Semi-Supervised Pulmonary Auscultation Analysis with Cross Pseudo Supervision

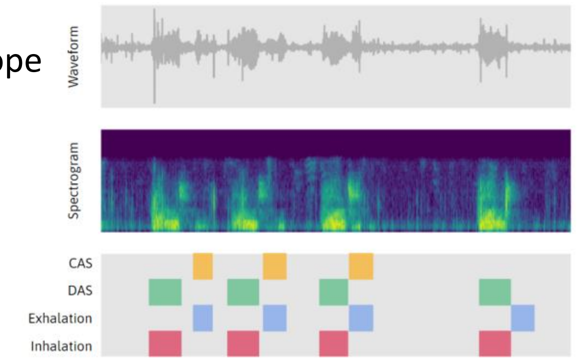
MSEF 2024

Jieruei Chang, *Princeton High School*

Motivation

Pulmonary Auscultation
- Listening to lung sounds with a stethoscope
- Crucial first step in respiratory disease diagnosis

The case for automation
- Reduce diagnostic errors
- Long-term patient monitoring

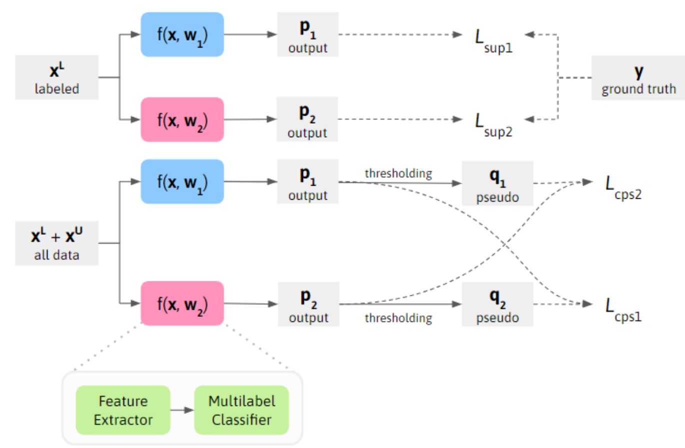


Problem Definition
- Labeling of respiratory sounds:
  - Inhalation / Exhalation
  - Continuous Adventitious Sounds (CAS): Wheeze, Rhonchi, Stridor
  - Discontinuous Adventitious Sounds (DAS): Crackles
- Multilabel segmentation problem: temporal position is important

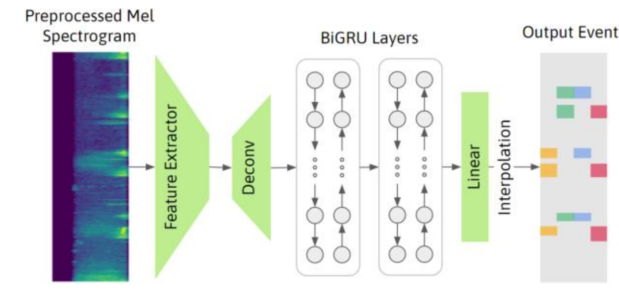
Challenges
- Limited dataset size; high intra-variation in lung sounds (difficulty in differentiation between different classes)

Approach

Cross Pseudo Supervision (CPS)
- Semi-supervised learning: Leverage both labeled and unlabeled data to address low-data situation
- Twin network setup: network outputs used for pseudo labels to supervise the other network



Single Network Architecture
- Feature extractor pretrained on AudioSet
- Bidirectional GRU to capture long-range temporal dependencies, understand more nuanced variations in lung sounds



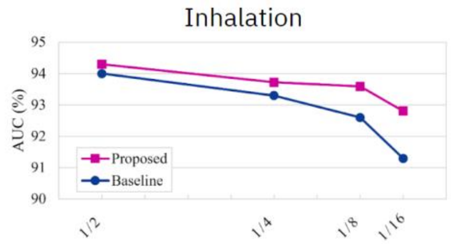
Loss Functions
- Binary Cross Entropy Loss between pseudo label and model predictions
- Weighted due to class imbalance
- Thresholding function as softmax/one-hot does not work due to multilabel classification problem

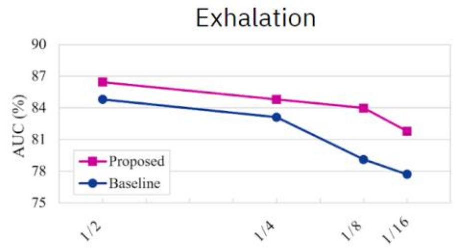
$$\mathcal{L}_{cps} = \frac{1}{N}\sum_{n=1}^{N} \ell_{bce}(f(\mathbf{x}_n, w_1), T_\tau(f(\mathbf{x}_n, w_2))) + \ell_{bce}(f(\mathbf{x}_n, w_2), T_\tau(f(\mathbf{x}_n, w_1))),$$

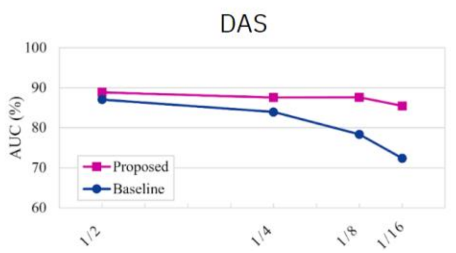
$$\ell_{bce}(\mathbf{p}, \mathbf{q}) = -\frac{1}{CD}\sum_{c=1}^{C} W_c \sum_{d=1}^{D} \mathbf{q}_{c,d} \cdot \log(\mathbf{p}_{c,d}) + (1 - \mathbf{q}_{c,d}) \cdot \log(1 - \mathbf{p}_{c,d}),$$

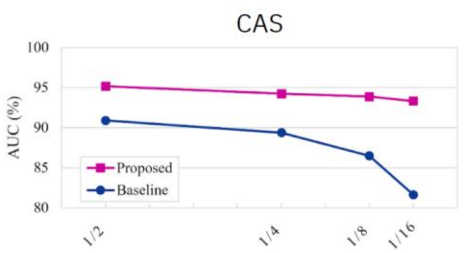
Experiments

Quantitative Performance
- AUC over varying labeled/unlabeled partitions
- Semi-supervised outperforms baseline on all partitions, and degrades much slower









Ablation Study
- BiGRU leverages strong temporal relationships (Inhalation/Exhalation cycle)
- Pre-training improves performance on classes with few examples (CAS, DAS)
- CPS leads to significant improvement across all categories

Component				AUC				F1			
CNN	BiGRU	Pretrain	CPS	Inhalation	Exhalation	DAS	CAS	Inhalation	Exhalation	DAS	CAS
✓				89.48%	74.22%	79.30%	84.36%	70.23%	32.61%	35.30%	32.17%
✓	✓			92.64%	79.10%	78.37%	86.49%	73.62%	38.57%	34.98%	34.08%
✓	✓	✓		92.44%	80.77%	83.79%	92.65%	73.43%	40.52%	44.19%	49.69%
✓	✓	✓	✓	**93.59%**	**83.99%**	**87.64%**	**93.88%**	**75.80%**	**45.76%**	**49.16%**	**53.57%**

Qualitative Performance
- proposed model provides results much closer to ground truth, esp. on classes with fewer examples (DAS/CAS)

1/2 labeled data   1/8 labeled data

Conclusion

Contributions
- Demonstrated that semi-supervised techniques are a viable approach in pulmonary auscultation analysis
  - Applied CPS in audio domain and demonstrated that it outperforms purely supervised baseline models
  - Showed that pretraining on large-scale audio datasets can further improve performance in this task

Future Work
- Use labeled and unlabeled data from different sources to more accurately reflect real world
- Create large-scale unlabeled datasets to further improve diagnostic accuracy and allow for noninvasive long-term patient monitoring