Learning Image-based Representations for Heart Sound Classification

Zhao Ren[1], Nicholas Cummins[1], Vedhas Pandit[1], Jing Han[1], Kun Qian[1,2], Björn Schuller[1,3]

[1]ZD.B Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany
[2] Machine Intelligence and Signal Processing Group, Technische Universität München, Germany
[3]GLAM – Group on Language, Audio & Music, Imperial College London, UK

26.04.2018 Lyon, France
2018 International Digital Health Conference

Authors

Zhao Ren

Dr. Nicholas Cummins

Vedhas Pandit

Jing Han

Kun Qian

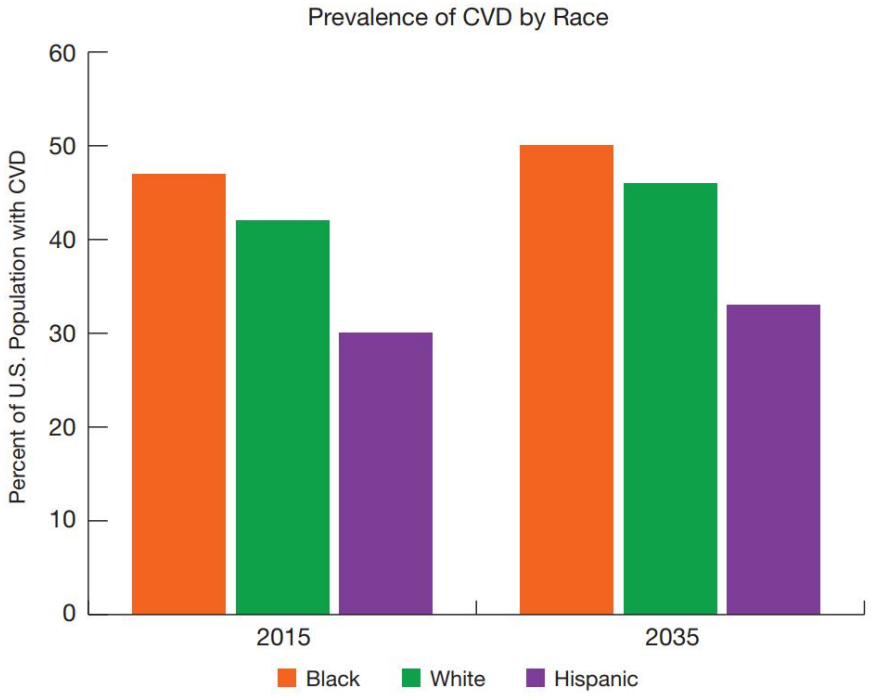
Prof. Björn Schuller

Outline

- Motivation
- Contribution
- Methodology
- Experiment
- Conclusions and Future Work

Motivation

American heart association cardiovascular disease (CVD) Burden Report 2017



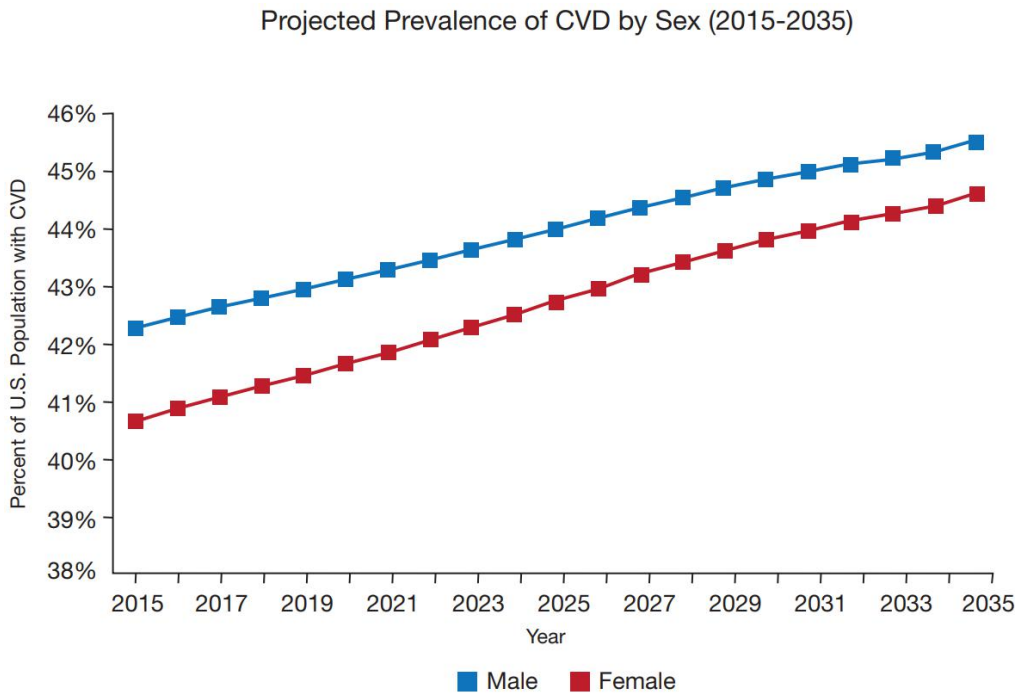


image source

https://healthmetrics.heart.org/wp-content/uploads/2017/10/Cardiovascular-Disease-A-Costly-Burden.pdf

Motivation

- Heart disease is a leading worldwide health burden
- Manual auscultation
  - The auscultating accuracy of primary care physicians is not high enough
  - The access to clinicians and medical care is limited in some areas

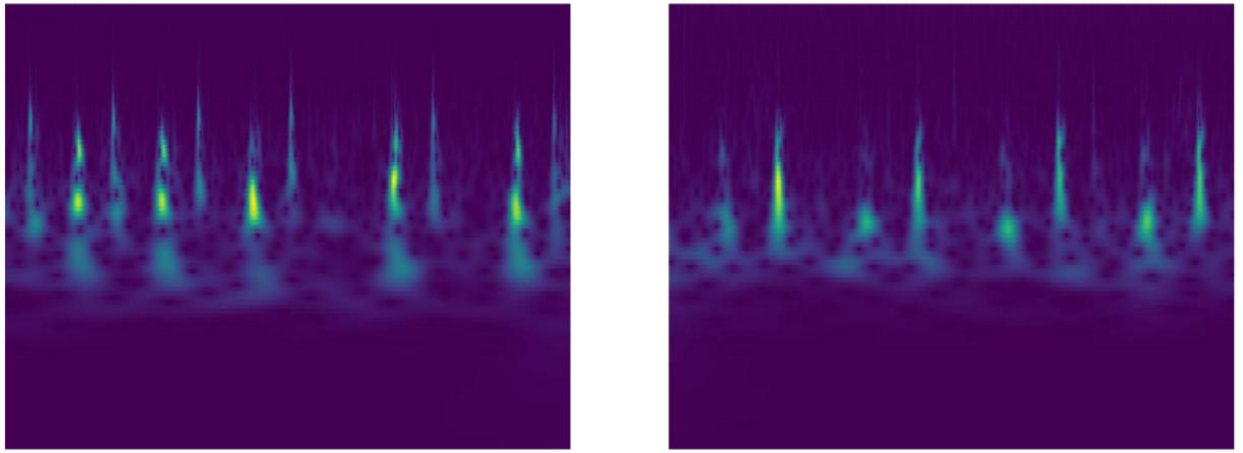
- Automated classification of heart sound by machine learning and deep learning
  - High accuracy
  - No restrain to the area

Contribution

- Scalogram images of Phonocardiogram (PCG) signals
- Pre-trained Convolutional Neural Networks (CNNs)
- transfer-learning based adaptation and updating of the Image Classification CNN (ImageNet) parameters

Methodology

Scalogram Representation



(a) Normal (*a0007.wav*) (b) Abnormal (*a0001.wav*)

Figure 1: The scalogram images are extracted from the first 4 s segments of normal/ abnormal heart sounds using the viridis colour map.

Methodology

VGG16

(1000-label classification)

convolutional layer:

conv<receptive field size>-<number of channels>

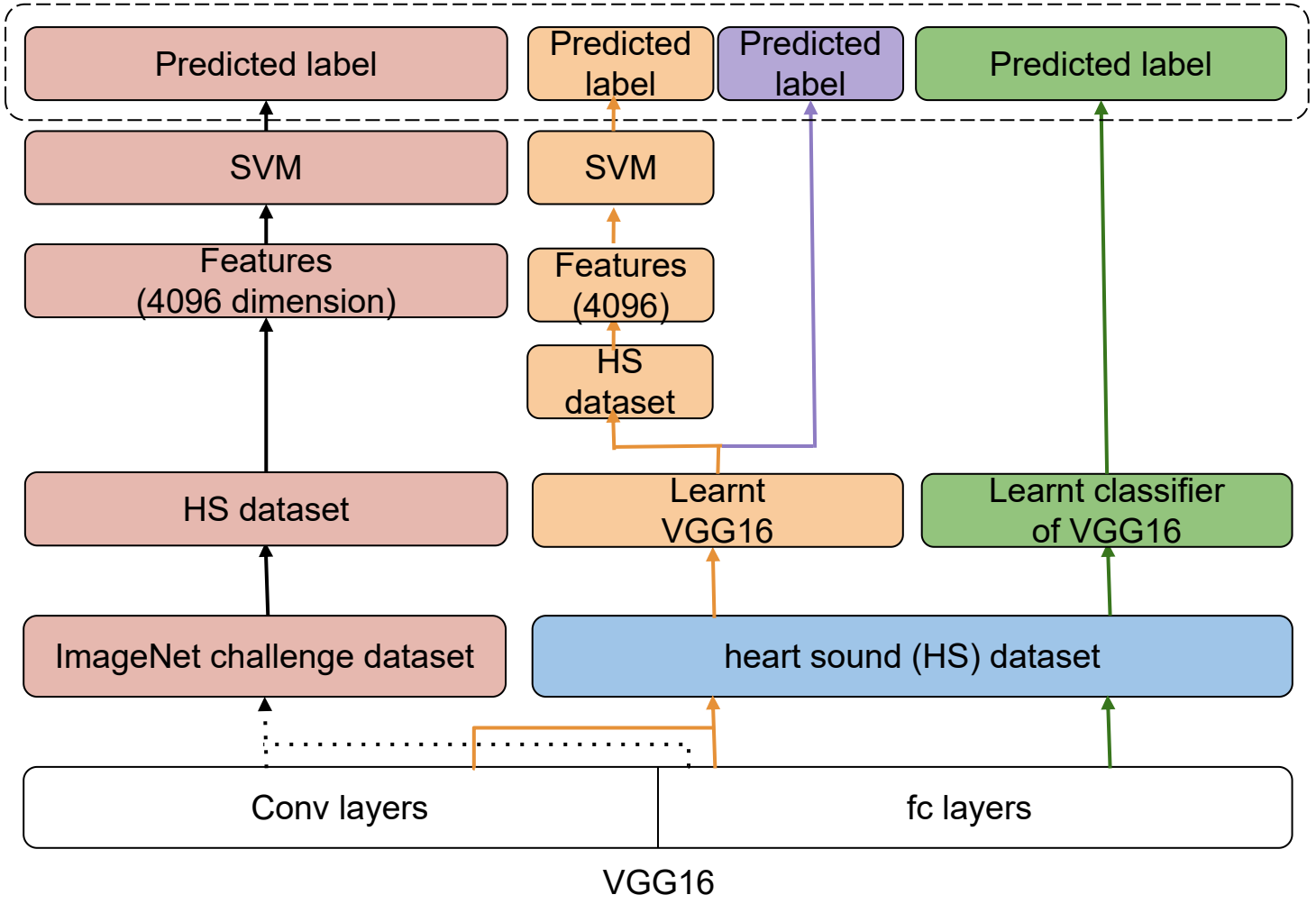
input: RGB image
2 × conv3-64; maxpooling
2 × conv3-128; maxpooling
2 × conv3-256; maxpooling
2 × conv3-512; maxpooling
2 × conv3-512; maxpooling
fully connected layer fc6-4096
fully connected layer fc7-4096
fully connected layer fc-1000
output: soft-max

Methodology

- Deep PCG Feature Representations
  - Pre-trained ImageNet + SVM
  - Learnt ImageNet + SVM

  Note: the deep features are extracted from the first fully connected layer fc6 of VGG16

- End-to-end ImageNet based Classification
  - Learning Classifier of ImageNet
  - Learning ImageNet



Database

- PhysioNet/ CinC dataset

Note: DLUT is always used in system training; a three-fold cross validation by excluding the databases MIT, AUTH, or UHA for fold 1, fold 2, or fold 3.

Dataset	Database	Recordings	Normal	Abnormal	Durations (s)	
					Min	Max
Training	MIT	409	117	292	9.27	36.50
a three-fold cross validation	AUTH	31	7	24	9.65	122.00
	UHA	55	27	28	6.61	48.54
training	DLUT	2141	1958	183	8.06	101.67
Total		**2636**	2109	527		
Test	AAD	490	386	104	5.31	8.00
	SUA	114	80	34	29.38	59.62
Total		**604**	466	138		

Experimental Setup

Baseline Classification System

Data

- Non-overlapping chunks of 4 seconds
- Late-fusion Strategy: the highest posterior probability $\{pi\}, i = 1,...,n$

Parameters

- VGG training
  - *learning rate = 0.001, batch size = 64, epoch = 50*
  - loss function: *entropy;* optimiser: *stochastic gradient descent*
- SVM
  - kernel: *linear;* $C \in [10^{-5};10^{+1}]$

Evaluation

Sensitivity (Se)	$Se = TP / (TP + FN)$	*TP:* the number of true positive abnormal samples
Specificity (Sp)	$Sp = TN / (TN + FP)$	*FN:* the number of false negative abnormal samples
Mean Accuracy (MAcc)	$MAcc = (Se + Sp) / 2$	*TN:* the number of true negative normal samples
		FP: the number of false positive normal samples

Experimental Results

Performances comparison of the proposed approaches with baseline. The methods are evaluated on the 3-fold development set and the test set. The experimental results are evaluated by Sensitivity (Se), Specificity (Sp), and the Mean Accuracy (MAcc).

	Development Set												Test set		
	fold 1			fold 2			fold 3			mean					
performance [%]	*Se*	*Sp*	*MAcc*	*Se*	*Sp*	*MAcc*	*Se*	*Sp*	*MAcc*	*Se*	*Sp*	*MAcc*	*Se*	*Sp*	*MAcc*
ComParE+SVM (baseline)	23.6	93.2	58.4	58.3	100.0	79.2	00.0	100.0	50.0	27.3	97.7	**62.5**	76.8	17.0	46.9
pre-trained VGG+SVM	57.2	70.9	64.1	41.7	85.7	63.7	17.9	81.5	49.7	38.9	79.4	59.1	24.6	87.1	55.9
learnt VGG+SVM	58.6	57.3	57.9	83.3	57.1	70.2	32.1	70.4	51.3	58.0	61.6	59.8	24.6	87.8	**56.2**
learning Classifier of VGG	68.2	51.3	59.7	79.2	14.3	46.7	35.7	40.7	38.2	61.0	35.4	48.2	33.3	63.7	48.5
learning VGG	83.6	40.2	61.9	95.8	28.6	62.2	53.6	44.4	49.0	77.7	37.7	57.7	12.3	95.7	54.0

p<.001 by one-tailed z test

Conclusions and Future Work

Conclusions:

- Scalogram images of Phonocardiogram (PCG)
- Pre-trained Image Classification Convolutional Neural Networks (ImageNet)
- VGG features + Support Vector Machine (SVM)

Future work:

- Data augmentation
- A new Imagenet topology based on the scalogram images will be developed and validated on a variety of heart sound datasets, e. g., AudioSet

Thank you for listening!

Q&A

zhao.ren@informatik.uni-augsburg.de

Chair of Embedded Intelligence for Health Care and Wellbeing