Arrhythmia Classification of 12-lead Electrocardiograms by Hybrid Scattering–LSTM Networks

Philip A. Warrick[1], Vincent Lostanlen[2], Michael Eickenberg[3], Joakim Andén[4], Masun Nabhan Homsi[5]

[1]PeriGen. Inc., Montreal, Canada
[2]New York University, New York, NY
[3]Flatiron Institute, New York, NY
[4]KTH Royal Institute of Technology
[5]Helmholtz Centre for Environmental Research (UFZ), Leipzig, Germany

Abstract

Electrocardiogram (ECG) analysis is the standard of care for the diagnosis of irregular heartbeat patterns, known as arrhythmias. This paper presents a deep learning system for the automatic detection and multilabel classification of arrhythmias in ECG recordings. Our system composes three differentiable operators: a scattering transform (ST), a depthwise separable convolutional network (DSC), and a bidirectional long short-term memory network (BiLSTM). The originality of our approach is that all three operators are implemented in Python. This is in contrast to previous publications, which precomputed ST coefficients in MATLAB. The implementation of ST on Python was made possible by using a new software library for scattering transform named Kymatio. This paper presents the first successful application of Kymatio to the analysis of biomedical signals. As part of the PhysioNet/Computing in Cardiology Challenge 2020, we trained our hybrid Scattering–LSTM model to classify 27 cardiac arrhythmias from two databases of 12–lead ECGs: CPSC2018 and PTB-XL, comprising 32k recordings in total. Our team "BitScattered" achieved a Challenge metric of 0.536±0.012 over ten folds of cross-validation but this result may be over-optimistic since we were not able to rank and score on the hidden test set.

1. Introduction

The World Health Organization estimates that cardiovascular diseases (CVDs) caused 17.9 million deaths worlwide in 2016, and may reach 23.6 million in the year 2030. In this context, electrocardiography (ECG) plays a vital role in CVD prevention, diagnosis, and treatment. This is because each electrode in an ECG can reveal cardiac abnormalities, which are risk factors for CVDs.

The main advantage of ECG is that its acquisition is inexpensive, painless, and non-invasive. However, the subsequent task of interpreting an electrocardiograph is tedious and time-consuming. To address this issue, the PhysioNet/Computing in Cardiology Challenge 2020 offers an evaluation benchmark for automatic detection and classification of cardiac abnormalities from 12–lead ECGs.

Prior literature on ECG classification exhibits a methodological divide: signal processing versus machine learning. On one hand, digital signal processing methods include low-pass filters, fast Fourier Transform, and discrete wavelet transform. On the other hand, machine learning methods include random forests, support vector machines, convolutional neural networks and long short-term memory (LSTM) networks. Both families of methods have their limitations: while feature engineering lacks flexibility to represent fine-grain class boundaries, a purely learned pipeline may lead to uninterpretable overfitting.

Our contribution to the Challenge aims to overcome the aforementioned methodological divide by combining insights from signal processing and machine learning. At a first stage, we extract time scattering transform (ST) coefficients for each ECG channel. Although this stage is not trainable, it offers numerical guarantees of stability to time warps. At a second stage, we train a depthwise separable convolution (DSC) neural network, followed by a bidirectional long short-term memory (BiLSTM) network. While the DSC combines scattering coefficients from multiple electrodes simultaneously, the BiLSTM can also capture longer-term trends in cardiac activity.

Our system is inspired from a previous publication, which aimed at detecting and classifying sleep arousals from polysomnographic recordings (1). However, whereas the original publication resorted to a combination of programming languages (MATLAB for scattering and Python for machine learning), we implement all stages of compu-

 ISSN: 2325-887X DOI: 10.22489/CinC.2020.462

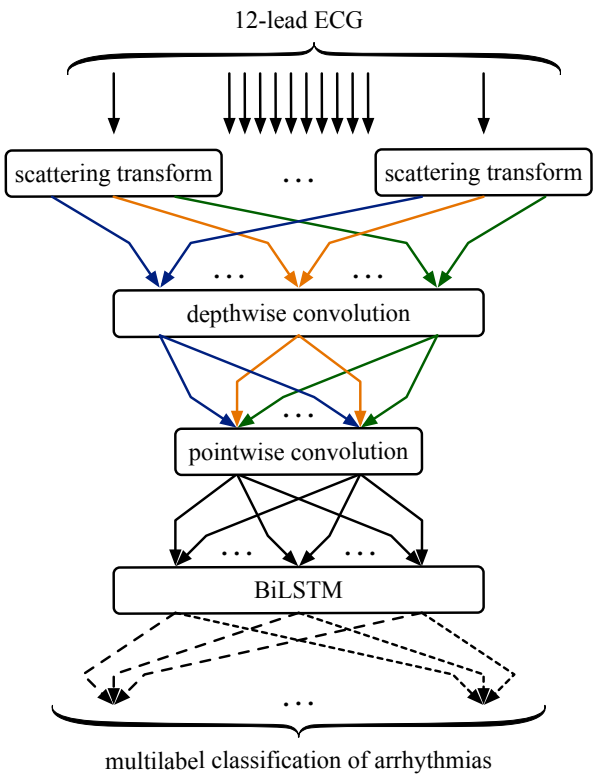

Figure 1. Overview of the proposed system. Top: channel-wise scattering transform of a 12-lead electrocardiogram (ECG). For simplicity, only two scattering transform blocks are shown. Arrow colors denote scattering paths. Middle: depthwise separable convolutional neural network (DSC), separated into a depthwise convolution layer and a pointwise convolution layer. For simplicity, only three scattering paths are shown and only two feature maps are shown. Bottom: bidirectional long short-term memory network (BiLSTM) followed by multilabel classification. Arrow styles denote output units. For simplicity, only three BiLSTM hidden units are shown, and only two arrhythmia classes are shown. See Section 2 for details.

tation in Python. This is possible thanks to a new library for scattering transforms in Python, named Kymatio.[1]

2. Methods

Figure 1 summarizes the different technical components of our proposed system. This section explains the role of each component in isolation.

[1] Official website of Kymatio: https://www.kymat.io

2.1. Scattering transform

The scattering transform is a deep convolutional network whose filters are defined a priori instead of being learned from data. Specifically, every layer in the scattering network contains filters of the form:

$$\boldsymbol{\psi}_j : t \longmapsto 2^{-j/Q}\boldsymbol{\psi}(2^{-j/Q}t), \tag{1}$$

where $\boldsymbol{\psi}$ is a wavelet, Q is a constant number of filters per octave, and the scale variable j is an integer ranging between 0 and J. Hereafter, we take the "mother wavelet" $\boldsymbol{\psi}$ to be a Morlet wavelet with a quality factor of $Q = 1$ and a center frequency of $\xi = 200\,\mathrm{Hz}$. The Morlet wavelet is a complex-valued function with a Gaussian envelope while being approximately analytic, i.e., with negligible Fourier coefficients outside of the half-line of positive frequencies ($\omega > 0$) . Furthermore, we set the maximum wavelet scale to $J = 11$ after a process of trial and error.

Let $\boldsymbol{\phi}_T$ be a Gaussian filter of cutoff frequency equal to $1/T$. The outputs of the scattering transform at orders one and two respectively are

$$\mathbf{S_1}\boldsymbol{x}(t, j_1) = |\boldsymbol{x} * \boldsymbol{\psi}_{j_1}| * \boldsymbol{\phi}_T(t) \quad \text{and}$$
$$\mathbf{S_2}\boldsymbol{x}(t, j_1, j_2) = \Big||\boldsymbol{x} * \boldsymbol{\psi}_{j_1}| * \boldsymbol{\psi}_{j_2}\Big| * \boldsymbol{\phi}_T(t), \tag{2}$$

where the vertical bars and the asterisk denote complex modulus and convolution product respectively.

The earliest application of the scattering transform to cardiology is due to (2), in the context of fetal heart rate classification. We refer to (3) for a mathematical introduction and to (1) for a recent review of the state of the art.

For every discretized value of time t, we concatenate first-order coefficients $\mathbf{S_1}\boldsymbol{x}(t, j_1)$ and second-order coefficients $\mathbf{S_1}\boldsymbol{x}(t, j_1, j_2)$ to produce a multidimensional time series $\mathbf{S}\boldsymbol{x}(t, p)$; where the multiindex p, known as scattering *path*, either denotes an singleton (j_1) or a pair (j_1, j_2). With $J = 11$, this results in 12 first-order and 63 second-order paths for a total number of $P = 75$ paths.

Hereafter, we set the time scale of Gaussian averaging to $T = 250\,\mathrm{ms}$. Note that T is not equal to $2^J/\xi$, as is customary. Rather, the filterbank $\{\psi_j\}_j$ covers the frequency range $[2^{-J}\xi; \xi] = [0.1\,\mathrm{Hz}; 200\,\mathrm{Hz}]$ whereas the scattering transform is discretized at a Nyquist rate of $2/T = 8\,\mathrm{Hz}$. This rate is chosen to be higher than the heart rate of patients (1–4 Hz) while being considerably lower than the acquisition rate of ECG (500 Hz).

We apply a pointwise compressive nonlinearity to the output of the ST, namely the inverse hyperbolic sine function: $\mathrm{asinh} : \mathrm{x} \mapsto \log(\mathrm{x} + \sqrt{\mathrm{x}^2 + 1})$. Previous literature has shown that such compressive nonlinearities can bring the empirical histogram of scattering transform magnitudes closer to Gaussian and improve classification accuracy (4). Figure 2 illustrates the scattering transform of an ECG channel sample, for the first two orders.

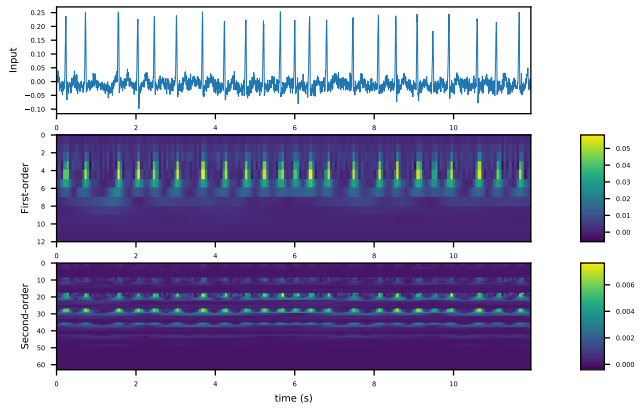

Figure 2. Scattering transform results for A0004 channel 1 ECG recording. From top to bottom: normalized input signal, 12 first-order ST paths, 63 second-order ST paths.

2.2. Depthwise separable convolution

A depthwise separable convolution (DSC) splits the computation into two operations: depthwise convolution applies a single convolutional filter per each ST input channel while the pointwise convolution linearly combines these transformed channels. Equations (3) and (4) describe the two steps mathematically.

$$X\left[p\right] = \sum_{e} S\left[e,\ p\right] F\left[p, e\right] \tag{3}$$

$$Y\left[n\right] = \rho\left[B\left[n\right] + \sum_{p} X\left[p\right] G\left[p, n\right]\right] \tag{4}$$

where E and P represent electrodes and paths, respectively. F and G refer to the filter maps, N number of paths, B is the bias and ρ represents the activation function. The total number of convolution coefficients including the bias weights is therefore $P \times E + (P+1) \times N$. This is often a reduction in parameters compared to regular convolution.

We used a DSC layer with $N = P = 66$ (chosen to be on the order of the number of paths) and the rectified linear activation function (ReLU).

2.3. Long-short term memory (LSTM)

An LSTM is a type of Recurrent Neural Network (RNN) specially designed to model temporal sequences. It preserves information from inputs that has already passed through it using the hidden state. An LSTM unit contains one or more self-connected memory cells and three gates for the input, output and forget units that provide continuous analogues of write, read and reset operations for the cells. Given a sequence $x = (x_1, ..., x_T)$, an LSTM computes the hidden vector sequence $h = (h_1, ..., h_T)$ and output vector sequence $y = (y_1, ..., y_T)$ by repeating the following equations from $t = 1$ to T:

$$\begin{aligned} h_t &= H(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \\ y_t &= W_{hy}h_t + b_h \end{aligned} \tag{5}$$

where W, b and H denote weights, bias vectors and hidden activation, repectively, implemented as follows:

$$\begin{aligned} i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \\ f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \\ c_t &= f_t c_{t-1} + i_t tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\ o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \\ h_t &= \sigma \tanh(c_t) \end{aligned} \tag{6}$$

where W_{xo} is the input-output gate matrix, W_{hi} is the hidden-input gate matrix, i, f, o and c represent respectively the input gate, forget gate, output gate and cell activation vectors, and σ denotes the logistic sigmoid function.

Bidirectional LSTMs (BiLSTM) process data in forward and reverse directions to capture both past and future contexts with two separate hidden layers, which are then fed forward to the same output layer. Our system composes two layers of BiLSTM, each containing 100 hidden units.

2.4. Prediction

The sequence learning applies the BiLSTM output to a final dense layer with sigmoid activation to calculate the final predictions. The Adam optimizer algorithm was used to handle sparse gradients on noisy ECG data.

The dense layer used binary cross-entropy loss during training to support multiple arrhythmia classes. Predictions were averaged over time and our decision rule chose any class that exceeded the probability threshold $p = 0.5$; otherwise the maximum probability class was chosen.

2.5. Data

The PhysioNet/CinC Challenge 2020 database includes 43,101 annotated ECG recordings. The training data was obtained from four sources publicly, while the hidden testing data comes from three sources (5):

- CPSC2018: the China Physiological Signal Challenge 2018 (CPSC2018). This source includes two databases: a public training dataset (CPSC) and unused data (CPSC2).
- St. Petersburg Institute of Cardiological Technics.
- PTB and PTB-XL: the Physikalisch Technische Bundesanstalt (PTB) Database, Brunswick, Germany.
- G12EC: The Georgia 12-lead ECG Challenge (G12EC) Database, Emory University, Atlanta, Georgia, USA.
- Undisclosed: this is hidden test data that comes from an American institution.

2.6. Implementation

Keras with Tensorflow as backend was used for building the neural networks. We used a machine with 32 GB of system memory and two GPUs with 12 GB of memory each, using Tensorflow class `MirroredStrategy` to distribute the training amongst the two GPUs.

We used the analog-to-digital gain (ADG) in the input headers to obtain physical (mV) units.

Although the longest ECG recording in CPSC dataset was $60\,s$, to reduce computational requirements, we reduced the time span of the learning batches to 30 s. ST paths of sequences longer than this were truncated at 30 s.

We collapsed the three pairs of equivalent classes, as described in (5), to obtain a total of 24 target classes for training. In addition, we applied a padding target for recordings of duration less than the batch size to remove their unused samples from participation in the loss function.

Training data was randomly split into training (90%) and testing datasets (10%). 10% of the training dataset was used as a validation set, using 20 early stopping epochs. Final model selection was based on ten-fold cross-validation results with various configurations. We used the Challenge metric $s_{normalized}$ described in (5) to assess performance.

We found that training converged in isolation for the CPSC, CPSC2 and PTB-XL datasets. The other datasets did not converge and so we used the composite dataset CPSC, CPSC2 and PTB-XL in subsequent cross-validation. Therefore, our experiments used 32,167 of the available 43,101 ECG recordings, or 75%. Due to time limitations, we were not able to submit a working submission using the docker environment and therefore we could not report results on the hidden test set.

3. Results

The results for our final model are shown in Table 1. Training stretched the limits of our machine configuration with per-fold training times of approximately 18 hours.

4. Discussion

Our approach achieved experimental success without need for feature engineering and with few parameters to select. Nevertheless, future research is needed to perform a wider search of key hyperparameters, especially the maximum ST scale, the downsampling factor, the number of DSC filters, and the number of LSTM units and layers.

We note that our approach did not generalize to all datasets. In particular, using ADG may not have been reliable for the large Georgia dataset, as was discussed on the Challenge discussion forum. In addition, one of the most frequent classes, namely sinus rhythm, has a poor classification rate: this fact merits further investigation.

Fold	Training	Validation	Testing
1	0.600	0.548	0.541
2	0.579	0.517	0.533
3	0.580	0.530	0.527
4	0.617	0.570	0.552
5	0.585	0.530	0.543
6	0.648	0.542	0.545
7	0.601	0.561	0.528
8	0.602	0.542	0.549
9	0.577	0.525	0.512
10	0.572	0.534	0.527
$\mu \pm \sigma$	0.569±0.014	0.540±0.016	0.536±0.012

Table 1. $s_{\text{normalized}}$ metric for each partition per fold; mean μ and standard deviation σ over all folds.

In future work we wish to complete debugging of the docker submission container.

Acknowledgments

The authors would like to acknowledge the computational facilities provided by PeriGen Inc. for this work.

References

[1] Warrick PA, Lostanlen V, Homsi MN. Hybrid scattering-LSTM networks for automated detection of sleep arousals. Physiological Measurement July 2019; 40(7):074001.

[2] Chudáček V, Andén J, Mallat S, Abry P, Doret M. Scattering transform for intrapartum fetal heart rate variability fractal analysis: A case-control study. IEEE Transactions on Biomedical Engineering 2014; 61(4):1100–1108.

[3] Mallat S. Understanding deep convolutional networks. Philosophical Transactions of the Royal Society A Mathematical Physical and Engineering Sciences 2016;374(2065):20150203.

[4] Lostanlen V, Lafay G, Andén J, Lagrange M. Relevance-based quantization of scattering features for unsupervised mining of environmental audio. EURASIP Journal on Audio Speech and Music Processing 2018;2018(1):15.

[5] Perez Alday EA, Gu A, Shah A, Robichaux C, Wong AKI, Liu C, Liu F, Rad BA, Elola A, Seyedi S, Li Q, Sharma A, Clifford GD, Reyna MA. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. Physiological Measurement 2020;under review.

Address for correspondence:

philip.warrick@perigen.com