A sparse, data-efficient ECG representation is predictive of myocardial infarction without expert knowledge

S. Rane
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
sunayana@mit.edu

Abstract

We present work showing that a sparse, data-efficient ECG representation provides predictive signal for myocardial infarction without any expert knowledge implicitly or explicitly used to aid the model. Our process for extracting this representation is scalabale, easily trained with small amounts of data, and ideal for deployment in the low- and middle-income countries that most need early-warning systems for myocardial infarction. We hope that the signal generated through this method can be combined as a predictive feature in diagnostic models and eventually lead to fully automated diagnostic systems in areas where access to trained cardiologists is scarce.

1 Introduction

According to the World Health Organization, Cardiovascular Disease (CVD) is the leading cause of death worldwide CVD [2018]. 3/4 of these deaths are in low- and middle-income countries, where access to trained experts (cardiologists and medical professionals) is scarce. Of those deaths, 85% are due to heart attack (in medical terms, myocardial infarction) and stroke. The vast majority of these deaths are preventable through early diagnosis and risk assessment, combined with lifestyle changes. However, most of the people at risk do not have the means or access to experts to get those early diagnoses.

Current state-of-the-art work in ECG analysis achieves high diagnostic accuracy for several disease types, including myocardial infarction. However, in order to achieve these accuracies, the current myocardial infarction models require heavy preprocessing which encodes expert knowledge. There has been progress in using deep neural networks with raw (unprocessed) ECG data, however this is primarily for arrhythmia diagnosis and requires datasets on the order of thousands of examples per class in addition to substantial computational resources for training. This is usually not feasible for practical diagnostics scenarios in low income countries. This research presents a different approach: a sparse representation of ECG data that requires very small amounts of data, computational power, and generates signal that is predictive for myocardial infarction.

2 Related Work

Automated ECG diagnosis has always been a difficult task, and any successful solution involves significant tradeoffs. Therefore, it is important to develop solutions with the intended application in mind, so that the tradeoffs we make are in line with our goals for deployment. Our intended application is to provide an early-warning diagnosis without any expert knowledge, for use in

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

countries and areas where access to experts is severely limited. Scalability and generalizability are key to learning from small amounts of data (data-efficieny), which is a critical aspect of achieving this goal.

The most successful traditional approach to ECG beat classification has been to use wavelet transforms to extract relevant features Saxena *et al.* [2002]. However, this type of analysis usually involves several hand-crafted variables, including specially made denoising and band-pass filters that are tuned to the specific frequency that is likely most predictive for a particular dataset Tripathy and Dandapat [2017]. For example, if the hypothesis is that the P-wave of an ECG will be particularly predictive for a specific disease or pathology, one would design a wavelet transform analysis to extract the frequency, magnitiude, and variance information of P-waves in the presence of noise masking Diery *et al.* [2011]. Successfully executing wavelet transforms usually requires some expert knowledge of how cardiologists interpret ECGs. They also require some dataset-specific crafting of filters and preprocessing, which is not easily transferable to new data and prediction problems. Moreover, they are often combined with traditional signal processing and expert systems for diagnosis, which (while they definitely have their advantages in interpretability), are also not easily generalizable or scalable without large hard-crafted collections of expert knowledge over which to reasonAl-Ani and Ayal Rawi [2013].

Within the last two years, there has also been promising progress in deep learning for ECG analysis, which has the benefit of not requiring the expert knowledge that is necessary for expert systems and and traditional wavelet transform analyses. Deep learning has been known to be a *tabula rasa* approach, where all relevant knowledge for prediction is extracted by the network itself. A downside of this approach is a lack of interpretability or clinical accountability for diagnostic results. However, interpretability is not as critical for our purposes of generating signal for an early-warning system which will then be examined by a professional, so deep learning is a promising direction. The most successful example of deep learning for ECG analysis used a 1-D convolutional neural network with residual connections (a 1-D ResNet) in order to classify various types of arrhythmias Hannun *et al.* [2019]. The model was able to achieve cardiologist-level performance. However, it requires many thousands of training examples and only works for arrhythmias, which are irregularities in the temporal domain, and are therefore easily capturable with 1-D convolutions over the time domain data.

There have also been some approaches to combining deep learning approaches with data-efficiency, such as transfer learning. Learning from small amounts of data is difficult and is key to the problem we are trying to solve. Therefore, we hope that deep transfer learning approaches can combine well with our extracted signal for completely automated diagnosis. The closest work to this approach takes a 1-D convolutional resnet trained to distinguish arrhythmia ECGs and uses the final layer to extract a deep representation of the ECG Kachuee *et al.* [2018]. In a form of transfer learning, this deep representation is then used to classify healthy vs myocardial infarction ECGs, with excellent performance and test accuracies above 90%. However, following the common theme in requiring extensive preprocessing, this work requires a detailed, 10-step preprocessing pipeline, which encodes expert knowledge of ECG morphology. For example, a particularly good-quality/predictive 10s subsegment of the larger ECG is chosen, and is then normalized and split into beats with an algorithm that identifies R-peaks (a particular morphological feature of ECGs). For each R-peak, a section of the signal of a particular (calculated and tuned) length is extracted around that peak. Only then does the classifier yield excellent transferable performance. We are excited about the potential of deep representations, but in this work they still require some amount of expert knowledge of ECG morphology and human intervention to be executed. We hope that our representation, which is both data-efficient and requires no expert knowledge, will work in tandem with deep representations as a diagnostic feature, adding new valuable information which will increase overall signal and reduce the need for complex preprocessing that is presently a bottleneck to generalizable, scalable, automated systems.

3 PTB Diagnostic Database

To obtain raw ECG diagnostic data, we used the PTB diagnostic database Bousseljot *et al.* [2004], available through PhysioNet Goldberger *et al.* [2000 June 13], for our work. A benefit of the PTB database is its inclusion of supplementary data, including demographics and medical annotations, which can be incorporated into future models. The database consists of 15-lead ECG data for 268

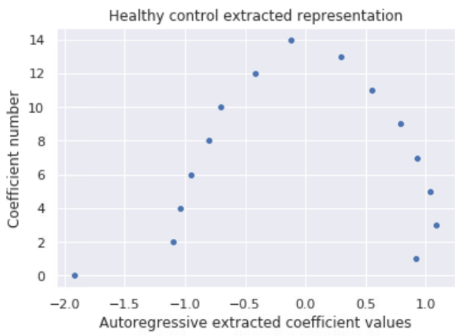

Figure 1: The extracted representation of a sample from a healthy control

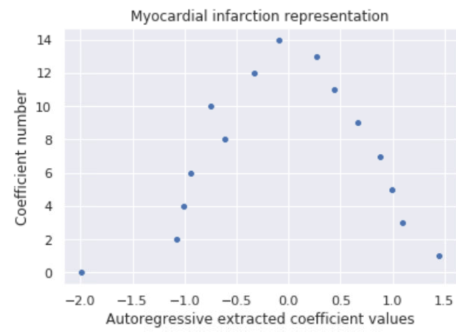

Figure 2: The extracted representation of a sample from a myocardial infarction patient

subjects, falling into 8 diagnostic classes: myocardial infarction, cardiomyopathy, bundle branch block, dysrhythmia, myocardial hypertrophy, valvular heart disease, myocarditis, healthy controls, and “miscellaneous”. However, each of the diagnostic classes other than myocardial infarction and healthy controls contain fewer than 20 samples. Myocardial infarction had 148 samples, and healthy controls had 52, all from different patients. For this reason and the fact that we wanted to address the leading (and most preventable) cause of cardiovascular death in low- and middle-income countries, we chose to focus solely on myocardial infarction. The 15-leads consist of the conventional 12 leads (i, ii, iii, avr, avl, avf, v1, v2, v3, v4, v5, v6) in addition to the 3 Frank lead ECGs. For our diagnostic models, we use only one of these leads (avr) as input. The recordings are of relatively high quality, sampled at 1000Hz with 16-bit resolution over a range of 16.384 mV.

The duration of recorded ECGs varies by patient in the PTB database. To ensure uniformity in input length for our model, we trimmed all ECGs to conform with the smallest recorded length, 38,400 samples. The trimming was automated and simply extracted the first 38,400 recorded data points for each patient. This corresponds to a little over 30s of raw ECG data per patient, which was the input to our representation algorithm.

4 Extracted representations

One of the biggest challenges of automatically extracting valuable information from an ECG signal is its high dimensionality. Convolutional neural networks have gained a great deal of recognition in recent years for their use in dimensionality-reduction. However, 1D convolutions for timeseries data, such as an ECG, primarily useful for extracting temporal patterns (e.g. arrhythmia). This is because they convolve over the time domain only. Autoregression helps solve this problem by providing an intuitively recurrent feature extraction framework, adaptable to multiple diseases and requiring orders of magnitude fewer data samples than a CNN.

Autoregressive parameter estimation essentially attempts to estimate the coefficients of a polynomial that is representative of the data. Autoregressive processes are so named because they incorporate their own output from a previous timestep into their input for a future timestep. The particular form of autoregression we use is called Burg’s method, initially developed for spectral estimation. It estimates the autoregressive coefficients best suited to the input signal by minimizing loss in forward and backward prediction errors, constrained to satisfy Levinson-Durbin recursion Al-Fahoum and A Al-Fraihat [2014]. The foundational Burg equation for power spectral density estimation (PSD) outputs a much smaller set of extracted features, with minimal levels of noise and better frequency resolution. This technique has been observed to have certain advantages over the fast-Fourier transform method of feature extraction in other physiological signals, including EEG Al-Fahoum and A Al-Fraihat [2014]. Most importantly, it is scalable, requires no preprocessing, and (particularly compared to deep learning) demonstrates valuable results even on small amounts of data.

5 Classification and evaluation

Due to class imbalance that can cause accuracy to be deceptively high, we opted instead for the area under the receiver operating characteristic (ROC) curve as our evaluation metric. ROC AUC

Table 1: Classification accuracies for various classifiers trained ("Representation" column indicates whether our extracted representation was used)

CLASSIFIER	ROC AUC	REPRESENTATION
RANDOM FORESTS	0.69 ± 0.03	$\surd$
GRADIENT BOOSTED TREE	$\mathbf{0.72}\pm 0.03$	$\surd$
FC NEURAL NETWORK	0.63 ± 0.04	$\surd$
LOGISTIC REGRESSION	0.67 ± 0.03	$\surd$
K NEAREST NEIGHBOR	0.65 ± 0.03	$\surd$
RNN	0.49 ± 0.01	$\times$
CNN	0.49 ± 0.01	$\times$

is known to be a metric more robust to class imbalance and therefore more realistic to evaluate for our purposes. We also calculated confidence intervals for our ROC AUC values, by calculating the standard deviation across 10-fold cross validation.

We evaluated several different classifers using the extracted representations as input, and compared the performance of each. Due to the well-established nature of these classification techniques, we have kept explanations brief, with references for additional information. We used two tree-based classifers, both ensemble methods: random forests Ho [1995] and gradient-boosted trees Friedman [2002]. The random forest consisted of 10 decision trees. The gradient-boosted trees consisted of 100 trees, and a learning rate of 0.1. We also evaluated logistic regression, as well as a fully connected neural network van Gerven and Bohte [2017] with 2 hidden layers of 100 units each.

It is well known that deep learning requires large amounts of data to learn classification. We hypothesized that our data would not be sufficient for deep learning techniques to yield predictive extracted features. However, we decided to test this by including two deep learning baselines: The first was a four-layer 1-D convolutional neural network (CNN) and the second was a recurrent neural network with 1024 hidden units (using LSTM cells). The deep learning baselines do not use our extracted representation, and are instead given the same raw ECG data that we use as the raw input to our representation-finding mechanism. The purpose is to investigate whether deep learning is able to conduct automated, predictive feature extraction with such a small amount of data. Architectural specifications/details for both of these networks are provided as an appendix.

6 Results and Discussion

The resulting ROC AUCs demonstrate that the extracted representations provide predictive signal for these models. Gradient boosted trees were most effective, and Random Forests came in a close second, indicating that CART methods are promising for this type of extracted feature set. As expected with such small amounts of data, both deep learning approaches fail to provide predictive signal. CART methods tend to require less computational power than neural networks, which also makes them ideal for deployment in low-income countries and settings where large amounts of computational power are unavailable.

Predictive signal is a valuable starting point, and we hope that this work will be combined with other predictive representations for stronger combined models that do not require large amounts of data or any expert knowledge. These attributes will allow the models to easily scale to new patients in low-income countries without access to cardiologists.

This is by no means the best classification performance achievable on this task–expert systems with encoded knowledge have been shown to achieve accuracies greater than 90% for specific datasets. However, this is a data-efficient representation, learned completely from scratch without need for expert knowledge, that can easily transfer across datasets. We hope it can be combined with other data-efficient techniques for finding representations, as in Kachuee et. al 2018, to improve predictive performance and greatly reduce the need for expert knowledge even when we only have small amounts of data. We hope that this will help bring diagnostic tools to settings where they are needed most, democratizing access to cardiovascular healthcare.

7 Appendix: Deployment Suggestions

To gain a better understanding of the grassroots hurdles in implementing an early-warning diagnostic system like this in the underprivileged areas of a low- or middle-income country, we observed and consulted with doctors in several hospitals in Mumbai, India. By and large, doctors in hospitals of the wealthier areas are not particularly interested in using an automated decision-aid for diagnosis. However, doctors in government and publicly-funded hospitals are incredibly overworked. A filtration system where patients are first screened through an autonomous system and then passed on to expert cardiologists as needed would be an excellent fit for these settings. Additionally, preventative screenings and lifestyle education will greatly reduce the number of patients with heart attacks and strokes (which account for 85% of CVD deaths) CVD [2018] that the hospitals have to treat, helping ease the burden of doctors and facilities.

An example of such a system in action was pointed out to us: the "hub and spoke" model of medical treatment has shown excellent results in the neighboring city of Chennai, India. In this system, rural villages surrounding the city had smaller, mobile medical clinics, often with trained volunteers (but no doctors) who conduct a series of routine tests for various diseases. Patients found to be at high risk in these "spoke" areas are then fed into the "hub", or the more advanced medical centers at the heart of the city. Cardiology is one area where expert knowledge is still very necessary to make any kind of a diagnosis, and trained volunteers can't provide the kind of screening patients need. We have shown that these methods can provide predictive signal without any expert knowledge or help from trained cardiologists. Systems derived from our methods, that are data-efficient, scalable, and require no expert knowledge, are an excellent fit to fill this gap and help address a preventable, worldwide cardiovascular disease epidemic.

8 Acknowledgments

This work is dedicated to two of the kindest, most inspiring professors MIT has known: Patrick Henry Winston and Woodie Flowers. Thank you for always encouraging us to live for others, and for showing us how it's done. You are dearly missed.
Thank you to the MIT Tau Beta Pi Fellowship committee for funding this research. Thank you to the members of the Szolovits and Stultz labs at MIT, particularly Marzyeh Ghassemi, Matthew McDermott, and Paul Myers for their advice and suggestions. Thank you to Hal Abelson, Amar Gupta, Patrick Henry Winston, Suri Bandler, Vivek Miglani, and Natalie Lao for their support and feedback. Thank you to the MIT-India team for making the hospital visits possible. Thank you to the many cardiologists in those hospitals who provided valuable feedback about the nature of deploying this technology for impact.

References

Muzhir Al-Ani and Atiaf Ayal Rawi. A rule-based expert system for automated ecg diagnosis. *International Journal of Advances in Engineering Technology*, 6:1480–1493, 09 2013.

Amjed Al-Fahoum and Ausilah A Al-Fraihat. Methods of eeg signal features extraction using linear analysis in frequency and time-frequency domains. *ISRN neuroscience*, 2014:730218, 02 2014.

Ralf-Dieter Bousseljot, D Kreiseler, and A Schnabel. The ptb diagnostic ecg database, 2004.

Cardiovascular diseases (cvds), Sep 2018.

A. Diery, D. Rowlands, T.R.H. Cutmore, and D. James. Automated ECG diagnostic p-wave analysis using wavelets. *Computer Methods and Programs in Biomedicine*, 101(1):33–43, January 2011.

Jerome H. Friedman. Stochastic gradient boosting. *Computational Statistics Data Analysis*, 38(4):367–378, 2002.

A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000 (June 13). Circulation Electronic Pages: http://circ.ahajournals.org/content/101/23/e215.full PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.

Awni Y. Hannun, Pranav Rajpurkar, Masoumeh Haghpanahi, Geoffrey H. Tison, Codie Bourn, Mintu P. Turakhia, and Andrew Y. Ng. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, 25(1):65–69, January 2019.

Tin Kam Ho. Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition*, 1995.

Mohammad Kachuee, Shayan Fazeli, and Majid Sarrafzadeh. Ecg heartbeat classification: A deep transferable representation. *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, 2018.

S. C. Saxena, V. Kumar, and S. T. Hamde. Feature extraction from ECG signals using wavelet transforms for disease diagnostics. *International Journal of Systems Science*, 33(13):1073–1085, January 2002.

Rajesh Kumar Tripathy and Samarendra Dandapat. Automated detection of heart ailments from 12-lead ECG using complex wavelet sub-band bi-spectrum features. *Healthcare Technology Letters*, 4(2):57–63, April 2017.

Marcel van Gerven and Sander Bohte. Editorial: Artificial neural networks as models of neural information processing. *Frontiers in Computational Neuroscience*, 11, December 2017.