Contrastive Learning Leveraging Patient Metadata for Sample-efficient Lung and Heart Sound Representations

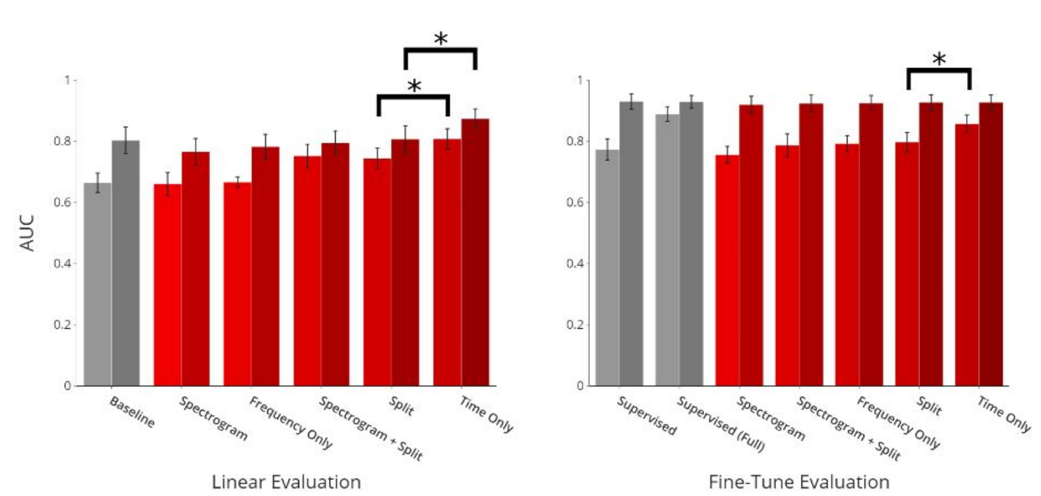
Pratham Soni[1], Siyu Shi[1], Pranav Sriram[1], Andrew Y. Ng[1], Pranav Rajpurkar[2]

1. Department of Computer Science, Stanford University; 2. Department of Biomedical Informatics, Harvard Medical School

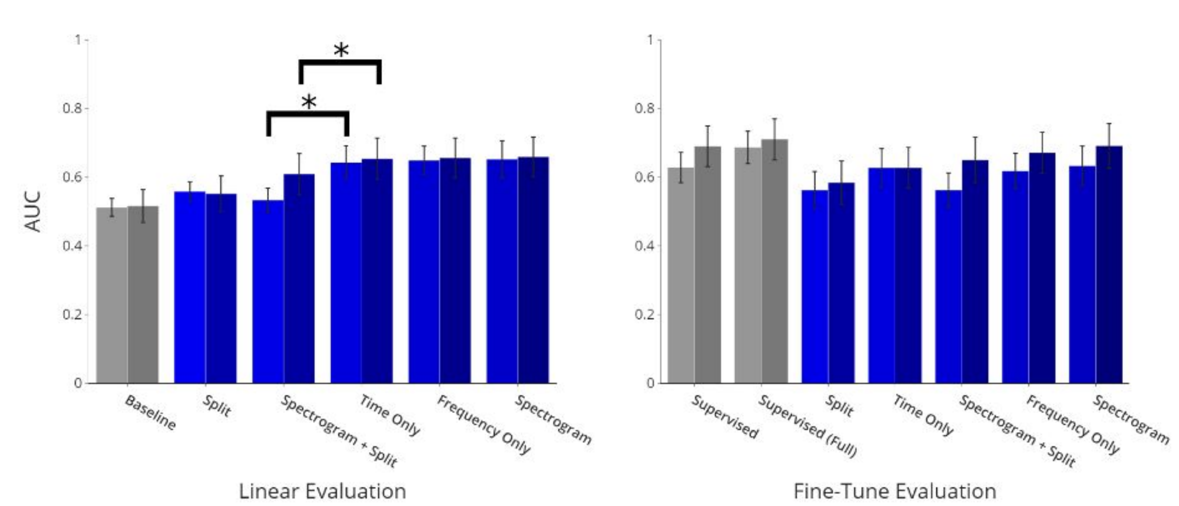
Motivation

1. Data labeling is often the limiting step in machine learning because it requires time from trained experts.
2. To address the limitation on labeled data, contrastive learning leverages unlabeled data to learn representations of data.
3. We propose a contrastive learning framework that utilizes metadata for selecting positive and negative pairs when training on unlabeled data.

Can contrastive learning better classify heart and lung sounds with limited data?

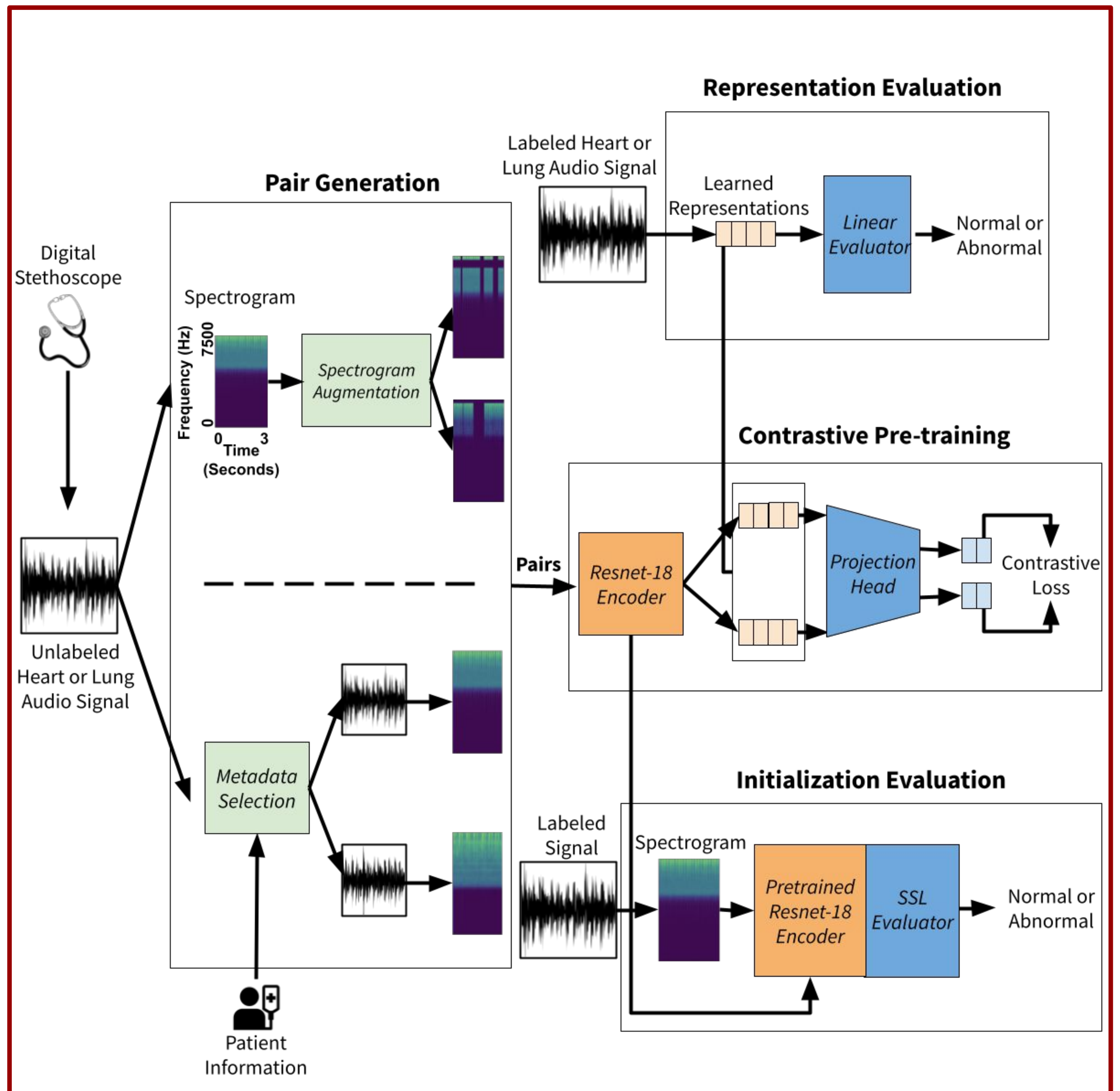


- Results for heart sounds classification
- All contrastive schemes match or surpass baseline performance at 10% with performance saturation for 100% in fine-tune evaluation

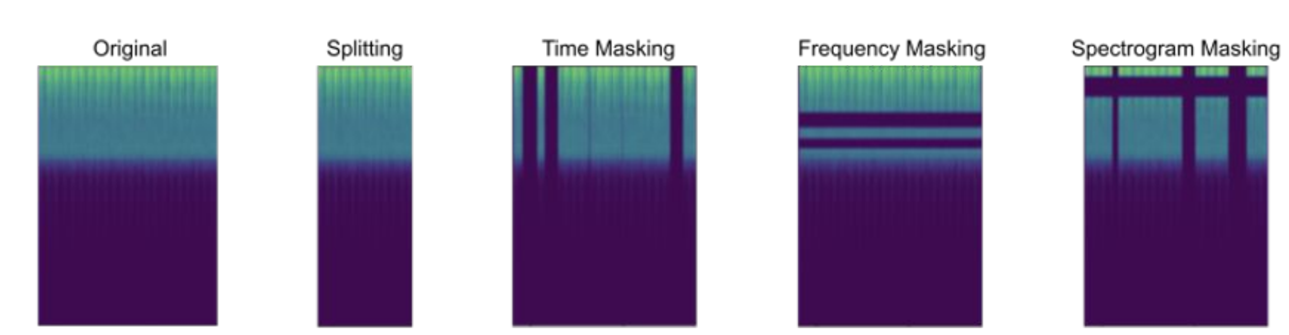


- Results for lung sounds classification
- For fine-tune evaluation, spectrogram masking provides the best performance, comparable to supervised training at both 10% and 100%.

Overview of the experimental setup



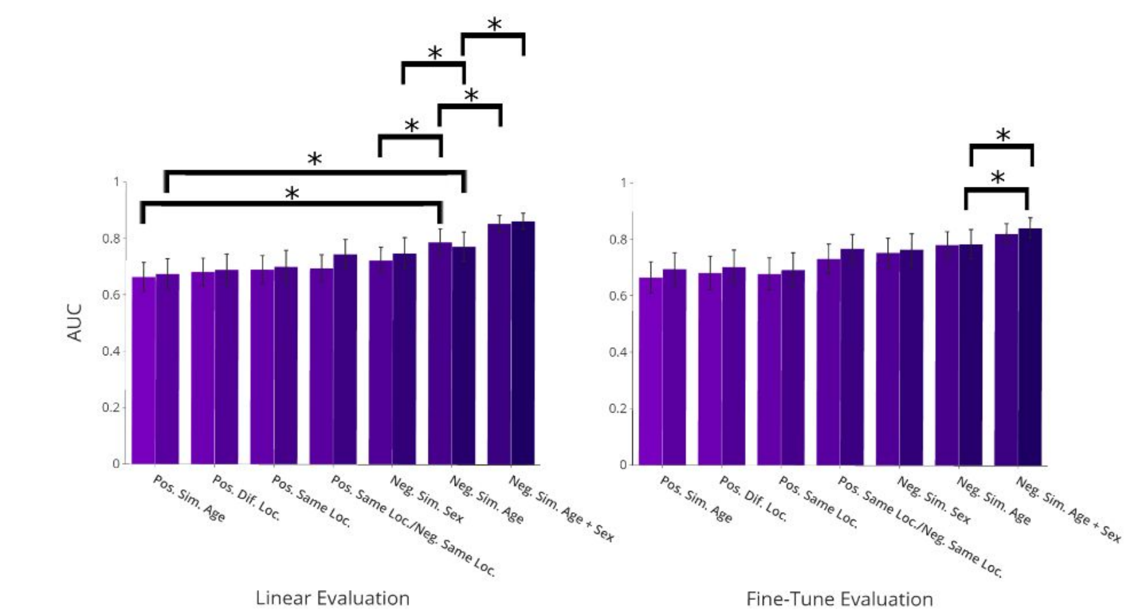
- We apply and extend self-supervised methodology in simCLR with views generated by applying randomized augmentations or selecting against recordings from the same sample
- We encapsulate the complexity of the unlabelled data for use as representations as well as using pre-training the encoder as an initialization point for further fine-tuning



Examples of spectrogram augmentation strategies. Note that in masking routines, masking is only applied to the only central audio sample, not left or right padding.

Can metadata further improved contrastive learning results?

Experimental results for lung sound classification with incorporated sample metadata.



- Performance at 10% and 100% training data levels are presented for each learning scheme. In both linear and fine-tune evaluation set-ups, there are marked differences between contrastive schemes, with the negative pair selection of age and sex providing superior performance in both set-ups.

Discussion

- We demonstrate that augmentation methods perform differently in different contexts and it is important to optimize contrastive learning frameworks according to the type of data.
- Results in lung sounds show that negative pair selection based on age improve downstream lung sound diagnosis tasks the most, followed by sex.

References

- Vu YNT, Wang R, Balachandar N, Liu C, Ng AY,, Rajpurkar P. MedAug: Contrastive learning leveraging patient metadata improves representations for chest X-ray interpretation.
- Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations.