Searching for Effective Neural Network Architectures
for Heart Murmur Detection from Phonocardiogram

Hao WEN[1], Jingsu KANG[2]

[1] *LIMB and School of Mathematical Sciences, Beihang University*

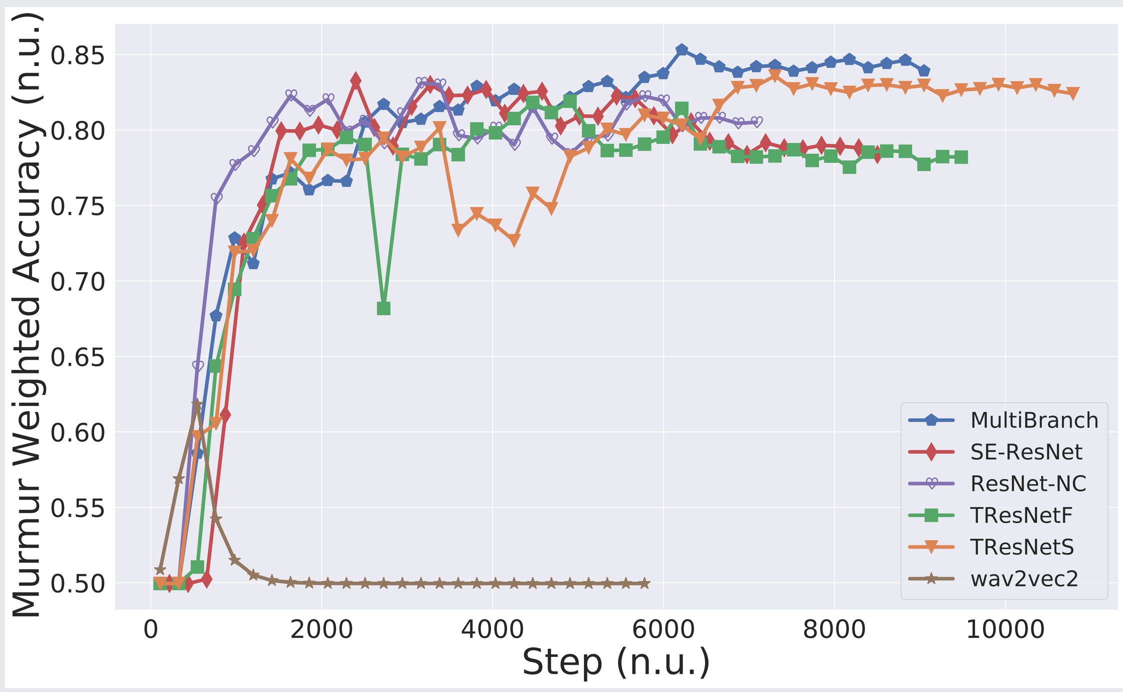
[2] *Tianjin Medical University*

The George B. Moody PhysioNet Challenge 2022, Team Revenger

Introduction

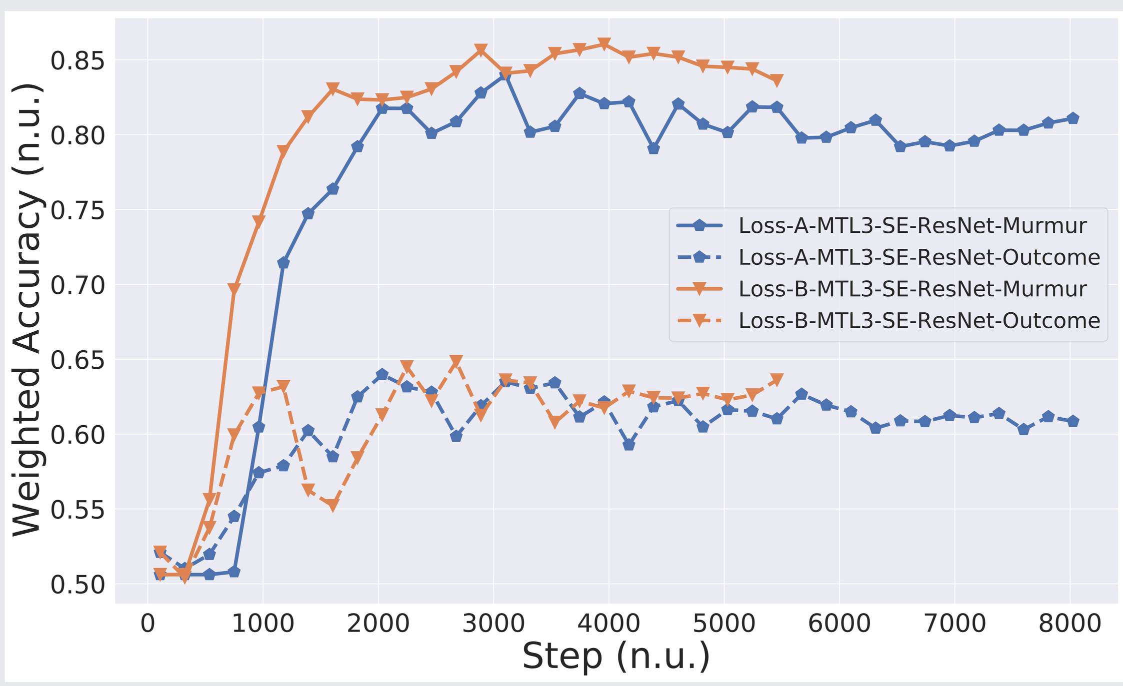
- We used **neural network models** to make per-recording predictions, then used a simple **greedy rule** to obtain per-patient predictions.
- We applied the **multi-task learning (MTL)** paradigm via hard parameter sharing to make predictions using **ONE** model. The model has 2 classification heads (murmur and outcome, denoted "MTL2") or 3 (denoted "MTL3") with an additional segmentation head.
- We performed extensive **architecture searching for the backbone (the shared parameters)** of the neural network model.

Neural Network Backbones



A part of the backbones (the shared representation) we experimented with.

Loss Function Choices

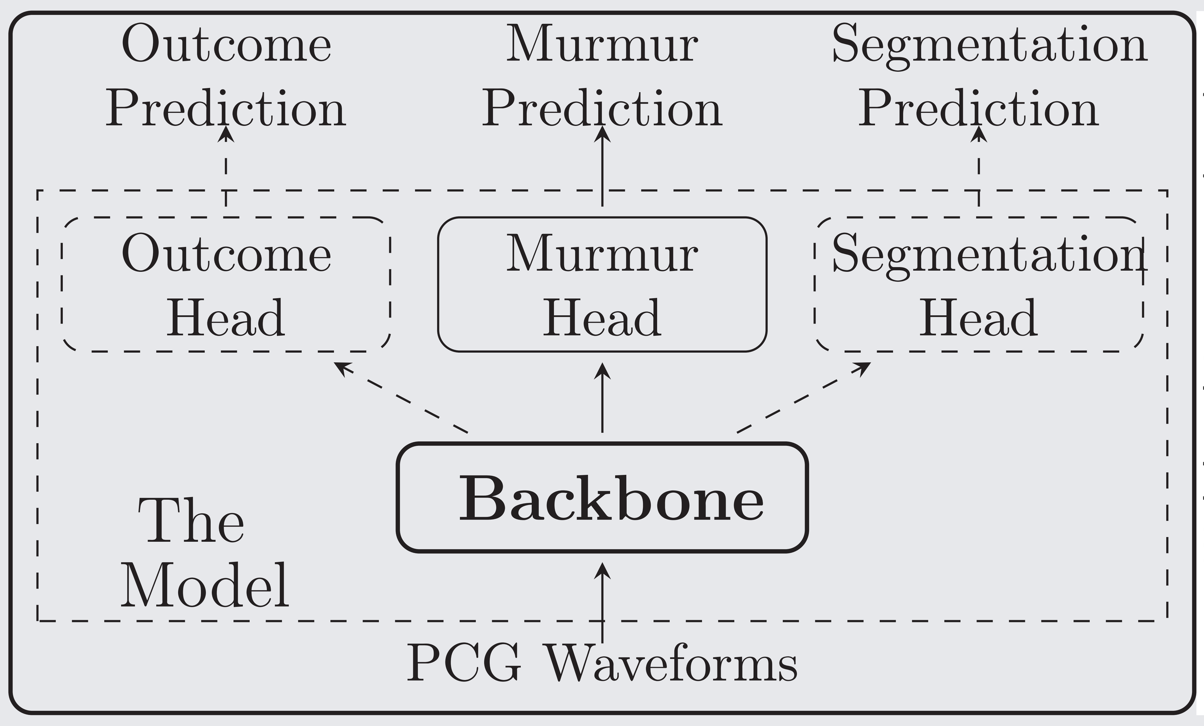


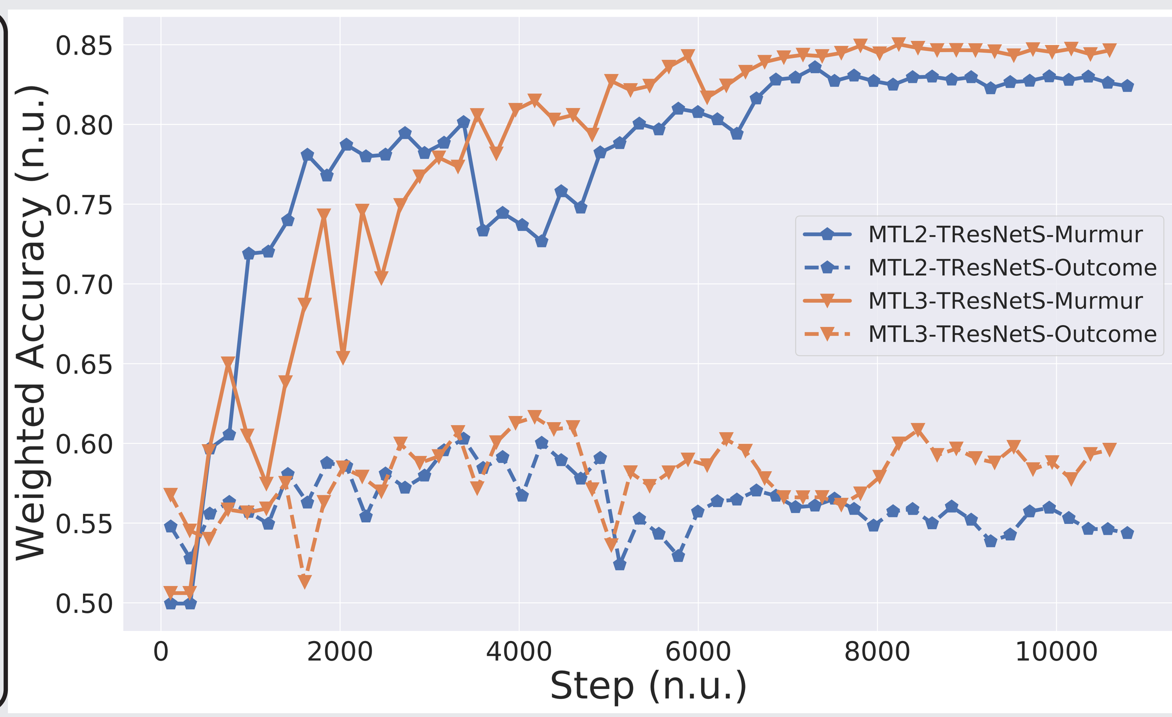
- Loss-A: Asymmetric Loss
- Loss-B: Weighted BCE Loss

Demographic Features

- Some demographic features are strongly correlated with "Outcome".
- A random forest classifier using demographic features and the murmur predictions as input improved outcome sores (reduced the outcome cost).
- Our final submission **DID NOT** use demographic features and related auxiliary (e.g. random forest) classifiers.

Multi-Task Learning





The multi-task learning (MTL) paradigm. Experiments indicates its efficacy.

Preprocess Pipeline and Augmentations

- Resampling to 1000 Hz.
- Butterworth bandpass filtering of order 3 and cutoff frequencies 25 - 400 Hz.
- Z-score normalization to zero mean and unit variance.

- Adding coloured noises.
- Polarity inversion (flipping).

Training Setups

- Optimizer: AMSGrad variant of **AdamW** with **OneCycleLR** scheduler.
- Stratified train-validation split: 80% – 20%.
- Batch size 32; epoch number ≤ 60 with early stopping.
- Freeze backbone from specific epoch (typically 30).

Submission Results

	Murmur	Outcome	
	weighted accuracy	**cost**	weighted accuracy
Train	0.88 ± 0.05	8452 ± 2324	0.88 ± 0.07
Train-val	0.86 ± 0.01	11341 ± 336	0.79 ± 0.05
Hidden val	**0.689**	**9471.652**	NA
Ranking	**79/303**	**21/303**	NA

Scores on the train, and train-val sets are provided with mean and standard deviation over most of our offline experiments. Scores on the hidden validation set are provided with the best scores out of the 10 submissions.

Discussions and Limitations

- Our multi-task learning (MTL) paradigm proved practical for the problems of heart murmur detection and clinical outcome identification from PCGs. The additional segmentation head makes the shared representation (the common backbone) learn more general features and thus improves the performances for the original two classification tasks the Challenge raised.
- All our models used **time-domain** signals, i.e. the **waveforms** as inputs. The derived **time-frequency-domain** signals, for example, the **spectrograms**, were not tested. Models that accept mixed-type inputs were not tested either.
- The need for Z-score normalization has to be reconsidered. More frequency-domain augmentation methods should be applied.
- The convolutional neural backbones proved effective, but still have room for improvements compared to top teams.
- The potential of models with the **transformer architecture** (e.g. the **wav2vec2** model) were not fully explored. As well as the powerfulness of model pretraining via self-supervised learning on larger datasets (e.g. the PhysioNet EPHNOGRAM dataset, etc.).

Acknowledgements

We would like to thank professor **Deren Han** from the School of Mathematical Sciences, Beihang University and professor **Wenjian Yu** from the Department of Computer Science and Technology, BNRist, Tsinghua University for generously providing GPU servers to help accomplish this work.

Code, configs, etc. available at https://github.com/DeepPSP/cinc2022.