Heart Murmur Detection Using Wavelet Time Scattering and Support Vector Machines

Adrian K Cornely, Grace M Mirsky

Benedictine University, Lisle, IL, USA

Abstract

The lack of infrastructure and specialized clinicians have led to reduced diagnoses and treatment of cardiac conditions for children in developing countries. An important tool for improved screening for cardiac abnormalities is the phonocardiogram (PCG). The PCG non-invasively records heart sounds, allowing for the observation of murmurs, which are abnormal heart sounds that may indicate underlying cardiac disease.

This study is part of the Heart Murmur Detection from Phonocardiogram Recordings: The George B. Moody PhysioNet Challenge 2022. For our approach, we employed wavelet time scattering and support vector machines to determine the classification of each PCG recording based on the features from the signals alone. The classification for each of the recording locations, as well as the demographic information (height, weight, age, sex), was then fed into a naïve Bayes classifier to determine the patient's overall classification.

For team Eagles, our murmur detection classifier received a weighted accuracy score of 0.467 on the validation set and a score of 0.525 (ranked 33 out of 40 teams) on the hidden test set. Our outcome classification achieved a cost of 10559 on the validation set, but we did not receive an official Challenge cost score for the hidden test set.

1. Introduction

The lack of infrastructure and specialized clinicians have led to reduced diagnoses and treatment of cardiac conditions for children in developing countries. Early detection is particularly important since increased morbidity and mortality are associated with delayed diagnosis of congenital heart disease (CHD) in children [1,2]. Consequently, insufficient clinical resources in these developing countries which lead to missed and/or delayed diagnoses have resulted in a significant public health concern [3-5].

An important low-cost tool for improved screening for cardiac abnormalities is the phonocardiogram (PCG). The PCG non-invasively records heart sounds, allowing for the observation of murmurs, which are abnormal heart sounds caused by turbulent blood that sometimes indicates underlying cardiac disease. Sample PCG signals are shown in Figure 1.
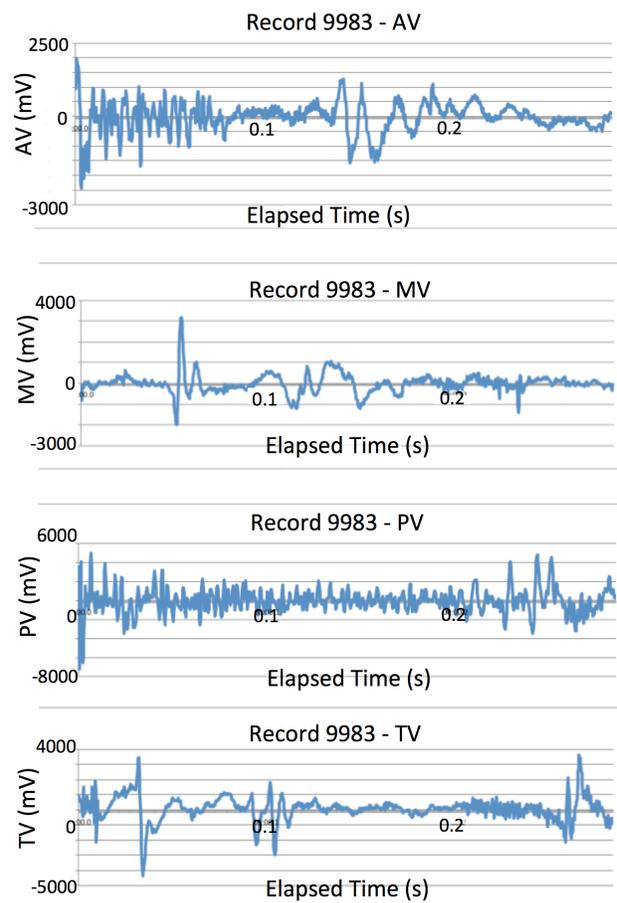
Figure 1. Phonocardiogram signals from four prominent recording locations: atrial valve (AV), mitral valve (MV), pulmonic valve (PV), and tricuspid valve (TV). Duration of each of the signals shown in the figure is 0.3 seconds; recordings are obtained sequentially from the different locations.

Murmurs in children are generally classified into two groups, according to whether or not the murmurs occur due to underlying structural issues of the heart. Murmurs that are not associated with underlying issues are called "innocent" or "functional" murmurs and do not typically require treatment. However, murmurs can also indicate the presence of defects in the septum, ductus arteriosus, and/or cardiac valves; these types of murmurs indicate structural issues and do typically require clinical intervention. Congenital heart valve disease frequently occurs due to malformed pulmonic or aortic valves [6]. These malformations may involve valves with an insufficient number of tissue flaps or an improper size or shape, or they may be missing a pathway through which blood may correctly flow. Congenital heart valve disease can cause regurgitation (valve does not completely close, resulting in backflow) or stenosis (valve does not completely open, resulting in insufficient blood flow through the valve).

The objective of the 2022 George B. Moody PhysioNet Challenge is to utilize multiple PCG recordings taken from a single patient to determine the presence, absence, or unclear case of murmurs for each patient as well as normal or abnormal clinical outcome [7,8]. By using a large publicly available dataset for training and a hidden test set, the robustness of the results is properly validated. Furthermore, by making the code submitted by all Challenge participants public, the results will be reproducible, allowing for rapid progress on this important research question.

2. Methods

2.1 Data Description

The training set consisted of 3163 recordings obtained from 942 patients in northern Brazil. All patients were 21 years old or younger at the time of the screening. Recordings were obtained sequentially from one or more of the typical auscultation locations: mitral valve (MV), aortic valve (AV), pulmonic valve (PV), and tricuspid valve (TV). An option for "other" location, meant to differentiate from these four valve locations, was also provided in the dataset. A complete description of the data collection process and the resulting dataset can be found in [9].

2.2 Pre-processing

In order to prepare the signals for analysis, several pre-processing steps were executed. First, we resampled the signals to 4000Hz to ensure that all the signals were analysed at a consistent rate. Next, we standardized the signal length to be five seconds; this involved either clipping longer signals or zero padding if the signals were too short. This pre-processing step is useful since fixed signal lengths are needed for various machine learning algorithms. Finally, we performed z-score normalization, with a range from -1 to 1, in order to have a consistent magnitude among the signals. These steps are summarized in Figure 2.
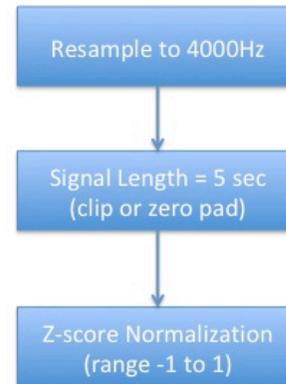
Figure 2. Pre-processing steps applied on the input signals prior to classification.

2.3 Challenge Algorithms

Wavelet scattering was used for feature extraction; this technique is useful in order to obtain low-variance features from time series data. These features tend to be stable with respect to time-warping deformations [10]. Furthermore, wavelet scattering has the benefit of only requiring the length of the scale invariant to be specified, which in this case we set equal to the length of the signal. The scattering network consists of two filter banks; the first contains eight wavelets per octave, and the second has one wavelet per octave.

The derived features were used to separately train four multiclass support vector machine (SVM) classifiers [11], one for each of recording locations based on the signals extracted from that location (MV, AV, PV, and TV). We note that not all patients had PCG signals recorded from all four locations. A linear kernel was utilized for these SVMs. The "other" location was excluded from this analysis, due to the potential variability in the actual physical recording location, making it unreasonable to group together signals from this class. SVMs were selected due to their ability to perform well for classification of small or medium sized datasets and scale well with the number of features, especially if they are sparse [12].

The results from each of the four SVM classifiers, as well as the corresponding patient's demographic information (height, weight, age group, sex), were then

fed into a naïve Bayes classifier in order to determine the patient's overall murmur classification. Demographic information can provide additional useful insights into appropriate patient classification, so they were included in this analysis. The same approach was taken in order to determine the clinical outcome. The algorithmic approach is summarized in Figure 3.

There are two scoring metrics used in the Challenge, which are described in detail in [7]. For murmur identification, a weighted accuracy score is computed. For this metric, higher scores indicate better performance. For outcome classification, a cost metric is computed, so lower values indicate better performance.
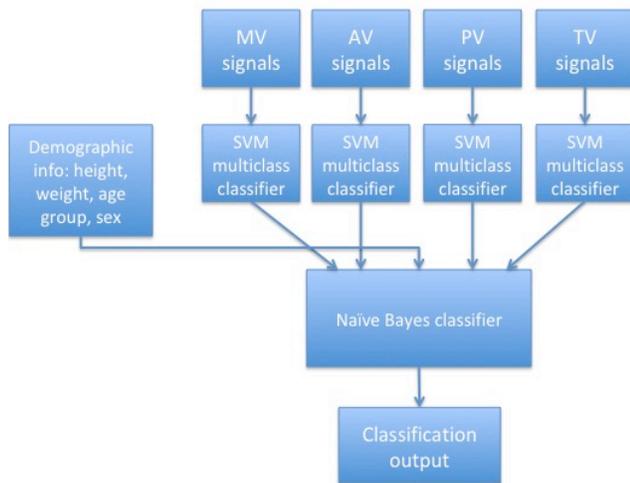
Figure 3. Algorithmic approach to the two Challenge classification tasks. Four signals (at most) are input into four separate SVMs; the outputs of the SVMs and the demographic info are fed into a naïve Bayes classifier to produce the final classification output.

3. Results

For murmur identification, the F-measure on the training set was 0.106, and the F-measure on the validation set was 0.102. For outcome classification, the F-measure on the training set was 0.933, and the F-measure for the validation set was 0.444. The training time for murmur identification was 4 minutes and 47 seconds; model run time was 28 seconds. The training time for outcome classification was 12 minutes and 37 seconds; model run time was 27 seconds.

In the Official Phase, the best performing entry on the validation set for team Eagles received a weighted accuracy score of 0.467 for murmur identification and a cost of 10559 for outcome classification. Our algorithm received a weighted accuracy score of 0.525 on the hidden test set for murmur identification, ranking 33rd

place out of 40 teams. Our algorithm failed on the hidden test set for the cost metric, so we do not have official results to report for that objective. These results are summarized in Tables 1 and 2.

Training	Validation	Test	Ranking
0.499	0.467	0.525	33/40

Table 1. Weighted accuracy metric scores for our final selected entry for the murmur detection task, including the ranking of our team on the hidden test set.

Training	Validation	Test	Ranking
5794	10559	N/A	N/A

Table 2. Cost metric scores for our final selected entry for the clinical outcome identification task. Our team did not receive a score on the hidden test set, so there is no official ranking.

Analysis of 305 successful entries from the leaderboard for murmur classification at the conclusion of the official phase revealed the following. The fastest training run was achieved by uestc-team, which computed in less than a second and is ranked as 178th out of 305 entries; the longest training run was submitted by PathToMyHeart, with a run time of over 68 hours and is ranked as 9th out of 305 entries. Furthermore, as expected, model run times were significantly shorter than training time, but there was still a very wide range. The fastest model run time was less than a second by team Simulab (182/305). The longest model run time was just over 13 hours by team listNto_urHeart (128/305). Weighted average scores ranged from 0.166 to 0.806, with a mean of 0.595 and standard deviation of 0.125, across the 305 submissions.

4. Discussion and Conclusions

The methods presented in this paper have the advantage of very fast compute time, which is particularly useful in terms of practical clinical utility. Specifically, for the validation set, the training time was 4 minutes and 47 seconds, and the model run time was 28 seconds. However, the classification accuracy for our method is much lower than most of the other teams who participated in the Challenge, so significant improvement is needed to increase the accuracy. Furthermore, we did not receive an official score for the cost metric for the test set; since we are unable to evaluate performance on this test set independently, the cause for the error is unknown.

References

[1] Murni, I. K., Wirawan, M. T., Patmasari, L., Sativa, E. R.,

Arafuri, N., & Nugroho, S. (2021). Delayed diagnosis in children with congenital heart disease: a mixed-method study. BMC pediatrics, 21(1), 1-7.

[2] Iqbal, S., Saidullah, S., Ahmed, R.I., Khan, M.A.A., Ahmed, N., & Khan, M. F. (2021). Factors Contributing to Delayed Diagnosis of Congenital Heart Disease in Pediatric Population. Age (Years), 2(184), 69-4.

[3] Meshram, R. M., & Gajimwar, V. S. (2018). Prevalence, profile, and pattern of congenital heart disease in Central India: A prospective, observational study. Nigerian Journal of Cardiology, 15(1), 45.

[4] Nazari, P., Davoodi, M., Faramarzi, M., Bahadoram, M., & Dorestan, N. (2016). Prevalence of congenital heart disease: a single center experience in southwestern of Iran. Glob J Health Sci, 8(10), 56421.

[5] Khoshhal, S. Q., Albasri, A. M., Morsy, M. M. F., & Alnajjar, A. A. (2020). The trends and patterns of congenital heart diseases at Madinah Cardiac Center, Madinah, Saudi Arabia. Saudi Medical Journal, 41(9), 977.

[6] Heart Valve Disease: Heart Valve Problems. UCSF Pediatric Cardiothoracic Surgery. https://pedctsurgery.ucsf.edu/conditions--procedures/heart-valve-disease.aspx [Accessed 31 August 2022].

[7] Reyna, M. A., Kiarashi, Y., Elola, A., Oliveira, J., Renna, F., Gu, A., Perez-Alday, E. A., Sadr, N., Sharma, A., Mattos, S., Coimbra, M. T., Sameni, R., Rad, A. B., Clifford, G. D. (2022). Heart murmur detection from phonocardiogram recordings: The George B. Moody PhysioNet Challenge 2022. medRxiv, doi: 10.1101/2022.08.11.22278688

[8] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220.

[9] Oliveira, J., Renna, F., Costa, P. D., Nogueira, M., Oliveira, C., Ferreira, C., & Coimbra, M. T. (2022). The CirCor DigiScope Dataset: From Murmur Detection to Murmur Classification. IEEE Journal of Biomedical and Health Informatics, doi: 10.1109/JBHI.2021.3137048.

[10] Andén, J. and Mallat, S. "Deep scattering spectrum." IEEE Transactions on Signal Processing 62.16 (2014): 4114-4128.

[11] Wavelet Time Scattering Classification of Phonocardiogram Data (MATLAB Help Center Tutorial). https://www.mathworks.com/help/wavelet/ug/wavelet-time-scattering-classification-of-phonocardiogram-data.html [Accessed 31 August 2022].

[12] Geron, A. *Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 2nd Edition. O'Reilly Media, Inc. pp. 153-174. (2019).

Address for correspondence:

Grace M. Mirsky
Benedictine University
5700 College Road
Lisle, IL 60532, USA
gmirsky@ben.edu