Heart murmur detection from phonocardiogram recordings: The George B. Moody PhysioNet Challenge 2022

Matthew A. Reyna1*, Yashar Kiarashi1, Andoni Elola2, Jorge Oliveira3, Francesco Renna4, Annie Gu1, Erick A. Perez Alday1, Nadi Sadr1,5, Ashish Sharma1,†, Jacques Kpodonu6, Sandra Mattos7, Miguel T. Coimbra4, Reza Sameni1, Ali Bahrami Rad1, Gari D. Clifford1,8

1 Department of Biomedical Informatics, Emory University, Atlanta, GA, USA
2 Department of Electronic Technology, University of the Basque Country UPV/EHU, Eibar, Gipuzkoa, Spain
3 REMIT, Universidade Portucalense, Porto, Portugal
4 INESC TEC, Faculdade de Ciências, Universidade do Porto, Porto, Portugal
5 ResMed, Sydney, Australia
6 Division of Cardiac Surgery, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA
7 Unidade de Cardiologia e Medicina Fetal, Real Hospital Português, Recife, Pernambuco, Brazil
8 Department of Biomedical Engineering, Emory University and the Georgia Institute of Technology, Atlanta, GA, USA

* Current Address: Department of Biomedical Informatics, Emory University, Atlanta, Georgia, United States
† Deceased

S1 Appendix. Additional scoring metrics

We defined additional scoring metrics to allow us to make more direct comparisons between methods and tasks.

In particular, we defined a weighted accuracy metric for the clinical outcome identification task as

$$a_{\text{outcome}} = \frac{5n_{TP} + n_{TN}}{5(n_{TP} + n_{FN}) + (n_{FP} + n_{TN})},$$

(1)

where Table 2 defines a two-by-two confusion matrix $N = [n_{ij}]$ for the clinical outcome abnormal and normal classes.

We also defined the total cost of diagnosis and treatment with algorithmic pre-screening of murmurs as

$$c_{\text{outcome}}^{\text{total}} = f_{\text{algorithm}}(n_{\text{patients}}) + f_{\text{expert}}(o_{PA} + o_{PN} + o_{UA} + o_{UN}, n_{\text{patients}}) + f_{\text{treatment}}(o_{PA} + o_{UA}) + f_{\text{error}}(o_{AA}),$$

(2)

where Table $A1$ defines a three-by-two confusion matrix $O = [o_{ij}]$ for the clinical outcome abnormal and normal classes, $n_{\text{patients}}$ is the total number of patients, and $f_{\text{algorithm}}$, $f_{\text{expert}}$, $f_{\text{treatment}}$, $f_{\text{error}}$ are defined above.
Table A1. Confusion matrix for murmur detection task with three classes (murmur present, murmur unknown, and murmur absent) using clinical outcomes with two classes (clinical outcome abnormal and clinical outcome normal). The columns are the ground truth labels from the human annotator, and the rows are the classifier outputs. The entries of the confusion matrix provide the number of patients with each classifier output for each ground truth class.

<table>
<thead>
<tr>
<th>Murmur Classifier</th>
<th>Clinical Outcome Expert</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Abnormal</td>
</tr>
<tr>
<td>Present</td>
<td>$o_{PA}$</td>
</tr>
<tr>
<td>Unknown</td>
<td>$o_{UA}$</td>
</tr>
<tr>
<td>Absent</td>
<td>$o_{AA}$</td>
</tr>
</tbody>
</table>