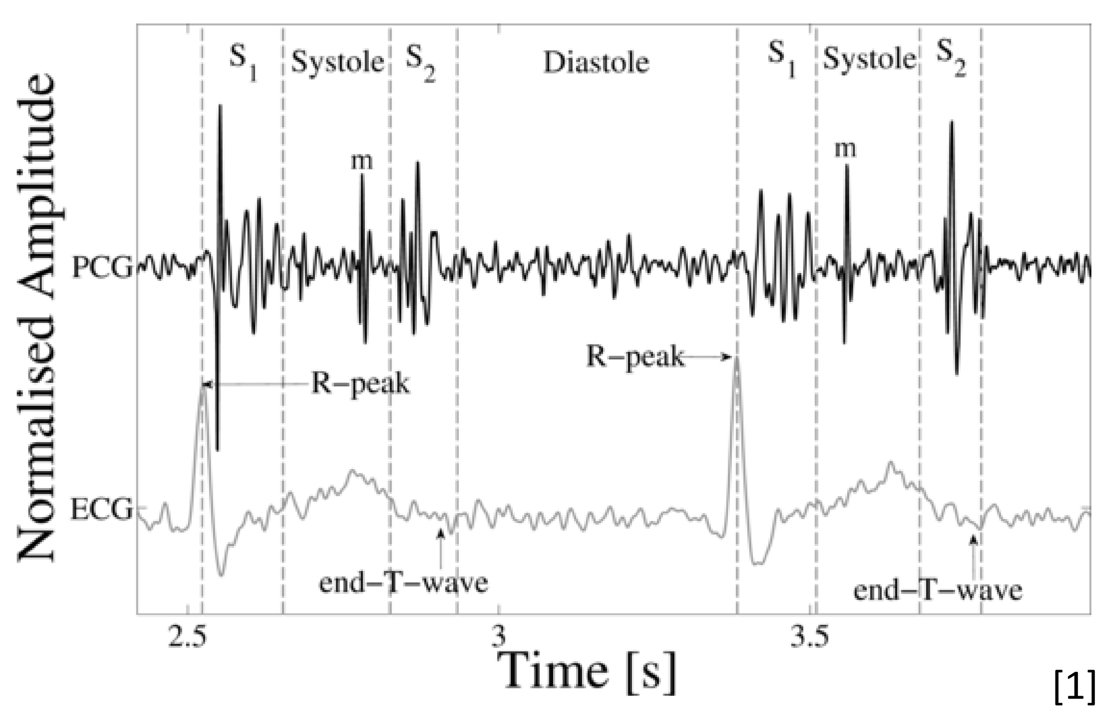
Application of Machine Learning Techniques for Heart Sound Recording Classification

Vincent Lee & Anatoly Yakovlev

vclee@stanford.edu, yakovlev@stanford.edu

Abstract

Compared to electrocardiogram (ECG), phonocardiograms (PCG) from heart sound recordings is ubiquitous and inexpensive. The goal is to create a heart abnormality classifier which takes heart audio recordings as input and provides a binary normal / abnormal indicator as output.



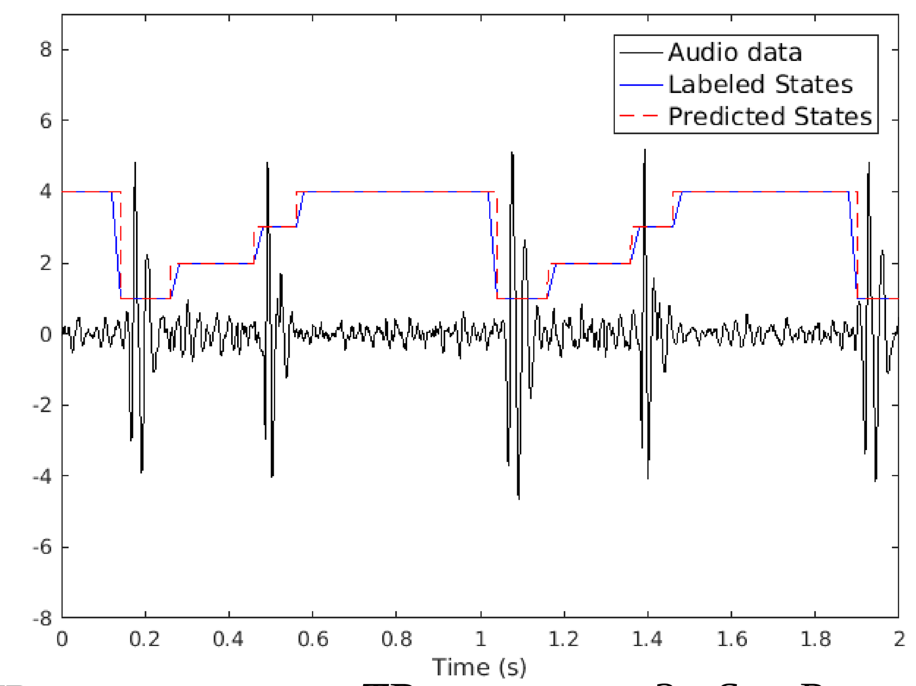
[1]

Dataset

Dataset consists of 3,541 heart sound recordings in .wav format, each lasting from 5 to over 120 seconds and recorded under various conditions. Sound recordings are provided by Physionet [2] with 2,725 normal and 816 abnormal sounds. After segmentation, dataset was balanced to have equal number of normal and abnormal recordings, shuffled, and split into 80% training and 20% test sets. Cross-validation set is 10% of training examples.

Data Segmentation Results

PCG segments of S1, systole, S2, and diastole are respectively encoded as 1, 2, 3, and 4. For evaluation, we adopt a tolerance parameter of 60 ms, which is used to define a window around the reference state label for which the predicted state onset is considered correct if it were to fall within that window.



TP: true positive
FN: false negative
FP: false positive

$$Se = \frac{TP}{TP+FN} \quad P_{+} = \frac{TP}{TP+FP} \quad Acc = \frac{TP}{TP+FP+FN} \quad F_1 = \frac{2\times Se\times P_{+}}{Se+P_{+}}$$

Training Set Performance of LR-HSMM (720 examples)

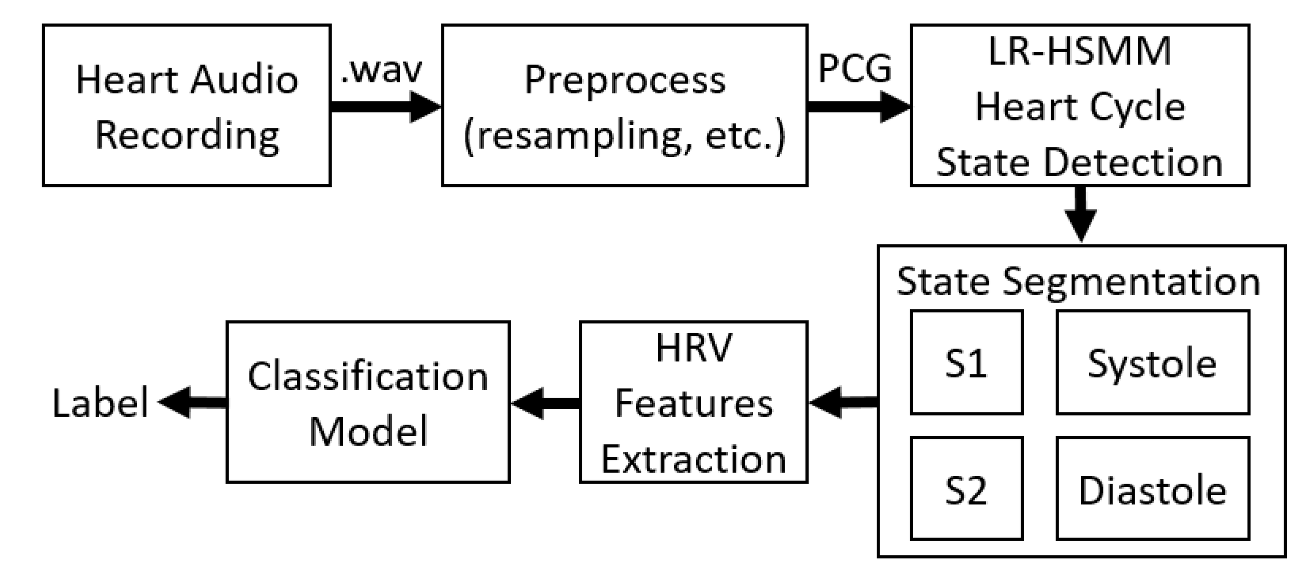
State	TP	FN	FP	Se	P+	Acc	F1
S1	9376	809	2	92.1	99.9	92.0	95.9
Systole	9685	478	64	95.3	99.3	94.7	97.3
S2	9642	389	40	96.1	99.6	95.7	97.8
Diastole	9368	641	13	93.6	99.9	93.5	96.6

Cross Validation Set Performance of LR-HSMM (72 examples)

State	TP	FN	FP	Se	P+	Acc	F1
S1	794	67	0	92.2	100	92.2	96.0
Systole	809	49	4	94.3	99.5	93.9	96.8
S2	803	40	8	95.3	99.0	94.4	97.1
Diastole	784	58	0	93.1	100	93.1	96.4

Overall Classification Flow

The overall classification flow consists of audio preprocessing into PCGs, data segmentation into the four heart states, HRV feature extraction, and finally heart abnormality classification based on the HRV features.



Classification Model

Model Name	Model Equation
Logistic Regression	$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$
K-Means	Color: $\forall i: c^{(i)} := arg\,\min_j \lVert x^{(i)} - \mu_j \rVert$ Adjust centroids: $\forall j: \mu_j := \frac{\sum_{i=1}^m 1\{c^{(i)}=j\}x^{(i)}}{\sum_{i=1}^m 1\{c^{(i)}=j\}}$ j ={1, 2} are {normal, abnormal} classes
Neural Net. 1	3-layers, h1=10 units, σ-activation
Neural Net. 2	4-layers, h1=10 units, h2=5 units, σ-activation
SVM with Gaussian Kernel	$h(x) = sign\left(\sum_{i=1}^m \alpha_i y^{(i)} K(x^{(i)}, x) + b\right)$ $K(z,x) = \exp\left(-\frac{\lVert z-x \rVert^2}{2\sigma^2}\right)$

Discussion

- Unbalanced distribution in training dataset leads to biased model predictions
- Extracted HRV features are the classification performance bottleneck:
  - HRV features derived from ground-truth segmented PCGs based on ECG waveforms yielded < 1% improvement in classification accuracy
  - Models using unmodified HRV features showed similar classification accuracy
- Gaussian Kernel SVM and NN models achieve better performance due to their ability to transform/synthesize new features

Future Work

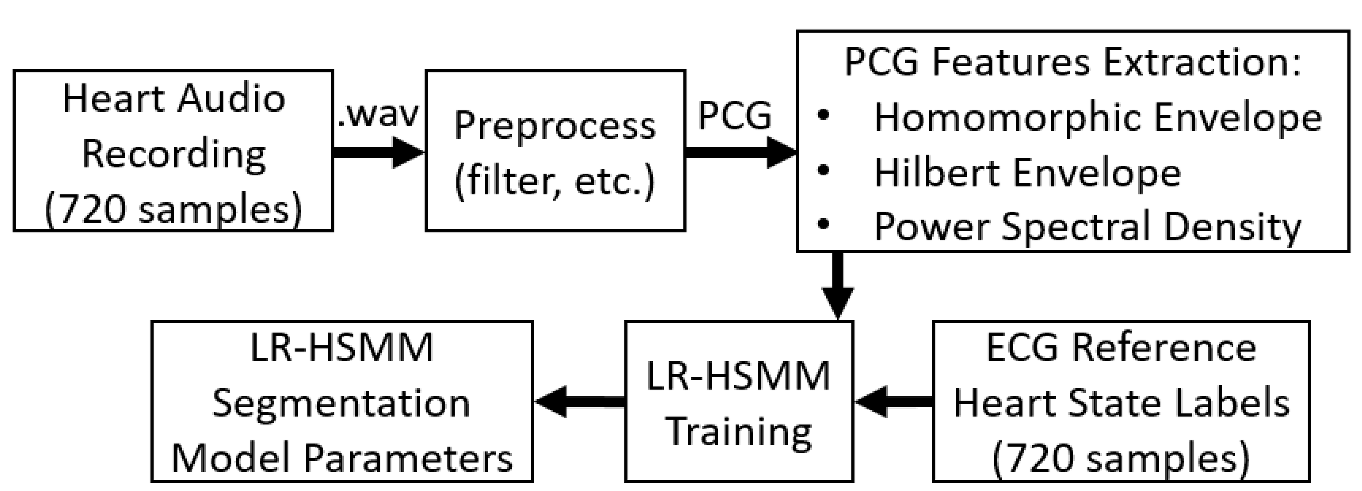
- Identifying HRV features that would improve classification accuracy on all models
- Incorporate ReLU activations for NNs
- Incorporate Kaggle Heartbeat Sounds dataset to add abnormal examples
- Train RNNs on sound waveforms directly

References

[1] D.B. Springer, et al., "Logistic Regression-HSMM-Based Heart Sound Segmentation," Trans. BE, April 2016.
[2] G.D. Clifford et al., "Classification of normal/abnormal heart sound recordings: PhysioNet/Computing in Cardiology Challenge 2016," CinC.
[3] C. Liu, et al., "Performance of an open-source heart sound segmentation algorithm on eight independent databases," Physio. Meas., 2017.
[4] A. Kampouraki, et al., "Heartbeat Time Series Classification With Support Vector Machines," TITB, July 2009.

Data Segmentation Methodology

Prior to heart sound classification, heart audio data is first segmented into S1, systole, S2, and diastole heart states using a logistic regression hidden semi-Markov model [1, 3].



HRV Feature Extraction

1) Heart rate variability (HRV) features are extracted from the segmented PCG [4]; 2) PCA is used to derive top 15 principal components to reduce input feature dimension; 3) Data is normalized for K-means and NNs; 4) SVM uses Gaussian kernel. Model input data has rows as examples and columns as features.

RR interval	Interval ratio of systole to RR	% of RR intvls > x
S1 interval	Interval ratio of diastole to RR	Entropy of RR intvls
S2 interval	Intvl ratio of systole to diastole	RMS of successive RR intvls
Systole interval	Ampl. ratio of systole to S1	% autocorr. of S1 intvls > x
Diastole interval	Ampl. ratio of diastole to S2	

Classification Results

The dataset is split into 1,306 training and 326 test examples. Each set consists of approximately 50% abnormal and 50% normal recordings. 10% of training examples are used for cross-validation. Model is retrained on entire training dataset prior to testing.

Model	Sensitivity		Precision		Accuracy		F1-score	
	Train	Test	Train	Test	Train	Test	Train	Test
LR	53.3	56.4	61.7	65.7	60.1	63.5	57.2	60.7
K-Means	43.6	94.5	70.4	52.0	62.6	53.7	53.9	67.1
Neural Net. 1	80.1	77.3	77.7	71.2	78.6	73.0	78.9	74.1
Neural Net. 2	80.6	82.8	70.8	68.5	73.6	72.4	75.4	75.0
Gaussian-K SVM		88.8		88.3		88.5		88.5