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

Alexander W Wong[1], Weijie Sun[2], Sunil V Kalmady[2], Padma Kaul[2], Abram Hindle[1]

[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 morphology related durations, amplitudes, and intervals. Additional signal representation techniques, including discrete Fourier transformations and polynomial function fitting, were also used to extract features.

We concatenate the features for all 12 leads and fit an ensemble of gradient boosting trees to predict probabilities of ECG instances belonging to each class. We evaluated using a 5-fold cross validation split of the provided dataset.

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]. Despite 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. Print font

Please use Times New Roman font. Use of this font by all authors ensures that our papers are produced consistently with a professional appearance.

1.2. Print size

Follow the information given in Table 1 for all font print sizes and styles.

2. Title block

On the first page, the title, authors, and institution(s) should be positioned and aligned as above.

2.1. Title

Do not use abbreviations in the title and keep to one or two lines. Remember that the title should be easily understood when cited as a reference in another publication. Make sure the title is centered. Note the use of capital letters, as shown in the example above. Do not place a period at the end of the title.

2.2. Author line

If it is essential to link authors to different institutions, use a small superscript at the end of each family name and link to each institution. No author degrees should be included. No periods should appear in the author line.

2.3. Institution address

This address identifies the institution(s). It is not a postal address and should not contain postal or zip codes. A full postal address is given at the end of the paper under 'Address for correspondence'. Do not put line spaces between address lines or use a period at the end.

3. Tables and figures

3.1. Tables

Table 1. Print sizes for different parts of the paper.

Text	Point size	Type
Title	14 point	Bold
Author line	12 point	
Institution line	12 point	
Abstract heading	11 point	Bold
All other headings	12 point	Bold
Abstract text	10 point	Italic
References and address	9 point	
All other text	10 point	

Keep the layout of tables simple. Avoid heavy lines and boxes.

3.2. Figures

Figures can fit across both columns if necessary. Include all figures, with captions, in the text document at their desired locations, i.e., not at the end. To make sure that all labels are legible, use at least 9-point font. Use a line thickness of at least 0.5 mm in the body of the figure.

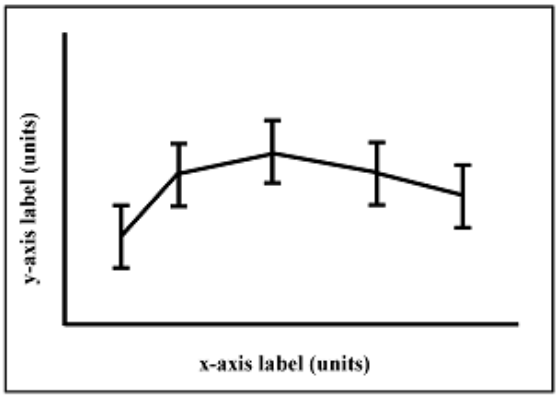

Figure 1. Put the figure legend here, at the bottom, clearly describing the figure.

Leave a line space after a figure legend. Be cautious with background colors in figures because they can make printed figures hard to read.

4. Final page

The text on the final page should be arranged so that both columns are approximately the same length.

5. Style of references

All references should be included in the text in square brackets in their order of appearance, e.g., [6], [6, 7], or [6–9]. In the reference list, use the Vancouver style (see IEEE Transactions on Biomedical Engineering or Annals of Biomedical Engineering). The main goal is to be consistent in your formatting of references.

References to *Computers in Cardiology* (or *Computing in Cardiology*) proceedings should include the volume and page numbers, unless no page numbers were included (post-2016). Check the style in the examples below.

6. Final checks

Print this template without resizing or centering and compare with your final paper to ensure that you have not accidentally changed margins and layout.

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] Author JK. Name of paper. Abbrev Title of Periodical Month 2018;x(xx):yyy–zzz.
[7] et al. MI. Application of amorphous oxide tft to electrophoretic display,. J Non Cryst Solids 2008;354(19):2777–2782.
[8] Fardel R, Nagel M, Nuesch F, Lippert T, Wokaun A. Fabrication of organic light emitting diode pixels by laser-assisted forward transfer. Appl Phys Lett Aug. 2007;91(6):Art. no. 061103.
[9] Buncombe J. Infrared navigation part i: Theory. IEEE Trans Aerosp Electron Syst Sep. 1944;AES-4(3):352–377.

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