; 359–370.
- [9] Goldberger Ary L., Amaral Luis A. N., Glass Leon, Hausdorff Jeffrey M., Ivanov Plamen Ch., Mark Roger G., Mietus Joseph E., Moody George B., Peng Chung-Kang, Stanley H. Eugene. *PhysioBank, PhysioToolkit, and PhysioNet*. *Circulation* June 2000;101(23):e215–e220. Publisher: American Heart Association.
- [10] BioSPPy: Biosignal processing in Python, 2015–. URL <https://github.com/PIA-Group/BioSPPy>.
- [11] Uludag U, Ross A, Jain A. Biometric template selection and update: a case study in fingerprints. *Pattern Recognition* July 2004;37(7):1533–1542. ISSN 0031-3203.
- [12] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD In-*

ternational Conference on Knowledge Discovery and Data Mining, KDD '16. San Francisco, California, USA: Association for Computing Machinery. ISBN 978-1-4503-4232-2, August 2016; 785–794.

Address for correspondence:

Alexander W. Wong
Department of Computing Science
2-32 Athabasca Hall, University of Alberta
Edmonton, Alberta, Canada
T6G 2E8
alex.wong@ualberta.ca