Phonocardiographic Murmur Detection by Scattering-Recurrent Networks

Philip A Warrick[1,2], Jonathan Afilalo[2]

[1]PeriGen Inc., Montreal, Canada, [2]McGill University, Montreal, Canada

Abstract

We describe an automatic detector of phonocardiogram murmurs. Our detector composes the scattering transform (ST) and a long short-term memory (LSTM) network. It is trained on PhysioNet/Computing in Cardiology Challenge 2022 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 transforms 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 classification problem with three murmur classes: present, absent or unknown. Additionally, we use the model to classify clinical outcome as normal or abnormal. These two classifications determine whether clinical followup should occur. Our team "PAWPCG" Challenge scores on the hidden validation set were a weighted accuracy of 0.557 for murmur classification and 13835.503 for the clinical outcome cost.

1. Introduction

The phonocardiogram (PCG) is a recording of heart sound that provides a low-cost, non-invasive diagnostic tool to assess turbulent blood flow patterns ("murmurs") suggestive of heart valve dysfunction. Physicians' subjective interpretation of PCG sounds has historically low sensitivity and specificity for the diagnosis of pathological heart murmurs. To improve diagnostic performance, we developed a PCG heart murmur detector using the Scattering Transform (ST) and Long Short-Term Memory (LSTM) networks. The PhysioNet/Computing in Cardiology Challenge 2021 offers a benchmark for automatic detection of heart murmurs from PCG.

Prior literature on PCG 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 PCG recording. 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 local scattering coefficients, the BiLSTM can capture longer-term trends in heart-sound activity. Our system is inspired from previous Challenge work on ECG arrhythmia detection(1).

2. Methods

2.1. Data

The PhysioNet/CinC Challenge 2022 data (2) provides multiple PCG recordings for each patient. Each patient in the Challenge data has one or more recordings from one or more prominent auscultation locations: pulmonary valve (PV), aortic valve (AV), mitral valve (MV), tricuspid valve (TV), and other (Phc). The recordings were collected sequentially (not simultaneously) from different auscultation locations using a digital stethoscope. The number, location, and duration of the recordings vary between patients. The Challenge labels consist of two types: murmur-related labels indicate whether an expert annotator detected the presence or absence of a murmur in a patient from the recordings or whether the annotator was unsure about the presence or absence of a murmur. Outcome-related labels indicate the normal or abnormal clinical outcome diagnosed by a medical expert.

The dataset include recordings from 1568 patients, of which 942 patients with 3163 PCG recordings are used as the training dataset. The remaining 626 patient records are reserved as a hidden validation set for Challenge scoring.

2.2. Overall Processing

We presented the PCG signals to an ST representation layer, which reduced the sampling rate from f_{in}=4 kHz to f_{out}=250 Hz, and then 2 bidirectional LSTM layers, which captured feature trajectories over time. The final dense layer used cross-entropy loss during training to support three target classes: murmur present, absent or unknown. During training, we used the "murmur location" information to assign targets to corresponding recordings. We split recordings into 5 s segments for training batches. During evaluation, we classified each recording by selecting the highest average class probability over all segments for that recording. If any recording was classified as "murmur present" for a given patient, that patient was assigned a diagnosis of "murmur present". The system was developed with TensorFlow and the Kymatio ST package.

2.3. 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 (4) 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 $\xi = 1490$ Hz and a frequential width of σ=155 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

$$\begin{aligned} \mathbf{S_1}\boldsymbol{x}(t, j_1) &= |\boldsymbol{x} * \boldsymbol{\psi}_{j_1}| * \boldsymbol{\phi}_T(t) \quad \text{and} \\ \mathbf{S_2}\boldsymbol{x}(t, j_1, j_2) &= \left||\boldsymbol{x} * \boldsymbol{\psi}_{j_1}| * \boldsymbol{\psi}_{j_2}\right| * \boldsymbol{\phi}_T(t), \end{aligned} \tag{2}$$

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 $T = 7.81$ ms. Note that this T is less than the customary $2^J/\xi$. Rather, the filterbank $\{\psi_j\}_j$ covers the frequency range $[2^{-J}\xi; \xi] = [0.73\,\text{Hz}; 1490\,\text{Hz}]$ whereas the scattering transform is discretized at a Nyquist rate of $2/T = 256$ Hz. This rate is chosen to be higher than typical patient heart rates yet considerably lower than the PCG acquisition rate (4 kHz).

2.4. Depthwise separable convolution

A depthwise separable convolution (DSC) splits the computation into two operations: depthwise convolution X linearly filters the PCG recording 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 $L \in \{1\}$ and P represent the number of recordings 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 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.

2.5. Implementation

The PCG recording lengths in the training set were of various durations. Therefore to reduce computational requirements, we reduced the time span of the learning batches to 5 s. Longer recordings were split into multiple training sub-sequences of 5 s. We applied a padding target for sub-sequences of duration less than 5 s to remove their unused samples from participation in the loss function.

We used two BiLSTM layers of 100 hidden units. The dense layer used cross-entropy loss to support multiple classes. Predictions were averaged over time for each subsequence. If any recording was classified as "murmur present" for a given patient, that patient was assigned a diagnosis of "murmur present". In this phase of system development, we simply set the outcome class to abnormal if a murmur present diagnosis was assigned.

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

Murmur	Local CV	Validation
AUROC	0.849±0.047	
AUPRC	0.695±0.06	
F-measure	0.569±0.037	
Accuracy	0.863±0.031	
WtAccuracy	0.753±0.078	0.557
Cost	18477±2626	

Outcome	Local CV	Validation
AUROC	0.323±0.035	
AUPRC	0.406±0.021	
F-measure	0.283±0.031	
Accuracy	0.359±0.05	
WtAccuracy	0.584±0.061	
Cost	15397±830	13836

Table 1. Challenge metrics (2) for murmur and outcome classification.

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 (20 epochs).

3. Results

As shown in Table 1, local 10-fold cross-validation test results achieved a weighted accuracy of 0.753±0.078 (mean±standard deviation) for murmur classification and 15397±830 for the clinical outcome cost. Our entry ("team PCGPAW") successfully trained a model on the Challenge server and obtained a weighted accuracy of 0.557 for murmur classification and 13836 for the clinical outcome cost on the hidden validation data. Our submitted entry used the GPU and completed training of the baseline models in just over $71\,\mathrm{min}$ and prediction of the hidden validation set in $2.5\,\mathrm{min}$, within the maximum allowable times of $48\,\mathrm{h}$ and $24\,\mathrm{h}$, respectively.

4. Discussion

Our classifier architecture showed promising results in this first phase of development. In future work, we hope to improve the system by including more of the available training annotations, notably, the segmentation of diastole and systole timing.

Extensions to our approach to explore include: using the demographic data, improving the decision rule, and searching hyperparameters.

Acknowledgements

We acknowledge computing resources provided by PeriGen Inc.

References

[1] Warrick PA, Lostanlen V, Eickenberg M, Homsi MN, Campoy Rodriguez A, Andén J. Arrhythmia classification of 12-lead and reduced-lead electrocardiograms via recurrent networks, scattering, and phase harmonic correlation. Physiological Measurement 2022;URL http://iopscience.iop.org/article/10.1088/1361-6579/ac77d1.

[2] Reyna MA, Kiarashi Y, Elola A, Oliveira J, Renna F, Gu A, Perez-Alday EA, Sadr N, Sharma A, Mattos S, Coimbra MT, Sameni R, Rad AB, Clifford GD. Heart murmur detection from phonocardiogram recordings: The George B. Moody PhysioNet Challenge 2022. medRxiv 2022;URL https://doi.org/10.1101/2022.08.11.22278688.

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

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

Address for correspondence: philip.warrick@perigen.com