

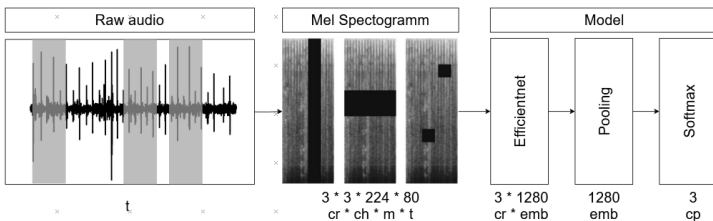
# Using Mel-Spectrograms and 2D-CNNs to detect Murmurs in Variable Length Phonocardiograms

Marius S. Knorr\*, Jan P. Bremer, Renate B. Schnabel

University Medical Center Hamburg Eppendorf, Department of Cardiology, University Heart & Vascular Center Hamburg, Hamburg, Germany

The Physionet Challenge 2022 requires participants to detect the occurrence of murmurs in phonocardiograms. Here, we deployed a carefully validated deep learning approach to variable length audio classification.

Raw audio data at a 4,000 Hz sampling rate was converted to time-frequency images with 224 mel-frequency bands. Next, the images were subdivided into three non-overlapping crops along the time axis. The crops were augmented utilizing dropout augmentations as well as a mixup augmentation shuffling the order. During model training the crops were fed into an EfficientNet-B0 neural network returning feature embeddings for each crop. An average pooling layer unifies the feature embeddings of the three crops, followed by three output neurons (present, unknown, absent) with a softmax activation function for classification. The model was trained utilizing categorical cross entropy for 20 epochs and was employed in a stratified k-fold validation scheme with 6 folds (4 training, 1 validation, 1 threshold selection) and a hold-out test dataset (20%; 189 patients). The performance was monitored per epoch on the validation data. The best epoch was chosen based on the validation loss and subsequently used. The final probabilities were derived by averaging the probabilities of available phonocardiograms per patient. The threshold for each class is optimized by iterating over multiple thresholds on the threshold selection fold for the final predictions. Metrics were calculated on the hold-out test dataset for each fold. On the k-fold validation scheme we scored a competition metric of 509 (CI: 486 - 532), multi-class area under the receiver operating curve of 0.773 (CI: 0.755-0.791) and area under precision recall curve of 0.57 (CI: 0.55 - 0.59). A single fold of our approach yielded a competition score of 498 on the leaderboard with the team name uke-cardio.



*t = time, cr = crops, ch = channels, m = mel-bands, emb = feature embedding, cp = class probabilities*