Arrhythmia Classification of Reduced-Lead Electrocardiograms by Scattering-Recurrent Networks

Philip A. Warrick[1,2], Vincent Lostanlen[3], Michael Eickenberg[4], Masun Nabhan Homsi[5]
Adrián Campoy Rodríguez[6], Joakim Andén[4,6]

[1]PeriGen Inc., Montreal, Canada [2]McGill University, Montreal, Canada
[3]LS2N, CNRS, École Centrale de Nantes, France, [4]Flatiron Institute, New York, NY, USA
[5]Helmholtz Centre for Environmental Research (UFZ), Leipzig, Germany
[6]KTH Royal Institute of Technology, Stockholm, Sweden

Abstract

We describe an automatic classifier of arrythmias based on 12-lead and reduced-lead electrocardiograms. Our classifier composes the scattering transform (ST) and a long short-term memory (LSTM) network. It is trained on PhysioNet/Computing in Cardiology Challenge 2021 data. The ST captures short-term temporal ECG modulations while reducing its sampling rate to a few samples per typical heart beat. We pass the output of the ST to a depthwise-separable convolution layer which combines lead responses separately for each ST coefficient and then combines resulting values across ST coefficients. At a deeper level, 2 LSTM layers integrate local variations of the input over long time scales. We train in an end-to-end fashion as a multilabel classification problem with a normal and 25 arrhythmia classes. We used canonical correlation analysis (CCA) for transfer learning from 12-lead ST representations to reduced-lead ones. For 12-, 6-, 4-, 3- and 2-leads, team "BitScattered" Challenge metrics on the hidden validation set were 0.46, 0.44, 0.45, 0.46 and 0.43; and on the hidden test set were 0.10, 0.11, 0.10, 0.10 and 0.10, respectively, ranking 34^{th} on the hidden test set.

1. Introduction

The World Health Organization estimates that cardiovascular diseases (CVDs) caused 17.9 million deaths worldwide 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 and non-invasive. However, the visual interpretation of ECG is tedious, time-consuming, and requires expert knowledge. To address this, the PhysioNet/Computing in Cardiology Challenge 2021 offers a benchmark for automatic classification of cardiac abnormalities from 12-lead and reduced-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 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. 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 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) network, followed by a bidirectional long short-term memory (BiLSTM) network. While DSC combines scattering coefficients from multiple leads simultaneously, the BiLSTM can also capture longer-term trends in cardiac activity. We also investigated transfer learning to the reduced-lead models using canonical correlation analysis (CCA). Our system extends previous Challenge work(1) and is inspired from previous publications, which aimed at detecting sleep arousals from polysomnographic recordings (2).

2. Methods

Figure 1 summarizes our proposed system; this section explains the role of each system component in isolation.

 ISSN: 2325-887X DOI: 10.22489/CinC.2021.210

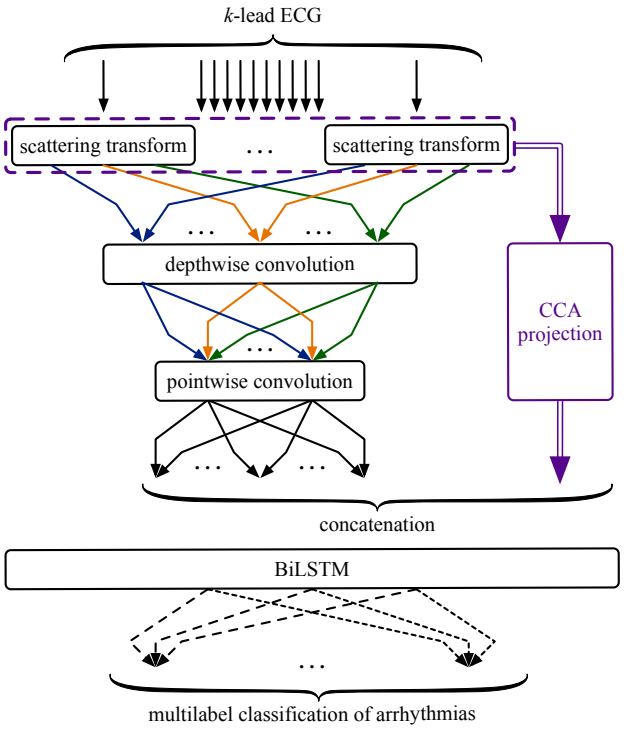

Figure 1. System overview for k-lead ECG. Top: channel-wise scattering transform. Arrow colors denote scattering paths. Middle left: depthwise separable convolutional neural network, separated into depthwise and pointwise convolution layers. Middle (optionally, in violet): CCA projection of k-lead scattering uses coefficients precalculated from 12-lead data before training. Bottom: bidirectional long short-term memory network (BiLSTM) followed by classification. Arrow styles denote output units.

2.1. Scattering transform

The scattering transform is a deep convolutional network whose filters are defined a priori instead of being learned from data. We refer to (3) for a mathematical introduction and to (2) for a recent review of the state of the art. Specifically, every layer contains filters of the form

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

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 first two orders of the scattering transform 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.

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.

To control the degree of time invariance, we modified the Python scattering package Kymatio[1] to set the time scale of Gaussian averaging to $T = 62.5\,\mathrm{ms}$. Note that this T is less than the customary $2^J/\xi$. 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 $2/T = 32\,\mathrm{Hz}$. This rate is chosen to be higher than typical patient heart rates yet considerably lower than the ECG acquisition rate ($500\,\mathrm{Hz}$).

We apply a pointwise compressive nonlinearity to the output of the ST, namely an offset log function: $\log(x + \epsilon)$ where $\epsilon = 1\mathrm{e}{-4}$. Then per-path normalization subtracted the mean and divided by the standard deviation. Figure 2 illustrates the scattering transform of normal and atrial fibrillation ECG recordings, for the first two orders.

2.2. Depthwise separable convolution

A depthwise separable convolution (DSC) splits the computation into two operations: depthwise convolution X linearly combines the leads for each ST path while the pointwise convolution Y linearly combines these transformed paths, as in equations (3) and (4)

$$X\,[p] = \sum_{l=1}^{L} S\,[l,\ p]\,F\,[p, l] \tag{3}$$

$$Y\,[n] = \beta\left[B\,[n] + \sum_{p=1}^{P} X\,[p]\,G\,[p, n]\right] \tag{4}$$

where $L \in \{12, 6, 4, 3, 2\}$ and P represent the number of leads and paths, respectively. F and G refer to the filter maps, N is the number of pointwise mixes, B is the bias and β represents the activation function. The total number of convolution coefficients including the bias weights is therefore $P \times L + (P + 1) \times N$. This is often a reduction

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

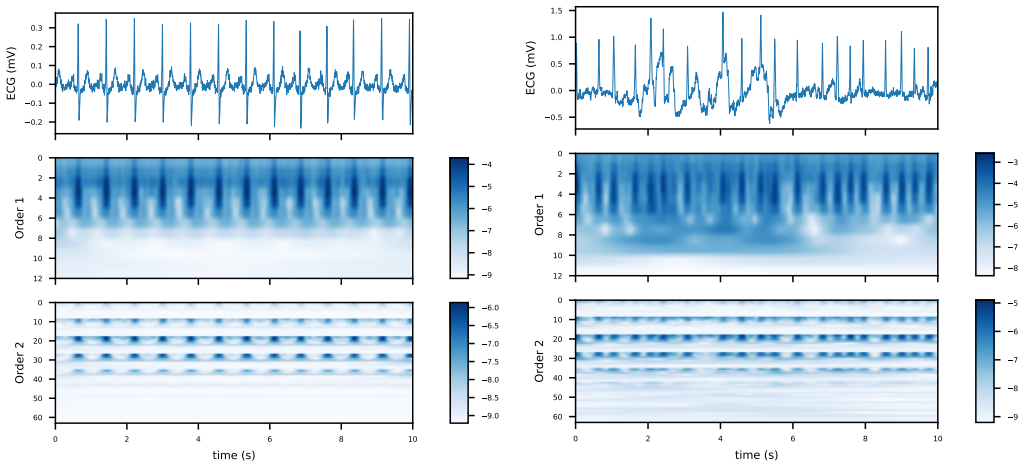

Figure 2. Scattering transform of ECG lead I 10 s recordings for (left) normal sinus rhythm (A0002) and (right) atrial fibrillation (A0023). Top to bottom: input ECG (linear scale), 12 first-order and 63 second-order ST paths (log scales).

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 ReLU activation.

2.3. Transfer learning for reduced-lead models

For reduced-lead models, we apply transfer learning from the 12-lead data using canonical correlation analysis (CCA). CCA finds a pair of linear transformations for two sets of multidimensional variables (views S_i), such that the linear projections of the two views, $(S_1 w_1, S_2 w_2)$ are maximally correlated (4). In our case, view S_i is the scattering of lead sets: S_1 corresponds to the lead set used for prediction (2, 3, 4 or 6 leads) and S_2 corresponds to the respective complements (10, 9, 8 and 6 leads). This is done by maximizing the following equation:

$$\begin{aligned} \rho &= \max_{w_1, w_2} \operatorname{corr}\left(S_1 w_1, S_2 w_2\right) \\ &= \max_{w_1, w_2} \frac{w_1^T \Sigma_{12} w_2}{\sqrt{w_1^T \Sigma_{11} w_1 w_2^T \Sigma_{22} w_2}} \end{aligned} \tag{5}$$

where Σ_{11}, Σ_{22} and Σ_{12} are the covariances and cross-covariance of S_1 and S_2; and w_1 and w_2 are determined by singular-value decomposition.

We calculate w_1 and w_2 from fold training data prior to network training. CCA uses S_1 and S_2 to find the projection vectors corresponding to the k highest left- and right-singular values, and $k = P \times L$ was chosen to include all the singular values.

During training and prediction, S_1 is projected with fixed w_1. This projection is intended to transfer information from (possibly unavailable) S_2, correlated with the complementary lead set, such that classification of reduced-lead ECG records is improved.

2.4. Data

The PhysioNet/CinC Challenge 2021 data (5) includes 88,000 public and 26,000 private ECG records. Each record is assigned one or more diagnosis by experts. We excluded subsets St. Petersburg having long durations (30 min) and PTB having non-uniform acquisition rates and low class coverage (5/26).

2.5. Implementation

Although the ECG recording lengths in the training set were as long as 120 s, the vast majority (78,181 of 87,663≈89%) were 10 s or less. Therefore to reduce computational requirements, we reduced the time span of the learning batches to 10 s. Longer recordings were truncated at 10 s or split into multiple training sub-sequences of 10 s. We applied a padding target for sub-sequences of duration less than 10 s to remove their unused samples from participation in the loss function. 24 Georgia and 388 Ningbo records were omitted from training because NaN values in the ECG recordings prevented convergence.

We used two BiLSTM layers of 100 hidden units. The dense layer used binary cross-entropy loss to support multiple classes. Predictions were averaged over time, and over sub-sequences if present. Our decision rule chose any class that exceeded probability threshold $p = 0.5$; otherwise the maximum probability class was chosen.

The 10-fold cross-validation data partitions were 90% training and 10% testing for each fold. The validation set, 10% of training, was used for early stopping (60 epochs).

Model	Cross-validation	Validation	Test
Base_{12}	0.601 ± 0.015	0.46 (23)	0.10 (34)
Base_{6}	0.584 ± 0.007*	0.44 (23)	0.11 (33)
CCA_{6}	0.573 ± 0.010		
Base_{4}	0.582 ± 0.015	0.45 (21)	0.10 (34)
CCA_{4}	0.581 ± 0.009		
Base_{3}	0.583 ± 0.006	0.46 (21)	0.10 (34)
CCA_{3}	0.576 ± 0.009		
Base_{2}	0.570 ± 0.008	0.43 (22)	0.10 (34)
CCA_{2}	0.564 ± 0.010		

Table 1. Challenge metric (5; 6) for baseline and CCA lead models on cross-validation and hidden validation and test sets (ranking out of 40 teams in parentheses). * indicate two-sided t-test $p < 0.01$ compared to previous row.

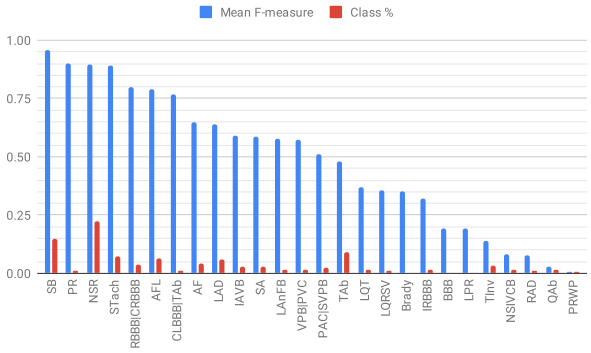

Figure 3. Class incidence (red) and cross-validation performance (F-measure, blue) for model Base_{12}. The highest incidence class was normal sinus rhythm (NSR).

3. Results

Table 1 shows cross-validation and hidden validation results for our baseline and CCA models using 10 s truncation. Our submitted entry completed training of the baseline models in just over 18 h and prediction of the hidden validation set in 18 min, within the maximum allowable times of 48 h and 24 h, respectively. Fig 3 shows the class incidence in the training data and cross-validation performance for the 12-lead model Base_{12}.

4. Discussion

Our approach achieved experimental success without need for feature engineering and with few parameters to select. We observe slight performance degradation for models with decreasing numbers of leads, suggesting that the correlation between leads is considerable. CCA did not significantly improve results but warrants further analysis.

Performance was only somewhat reduced on the hidden validation set compared to cross-validation, indicating good generalization; however hidden test set performance was drastically low in comparison, suggesting that the test set was quite different from the training and validation sets or that there was a systemic error in our processing of the test set. We anticipate that the organizers will provide further information to better assess these results.

Splitting rather than truncating records did not affect results, although benefits may have been masked by the preponderance of 10 s recordings.

We note that higher incidence classes tended to perform better, especially normal sinus rhythm (28,891 of 87,663 training records≈33%). Considering this imbalance could improve results for low incidence classes.

Extensions to our approach to explore include: improving the decision rule; exploring alternate loss functions; searching hyperparameters; and using age and sex demographic data, recognized risk factors for cardiac pathology.

Acknowledgements

We acknowledge resources provided by the Swedish National Infrastructure for Computing and PeriGen Inc.

References

[1] Warrick PA, Lostanlen V, Eickenberg M, Andén J, Homsi MN. Arrhythmia Classification of 12-lead Electrocardiograms by Hybrid Scattering-LSTM Networks. Computing in Cardiology 2020;47:1–4.

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

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

[4] Hardoon DR, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: An overview with application to learning methods. Neural Computation 2004; 16(12):2639–2664. ISSN 08997667.

[5] Reyna MA, Sadr N, Perez Alday EA, Gu A, Shah A, Robichaux C, Rad BA, Elola A, Seyedi S, Ansari S, Li Q, Sharma A, Clifford GD. Will Two Do? Varying Dimensions in Electrocardiography: the PhysioNet/Computing in Cardiology Challenge 2021. Computing in Cardiology 2021;48:1–4.

[6] 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;41.

Address for correspondence: philip.warrick@perigen.com