Detecting Murmurs in Phonocardiogram Recordings Using Feature Based Ensemble Learning

Mohammed H Baydoun[(1)], Hassan M Ghaziri[(1)], Ali El-Hajj[(2)] , Lise E Safatly[(2)]

[(1)] Beirut Research and Innovation Center, Beirut, Lebanon
[(2)] American University of Beirut, Beirut, Lebanon

This work addresses the Physionet 2022 challenge which targets detecting murmurs in Phonocardiogram recordings. The sound data was collected from different prominent auscultation locations using a digital stethoscope with the main locations being the pulmonary valve, the aortic valve, the mitral valve and the tricuspid valve. The data further includes some standard information about the participants such as age and gender. The labels denote the presence or absence of a murmur noting that some samples are considered unclear.

The initial proposed approach relies on our previous related work [1] which was tested on the Physionet 2016 challenge data [2]. The technique is feature based where the data is first preprocessed to reduce the noise, and then segmented using a basic segmentation technique that divides the data into 0.6 second intervals. This is followed by extracting several statistical, time and wavelet based features. The work adds to these several similarity based features to better detect murmurs in addition to applying feature selection. The features are then used for classification using ensemble based techniques. In particular, four distinct bagging classifiers are trained and used to classify the hidden data where each classifier concentrates on a different valve. The classifiers utilize oversampling for the data labeled with murmurs to better handle the utilized cost metric. Each recording is assigned to the classifier for the corresponding valve and if a participant has multiple recordings then the maximum score is selected.

The proposed technique was first checked locally by utilizing 2-fold cross-validation and afterwards on the initial validation set of the challenge. The best validation score result is currently 540.

Initial 2-Fold results

Utilized Score	Value
F-score	0.33
Proposed metric	**547**

Initial Validation Results

Utilized Score	Value
Proposed metric	**540**

[1] Baydoun, M., Safatly, L., Ghaziri, H. and El Hajj, A., 2020. Analysis of heart sound anomalies using ensemble learning. Biomedical Signal Processing and Control, 62, p.102019.

[2] Liu, C., Springer, D., Moody, B., Silva, I., Johnson, A., Samieinasab, M., Sameni, R., Mark, R. and Clifford, G., 2016. Classification of Heart Sound Recordings-The PhysioNet Computing in Cardiology Challenge 2016. *PhysioNet*.