AI-empowered mHealth:

Pioneering Applications and Overcoming Key Challenges

Tong Xia

https://xtxiatong.github.io | tx229@cam.ac.uk

Postdoc & PhD @ University of Cambridge

Master @ Tsinghua University

Bachelor @ Wuhan University

UNIVERSITY OF CAMBRIDGE

清華大學 Tsinghua University

武漢大學 WUHAN UNIVERSITY

Content

- Research background
- Research overview
- Audio-based mHealth
- Future work

Content

Public health challenges

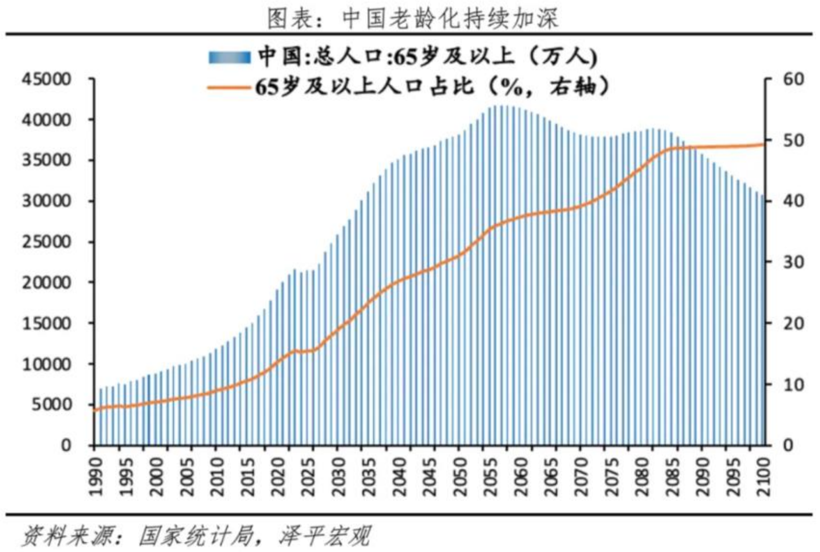
Cost-effective and scalable mHealth solutions

- Shortage of medical resources

- About **47%** of the global population lacks access to adequate diagnostic services
- Pandemic and epidemic

- Aging population

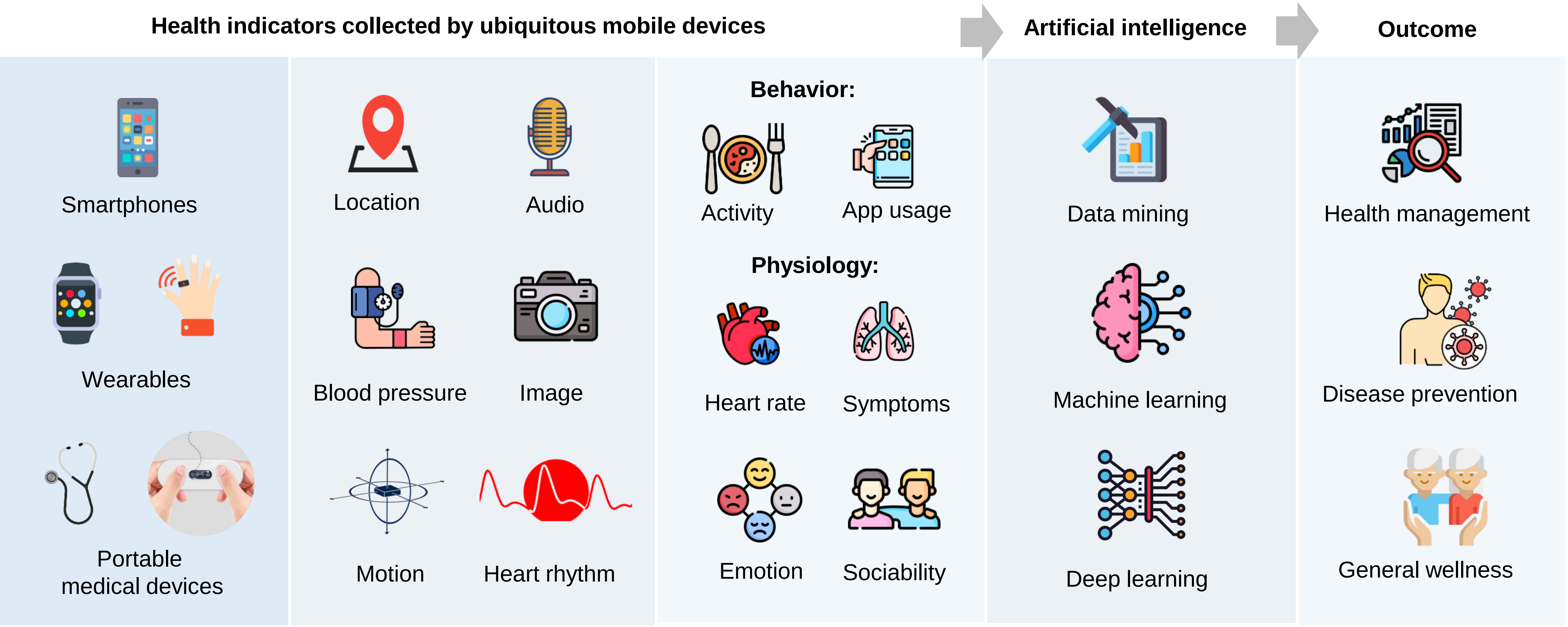
- It is estimated that by 2100, people over the age of 65 will make up half of China's population.



[1] Kenneth A Fleming, et al. The lancet commission on diagnostics: transforming access to diagnostics. The Lancet, 398(10315): 1997–2050, 2021

[2] 中国老龄化报告2024_腾讯新闻 (qq.com)

AI-empowered mHealth

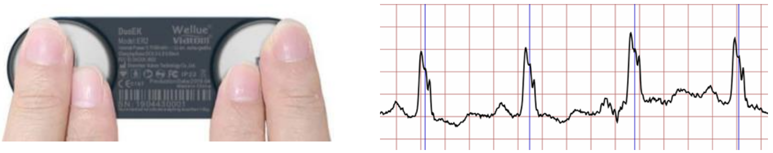


Content

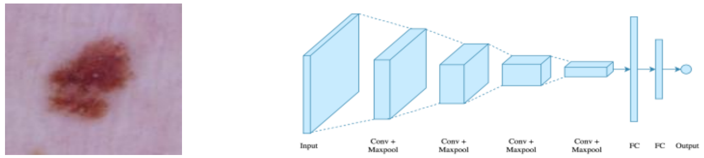
Applications

Active sensing

Audio-driven respiratory health screening
(KDD’20, NPJ DM’21)

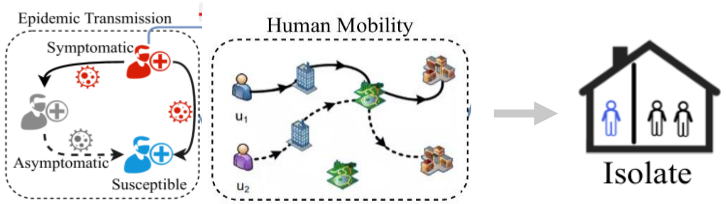


ECG(electrocardiogram)-based heart arrythmia detection
(WHI’22, IEEE JBHI’24)



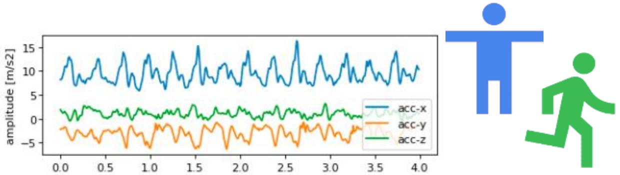
Dermoscopic image-based skin lesion prediction
(KDD FL4Data’23, IEEE JBHI’24)

Passive sensing



Precise mobility intervention for epidemic control
(BigData’21, KDD’22)

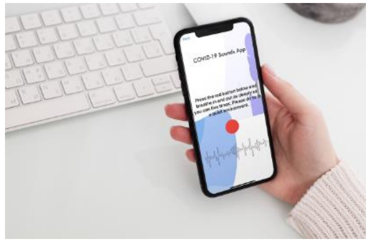
Predicting hospital visits from GPS records
(UbiComp’21)



Accelerometer-based human activity recontinuation
(KDD’24)

Key Challenges

Active sensing



1. Difficult to collect large data
2. Limited health annotation

- **Collecting 550h+ dataset *COVID-19 Sounds***
- **Proposing acoustic foundation models OPERA**

(NeurIPS'21, arXiv' 24)

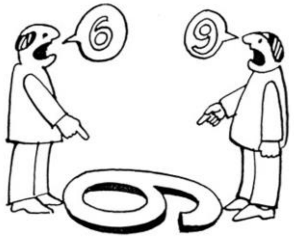
Passive sensing

3. Data complexity and missingness
4. Implicit association with health

- **Mobile sensing data generation method**
- **Dynamic graph neural network approach**

(AAAI'20, KDD'22, UbiComp'20)

Both active and passive sensing

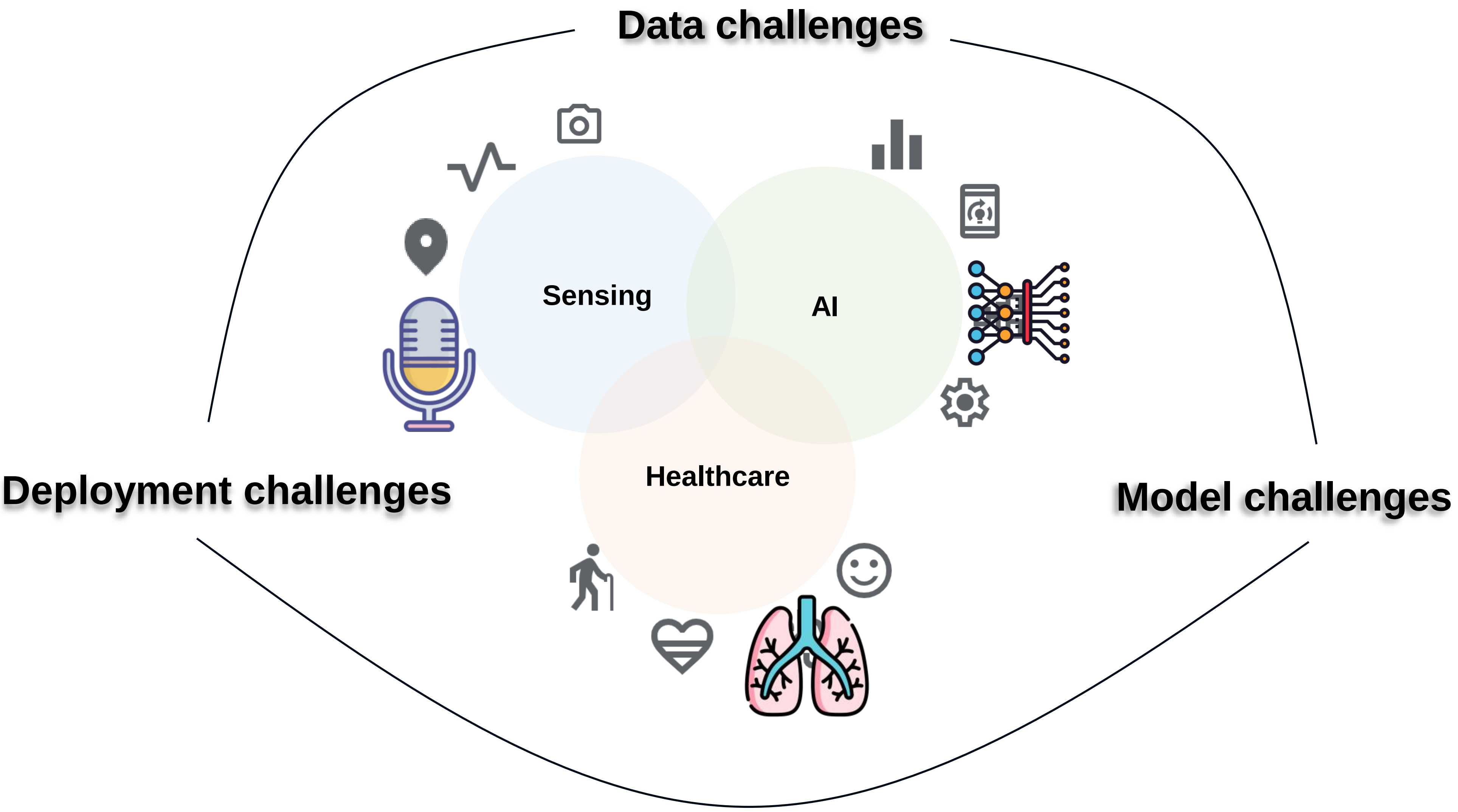


5. Model robustness and efficiency
6. Individual health data privacy

- **Efficient uncertainty quantification for mHealth**
- **Decentralised model learning for mobile sensing data**

(INTERSPEECH'21, ICASSP'23, IEEE JBHI'24, KDD'24)

AI-empowered mHealth



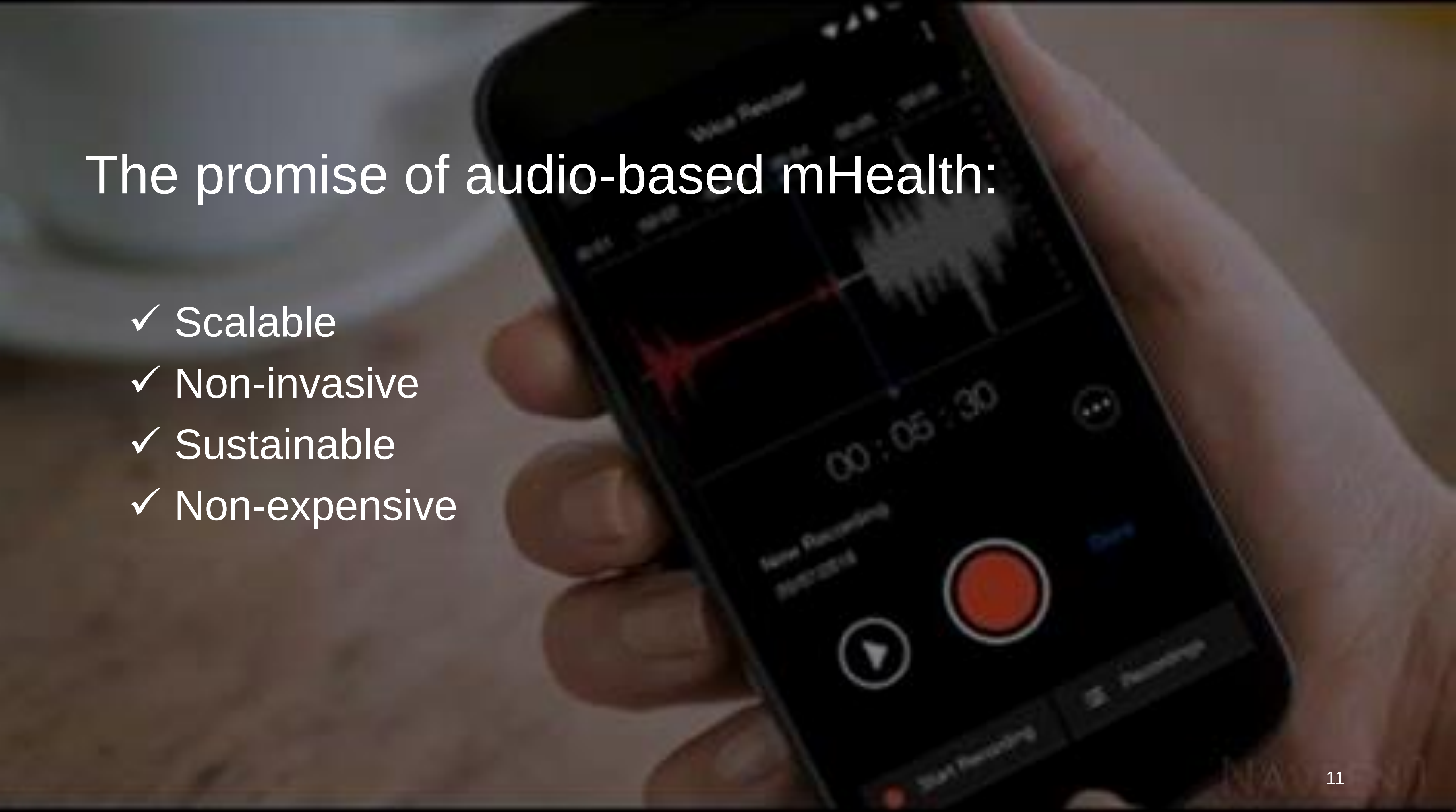
Content

Research background

Research overview

Audio-based mHealth

Future work



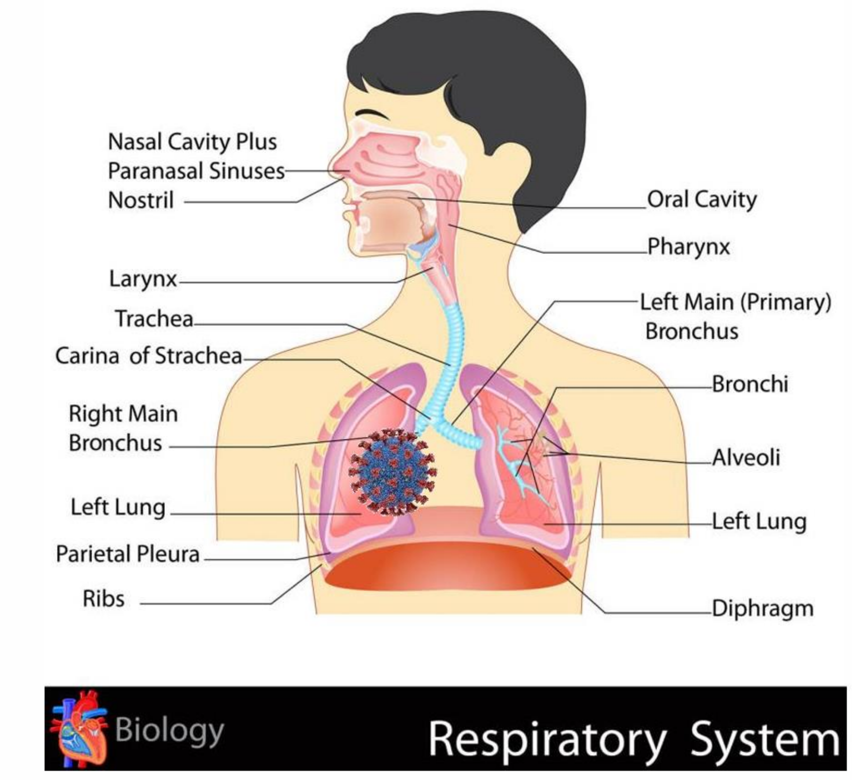
The promise of audio-based mHealth:

- ✓ Scalable
- ✓ Non-invasive
- ✓ Sustainable
- ✓ Non-expensive

Audio for respiratory health

A non-invasive and ubiquitous **screening protocol**, which would allow individual prescreening ‘anywhere’, ‘anytime’, in real time, and available to ‘anyone’.

- ✓ Lung is the main organ involved and **infected** by virus, which leads to some **changes** in respiratory sounds[3].
- ✓ Audio-based methods are promising in detecting cough-related disease like **pertussis**[4], **croup**[5], and **tuberculosis**[6].



[3] Hui Huang, et al. The respiratory sound features of COVID-19 patients fill gaps between clinical data and screening methods[J]. medRxiv, 2020.
[4] R. X. A. Pramono, et al, “A cough-based algorithm for automatic diagnosis of pertussis,” PloS one, vol. 11, no. 9, 2016
[5] R. V. Sharan, et al “Automatic croup diagnosis using cough sound recognition,” IEEE Transactions on Biomedical Engineering, vol. 66, no. 2, pp. 485–495, 2018. 4
[6] G. Botha, et al, “Detection of tuberculosis by automatic cough sound analysis,” Physiological measurement, vol. 39,no. 4, p. 045005, 2018

Audio-based respiratory health screening

How to **collect data** for mHealth research?

- A large-scale crowdsource dataset *COVID-19 Sounds*

(T. Xia*, et al. NeurIPS'21 🏆)

How to **design and train AI models** for mHealth applications?

- A CNN model for respiratory health screening

(T. Xia*, et al. Nature npj Digital Medicine' 22)

How to ensure **generalizability** of AI for mHealth?

- Pretrained acoustic foundation models for mHealth

(T. Xia*, et al. Under review)

Data collection through smartphones

✓ **A free, open, and easy-to-use data collection app**

covid-19-sounds.org

Sign-up survey

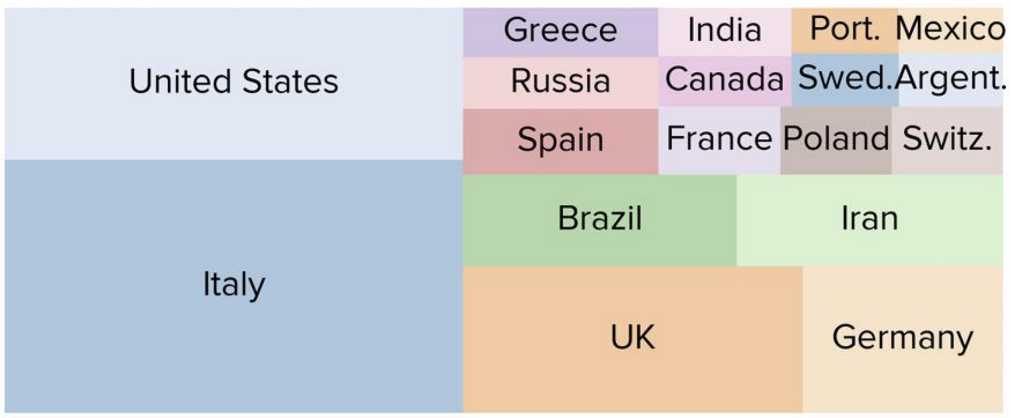
Demographics, Medical history, Smoking history, etc.

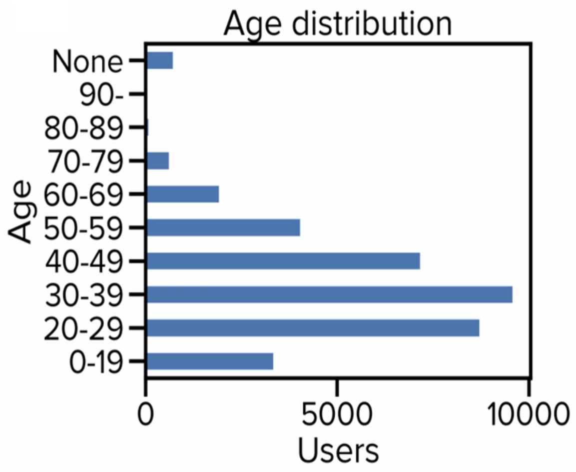
Daily survey

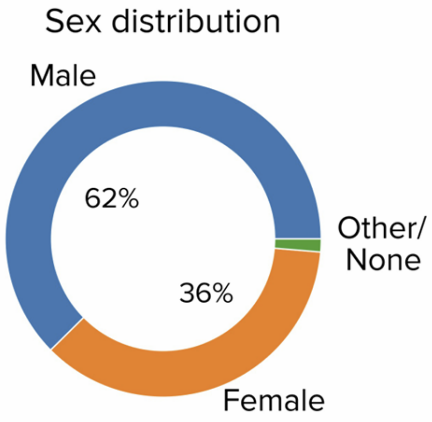
- ✓ **Symptoms**: fever, loss of taste, etc.
- ✓ **COVID-19 test**: positive, negative, never tested, etc.
- ✓ **Audio**: breathing, cough and some voice.

[7] **T. Xia***, D. Spathis*, C. Mascolo, et al. COVID-19 Sounds: A Large-Scale Audio Dataset for Digital Respiratory Screening. **NeurIPS** Datasets and Benchmarks Track 2021

Data collection through smartphones







	Android App	IOS App
Asthma	2,381	1,792
COPD^	331	119
COVID-19 positive	708	674
COVID-19 negative	3,497	2,953
COVID-19 not tested	16,610	13,473
Total samples	28,286	19,759

OPEN DATA

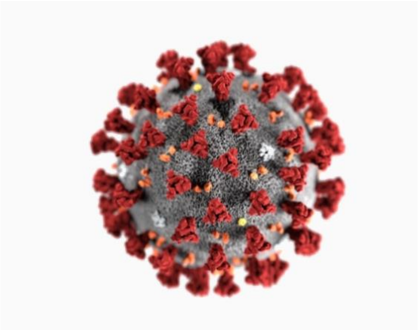
Comprehensive:

- ✓ Worldwide reach
- ✓ Multiple languages
- ✓ Demographic range
- ✓ Diverse health conditions
- ✓ **Multiple audio modalities**
- ✓ **Large scale (552 hours)**

[8] **T. Xia**⋆, D. Spathis⋆, C. Mascolo, et al. COVID-19 Sounds: A Large-Scale Audio Dataset for Digital Respiratory Screening. **NeurIPS** Datasets and Benchmarks Track 2021

^COPD: Chronic obstructive pulmonary disease

A CNN model for respiratory health screening



Data collection → Data curation → Model development → Performance evaluation

An application:

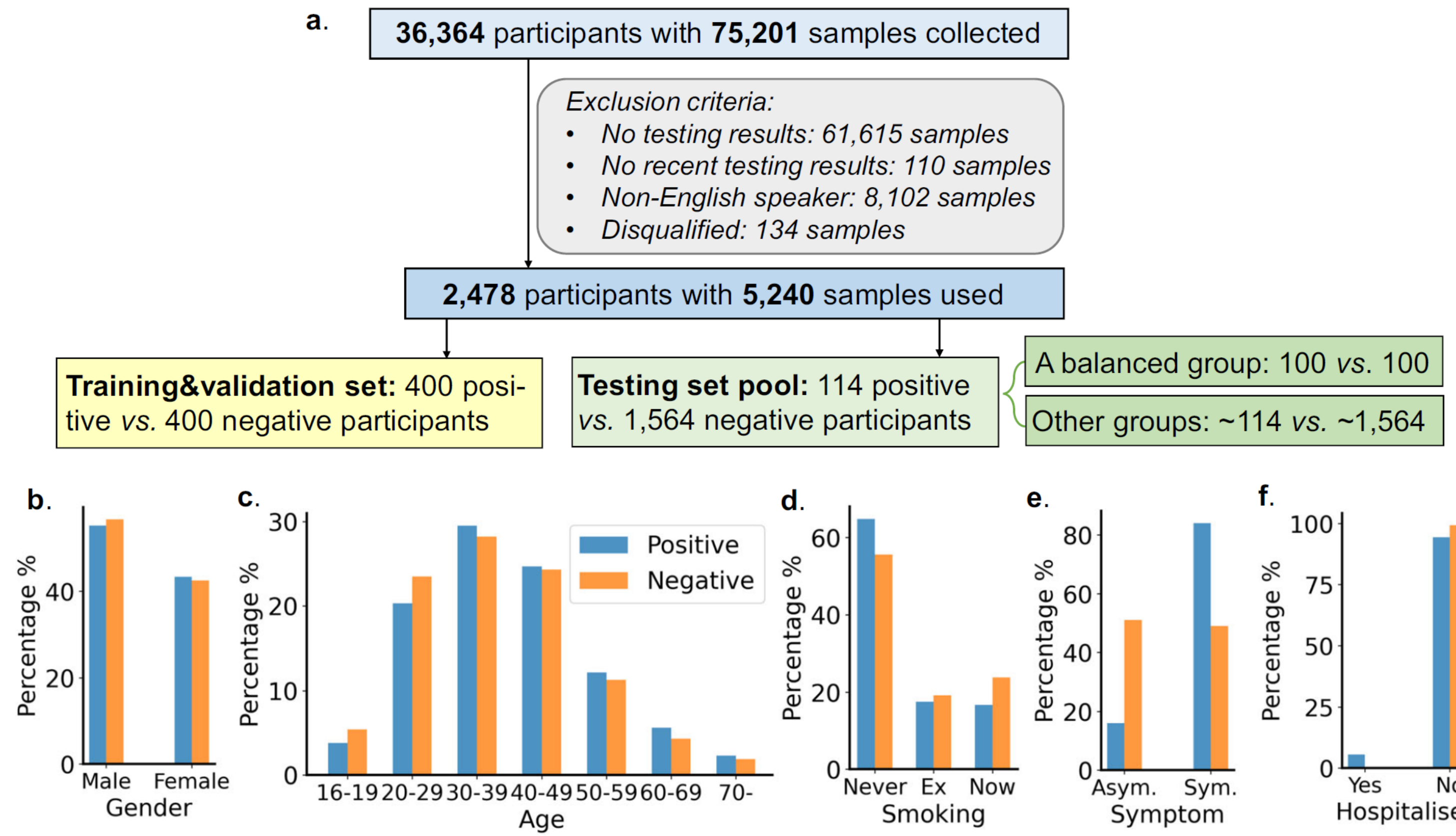
COVID-19 prediction

distinguish between **tested-negative** controls, including symptomatic cases which could be caused by another infection and **tested-positive** patients, including asymptomatic cases

[8] J. Han*, **T. Xia***, et al. Sounds of COVID-19: Exploring Realistic Performance of Audio-based Digital Testing. **Nature NPJ Digital Medicine** 2022

A CNN model for respiratory health screening

Data collection → Data curation

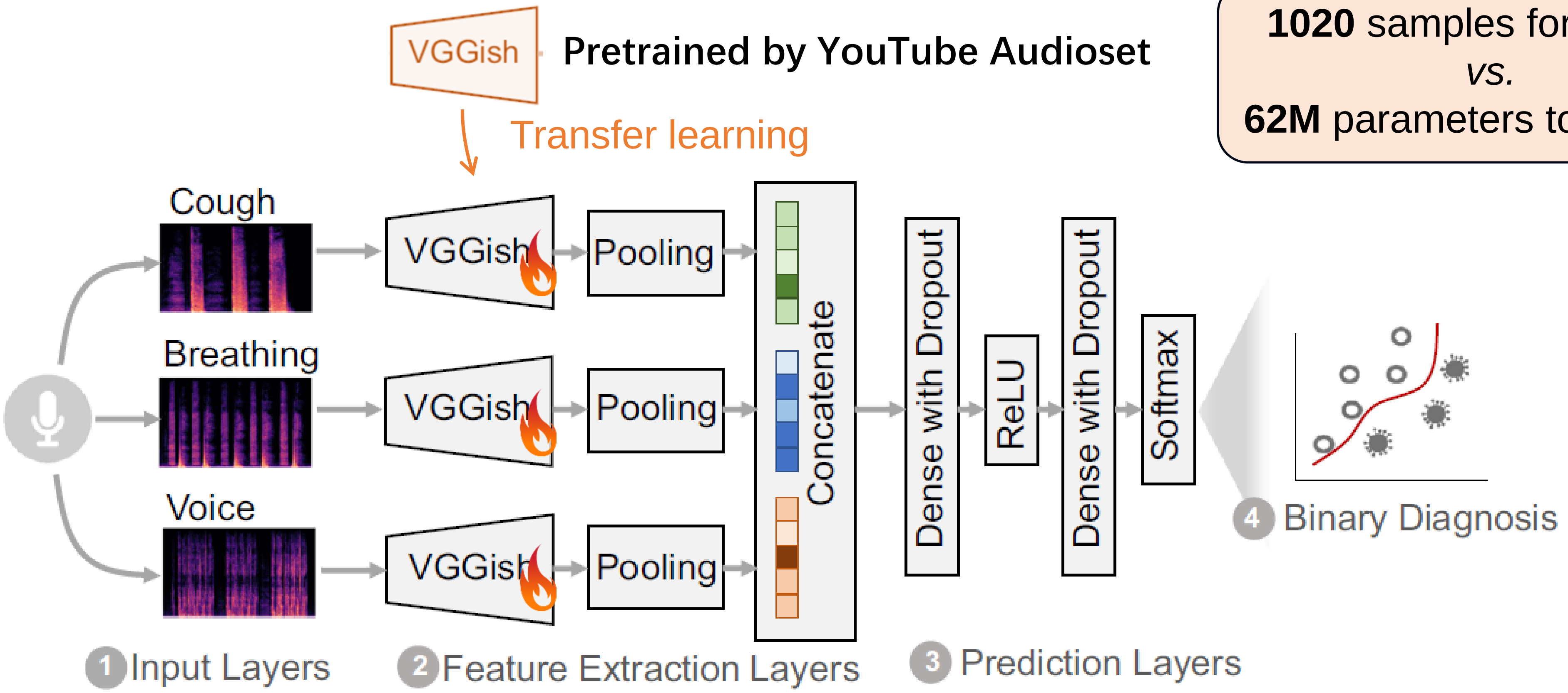


[9] J. Han*, **T. Xia***, et al. Sounds of COVID-19: Exploring Realistic Performance of Audio-based Digital Testing. **Nature NPJ Digital Medicine** 2022

A CNN model for respiratory health screening

Data collection → Data curation → Model development

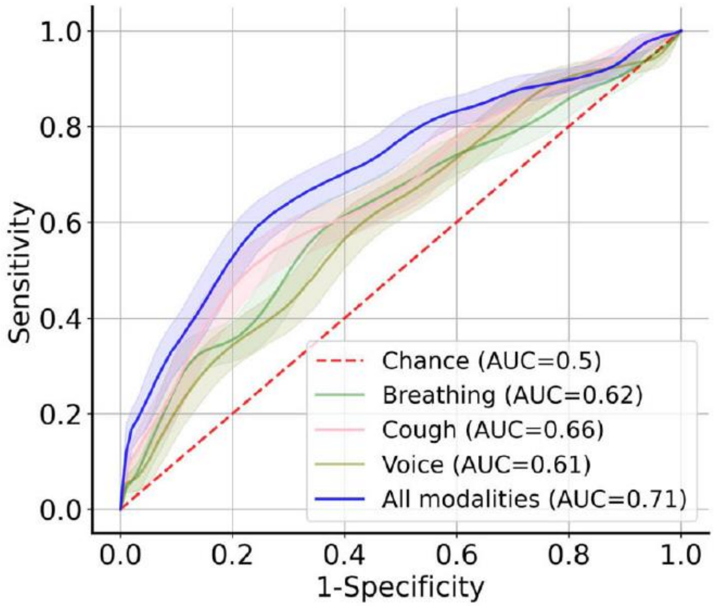
1020 samples for training
vs.
62M parameters to optimize

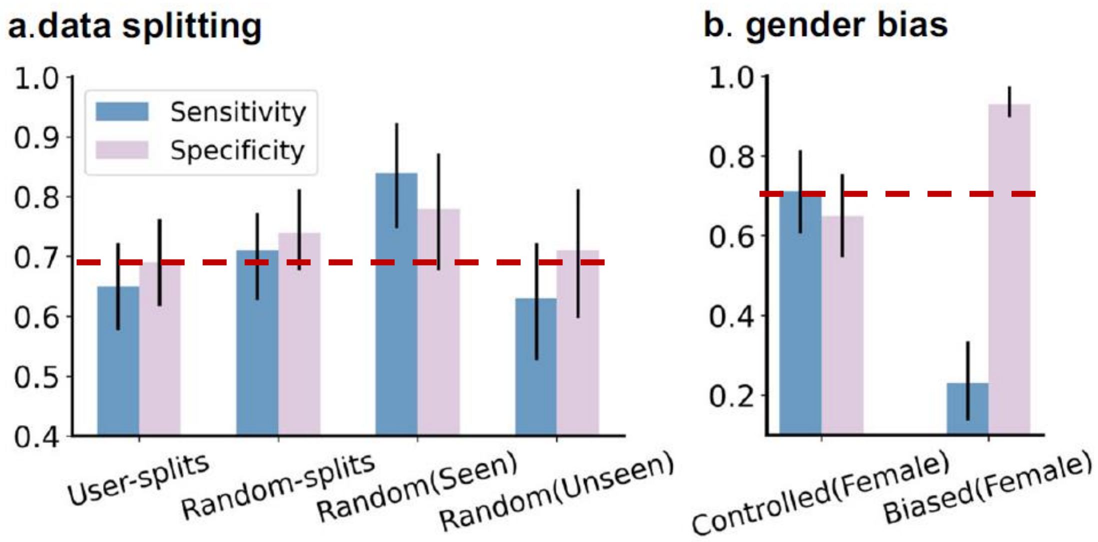


[9] J. Han*, **T. Xia***, et al. Sounds of COVID-19: Exploring Realistic Performance of Audio-based Digital Testing. **Nature NPJ Digital Medicine** 2022

A CNN model for respiratory health screening

Data collection → Data curation → Model development → Performance evaluation



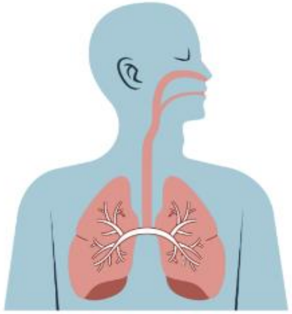


Audio: AUROC of 0.71, Sensitivity of 0.65, Specificity of 0.69
Fast flow test: Sensitivity ranges from 0.37 to 0.99

[9] J. Han*, **T. Xia***, et al. Sounds of COVID-19: Exploring Realistic Performance of Audio-based Digital Testing. **Nature NPJ Digital Medicine** 2022

How to ensure generalizability for mHealth?

AI-empowered acoustic mHealth application



- ✓ Asthma diagnose
- ✓ COPD prediction
- ✓ Smoking history estimation

- ✓ Spirometry inference
- ✓ Vital capacity prediction
- ✓ Respiratory rate estimation

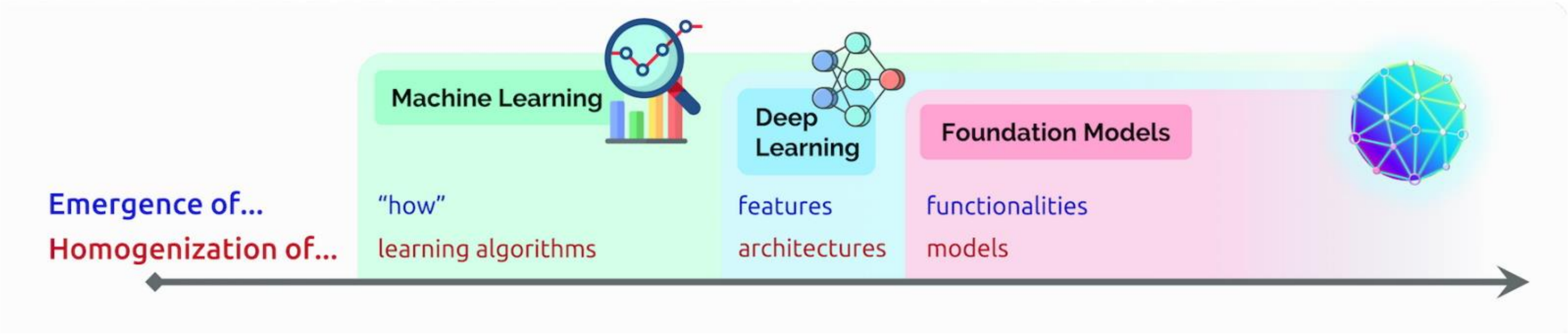
- ✓ Murmur prediction
- ✓ Heart abnormity detection

- ✓ Snoring recognition
- ✓ Body position prediction
- ✓ Sleep apnea detection

Task specific model for each application?

[10] E. Zhang*, **T. Xia***, et al. Towards Open Respiratory Acoustic Foundation Models: Pretraining and Benchmarking. https://arxiv.org/abs/2406.16148

How to ensure generalizability for mHealth?



Task-specific → Task-agnostic

[Beginner's Guide: Using Foundation Models in ML Projects (labellerr.com)](https://labellerr.com)

Pretrained acoustic foundation models for mHealth

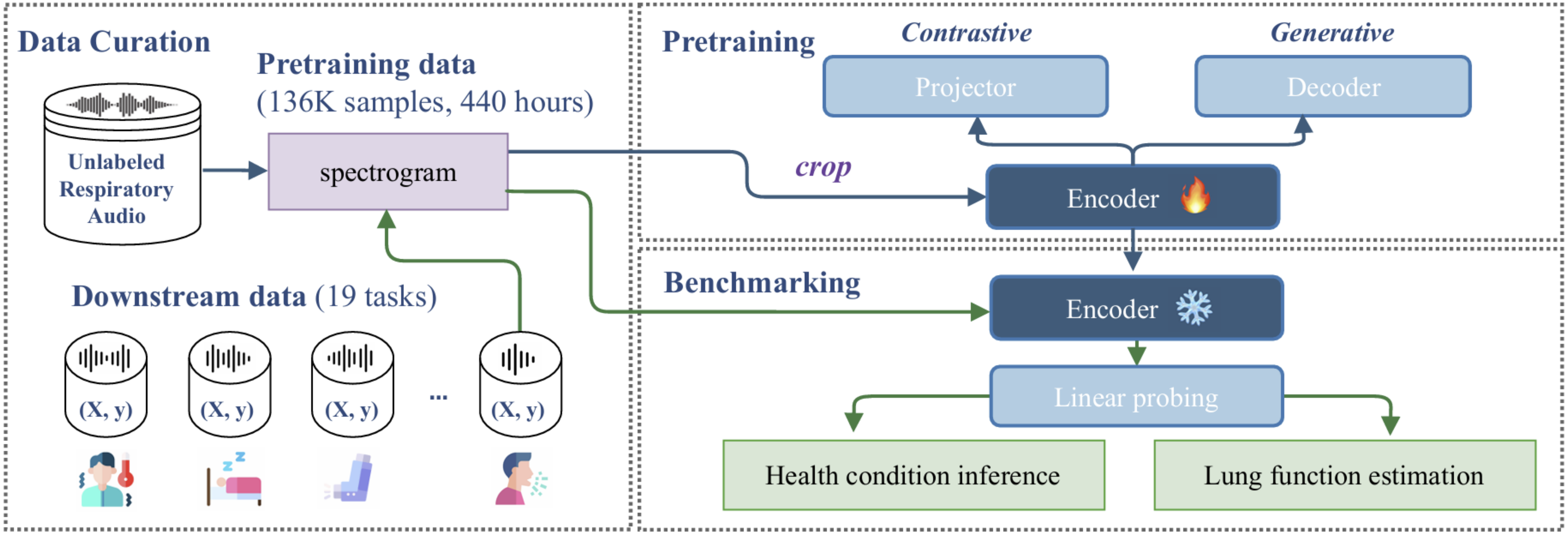


Figure 1: System overview of OPERA. After data curation, respiratory acoustic foundation models (Encoder) are pretrained and then evaluated on various downstream health tasks.

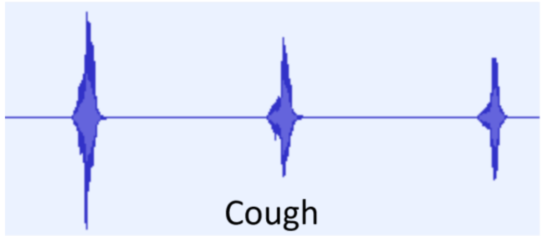
[10] E. Zhang*, **T. Xia***, et al. Towards Open Respiratory Acoustic Foundation Models: Pretraining and Benchmarking. https://arxiv.org/abs/2406.16148

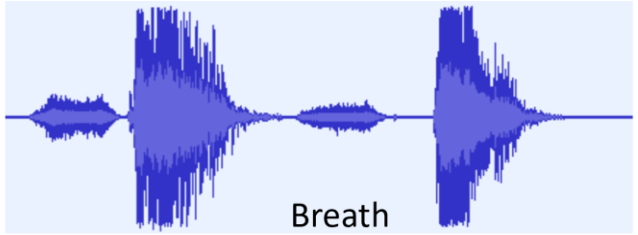
Pretrained acoustic foundation models for mHealth

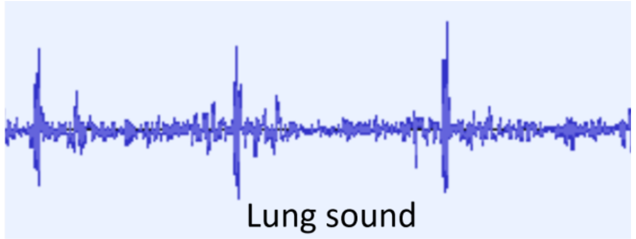
Data curation

- We curate a unique large-scale (~136K samples, 440 hours), multi-source (5 datasets), multi-modal (breathing, coughing, and lung sounds) respiratory audio dataset for foundation model pretraining

Data name	Collected by	SR	Modality	#Sample	Duration (s)	Crop (s)
COVID-19 Sounds [69]	Microphone	16∼44.1kHz	Induced cough (3 times)	40866	6.1[2.6∼11.2]	2
			Deep breath (5 times)	36605	20.5[9.7∼31.6]	8
UK COVID-19 [12]	Microphone	48kHz	Induced cough (3 times)	19533	4.1[2.1∼9.2]	2
			Exhalation (5 times)	20719	7.7[4.2∼15.6]	4
COUGHVID [47]	Microphone	48kHz	Induced cough (up to 10s)	7179	6.9[2.4∼9.9]	2
ICBHI [51]	Stethoscope	4∼44.1kHz	lung sound (several breath cycles)	538	22.2[20.0∼65.9]	8
HF LUNG [31]	Stethoscope	4kHz	lung sound (several breath cycles)	10554	15.0[15.0∼15.0]	8

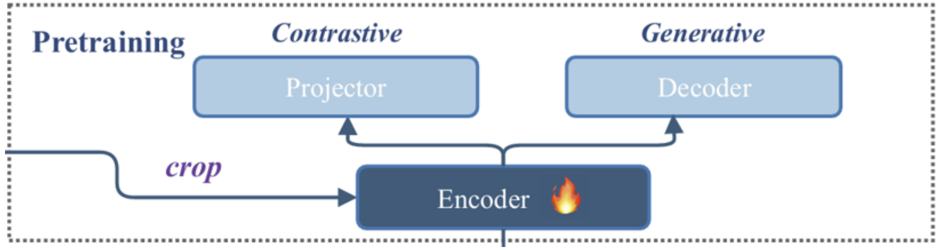


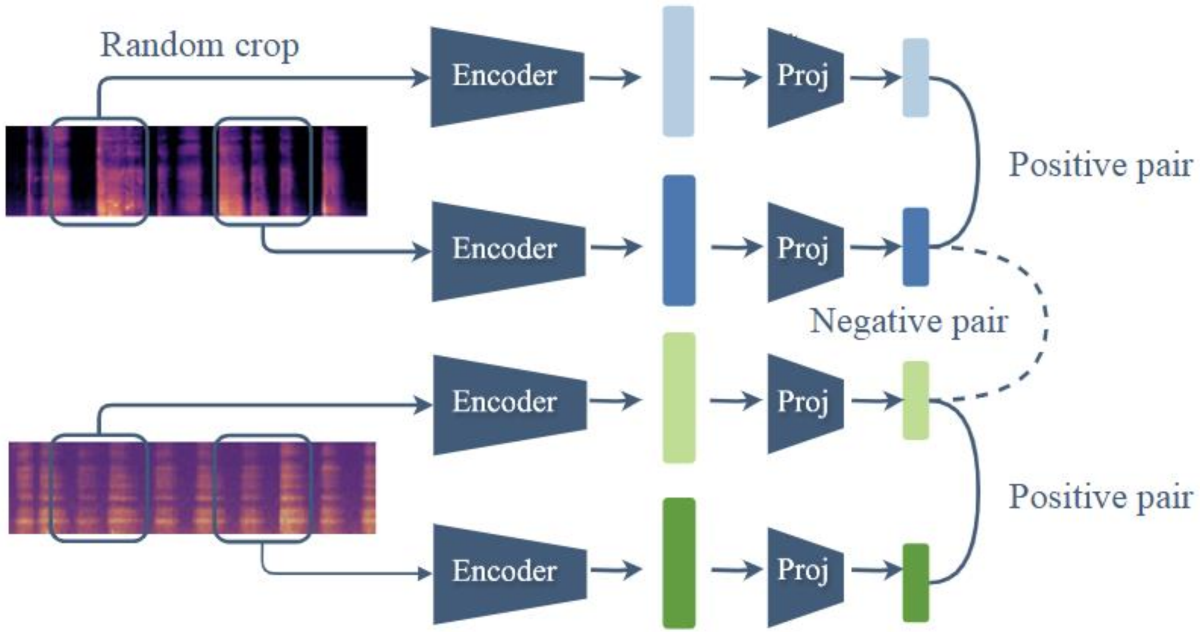




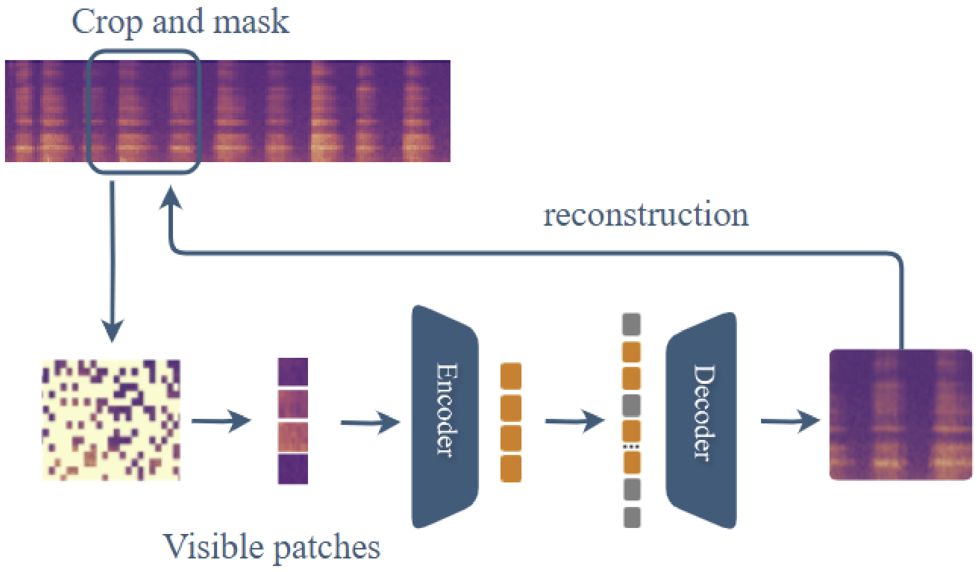
Pretrained acoustic foundation models for mHealth

Data curation → Model pretraining





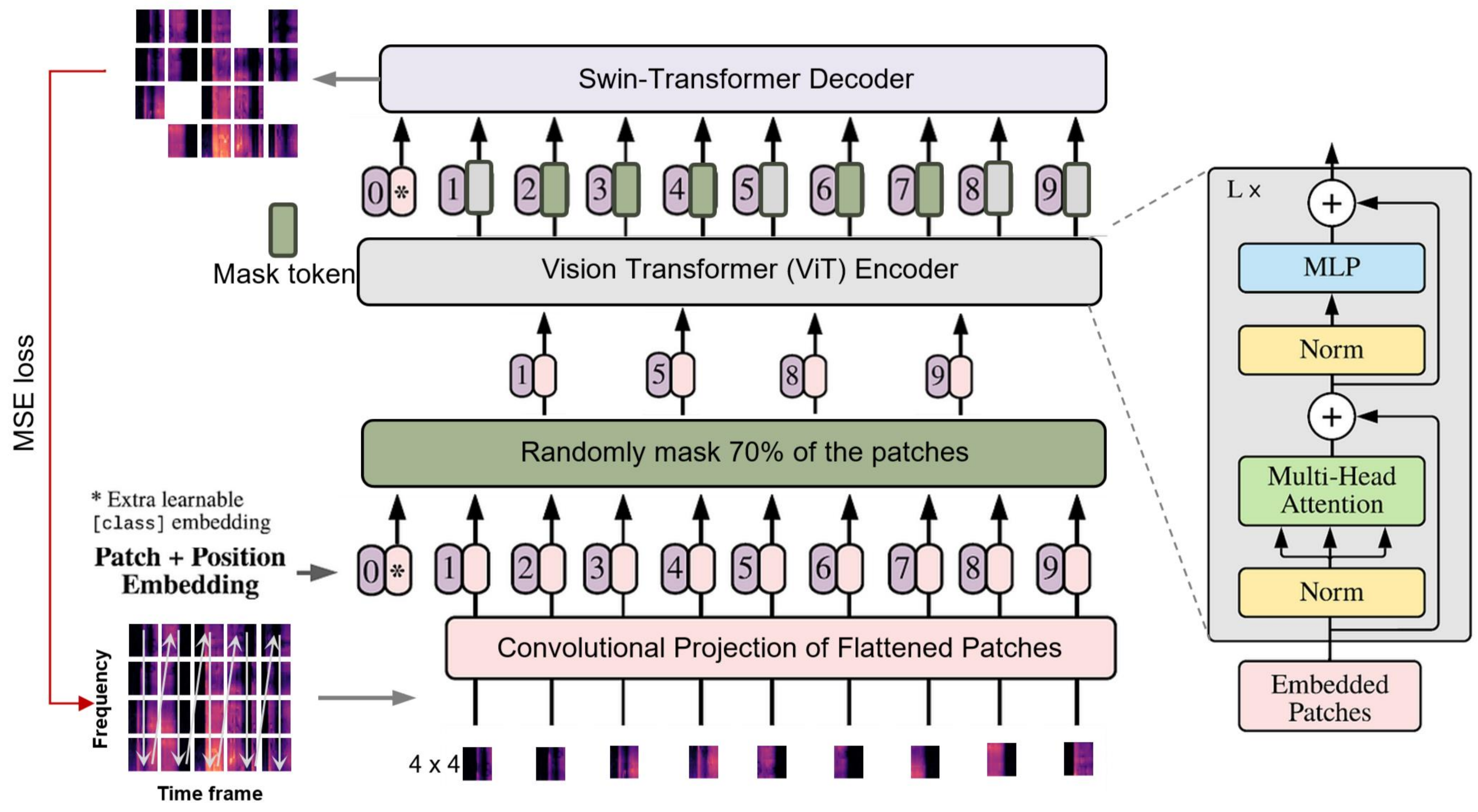
(a) Contrastive (OPERA-CT, OPERA-CE)



(b) Generative (OPERA-GT)

Pretrained acoustic foundation models for mHealth

Data curation → Model pretraining



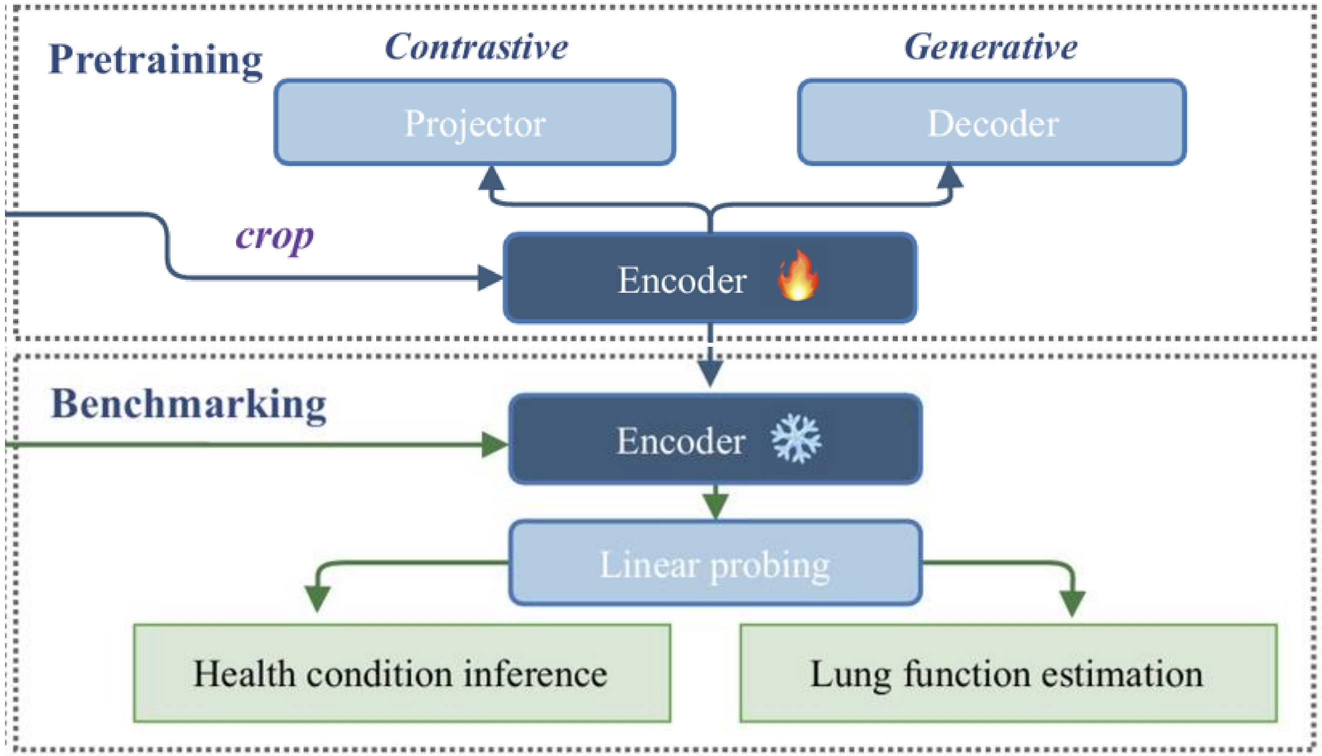
Mask Auto Encoder (MAE) pre-training

Pretrained acoustic foundation models for mHealth

Data curation → Model pretraining → Performance evaluation

Large-scale unlabeled data

Small labeled data

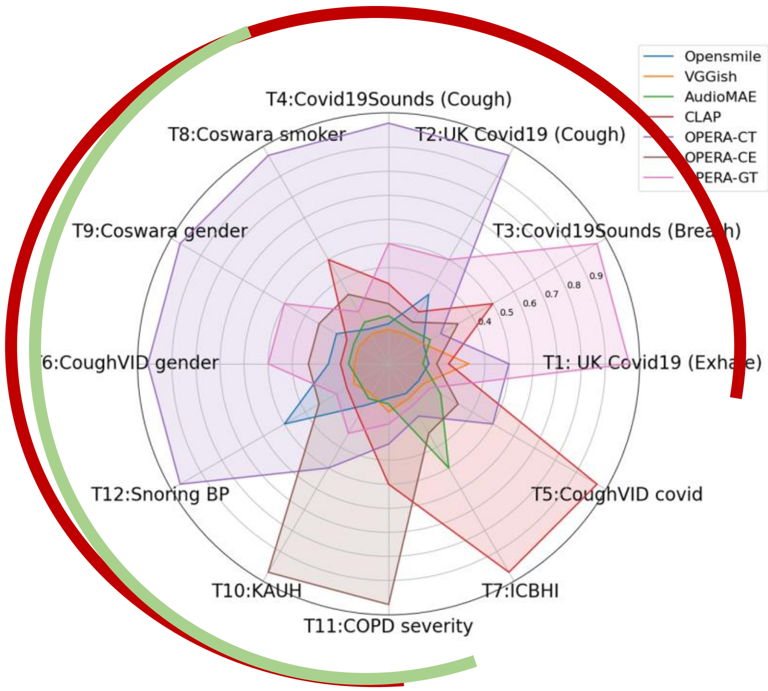


x12 (categorical)

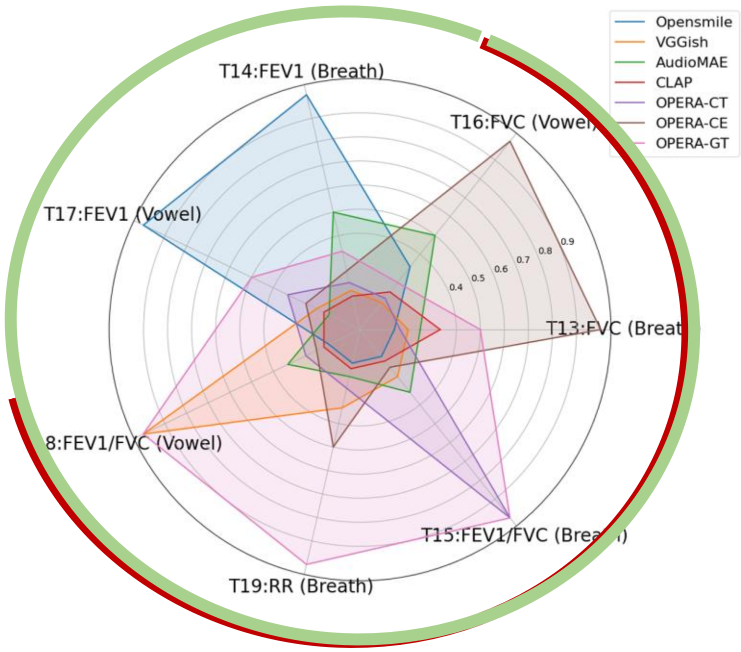
x7 (continues)

Pretrained acoustic foundation models for mHealth

Data curation → Model pretraining → Performance evaluation



(a) Health Condition Inference



(b) Lung Function Estimation

- ✓ Outperform baselines on 15 (10+5) out of 19 tasks
- ✓ Generalizable to unseen data and new respiratory audio modalities

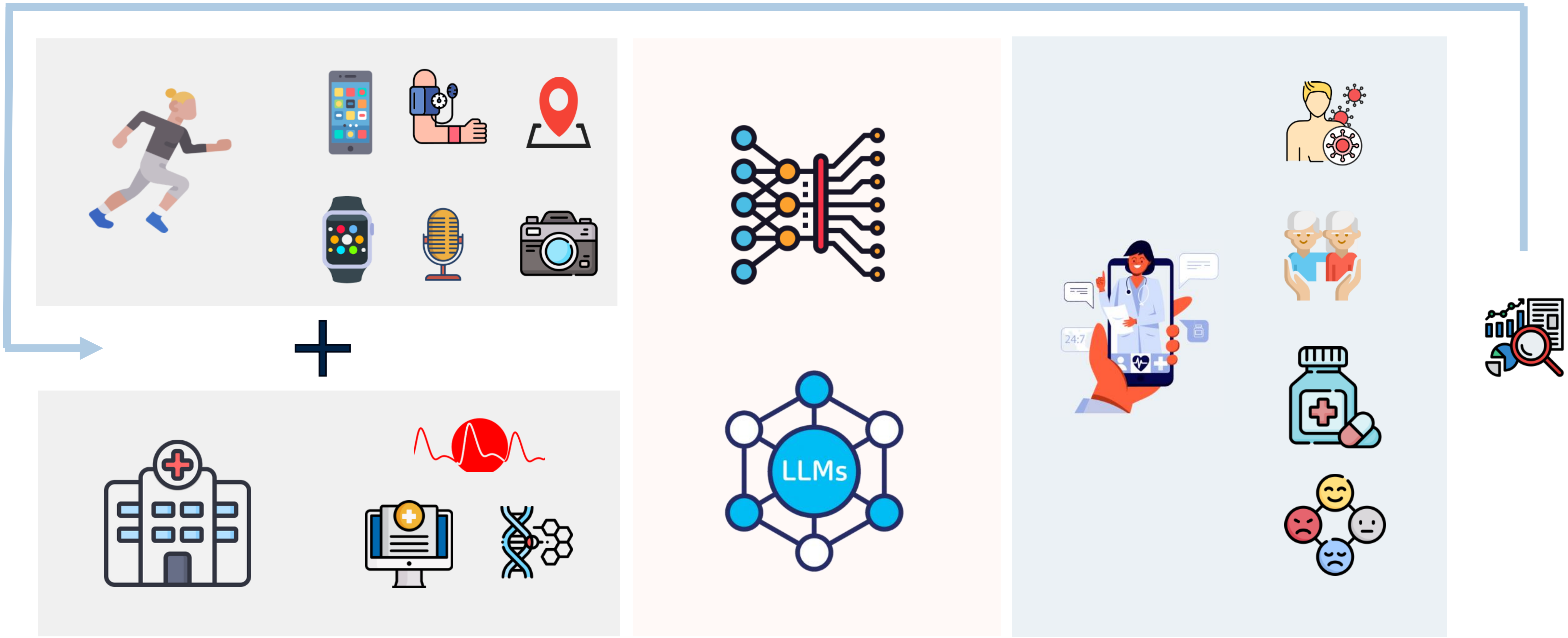
OPERA

We make everything open for research:
https://github.com/evelyn0414/OPERA.git

mHealth Applications

What is the future?

- Intervention effects loop back to improve the system



- Multi-modal mobile data and clinical data reunited
- Collaboration with doctors to define clinical meaningful tasks

- Powerful foundation models and LLM agents
- Efficient and decentralized model fine-tuning

- Delivery explainable and reliable healthcare on personal mobile devices

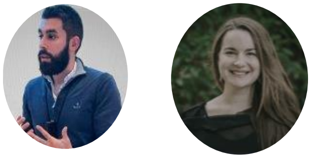
Acknowledgement

Academia collaborators

...

Industry collaborators

NOKIA BELL LABS

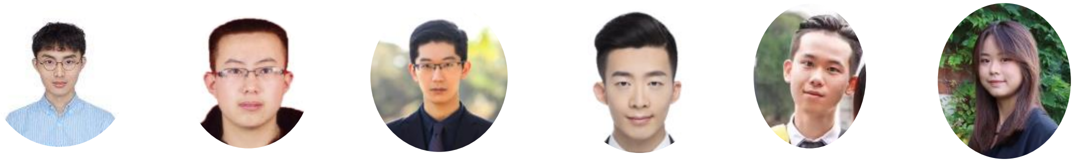


腾讯 Tencent

Flower Labs

HITACHI Inspire the Next

Mentored students



Examiners' Prize for a Highly Commended MPhil Project Report (Top 2 in 55 students)

...

Grants

国家自然科学基金委员会 National Natural Science Foundation of China

HUAWEI

erc European Research Council Established by the European Commission

Reference

PhD thesis:

- **T. Xia.** Reliable and decentralised deep learning for physiological data. PhD Thesis 2024.

Publications (^equal contribution):

- C. Brown^, J. Chauhan^, A. Grammenos^, J. Han^, A. Hasthanasombat^, D. Spathis^, T. Xia^, P. Cicuta, and C. Mascolo. Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data. **KDD** 2020 (Google scholar citation 487, Cambridge Frame Hall **'Better Future' award**)
- J. Han^, **T. Xia**^, et al. Sounds of COVID-19: Exploring Realistic Performance of Audio-based Digital Testing. **Nature NPJ Digital Medicine** 2022 (Google scholar citation 78)
- **T. Xia**^, D. Spathis^, C. Mascolo, et al. COVID-19 Sounds: A Large-Scale Audio Dataset for Digital Respiratory Screening. **NeurIPS** Datasets and Benchmarks Track 2021 2020 (**the 2nd poster award** at the Precision Health Initiative Launch Symposium Cambridge, Google scholar citation 73)
- **T. Xia**, J. Han, L. Qendro, T. Dang, and C. Mascolo. Uncertainty-aware COVID-19 Detection from Imbalanced Sound Data. **INTERSPEECH** 2021 (**Student Travel Grant award**, Google scholar citation 30)
- **T. Xia**, et al. Attnmove: History Enhanced Trajectory Recovery via Attentional Network. **AAAI** 2021. (Google scholar citation 43)
- **T. Xia**, J. Han, A. Ghosh, and C. Mascolo. Cross-device Federated Learning for Mobile Health Diagnostics: A First Study on COVID-19 Detection. **ICASSP** 2023
- **T. Xia**, A. Ghosh, X. Qiu, and C. Mascolo. FLea: Addressing Data Scarcity and Label Skew in Federated Learning via Privacy-preserving Feature Augmentation. **KDD** 2024
- **T. Xia**, J. Han, C. Mascolo. Benchmarking Uncertainty Quantification on Biosignal Classification Tasks under Dataset Shift. Workshop on Health Intelligence, AAAI 2022
- **T. Xia^** and A. Ghosh^. Mobility-based Individual POI Recommendation to Control the COVID-19 Spread. IEEE **Big Data** 2021.
- T. Feng, **T. Xia**, et al. Precise Mobility Intervention for Epidemic Control Using Unobservable Information via Deep Reinforcement Learning. **KDD** 2022
- E. Bondareva, **T. Xia**, J. Han, C. Mascolo. Towards Uncertainty-Aware Murmur Detection in Heart Sounds via Tandem Learning. CinC 2022

Reference

- **T. Xia**, T. Dang, J. Han, L. Qendro, and C. Mascolo. Class-balanced Evidential Deep Learning for Health Diagnostics. IEEE JBHI 2024
- **T. Xia**, J. Han, L. Qendro, and C. Mascolo. Exploring Machine Learning for Audio-based Respiratory Condition Screening: A Concise Review of Databases, Methods, and Open Issues. JEBM 2022

Under review:

- J. Han, **T. Xia**, C. Mascolo. Audio-based Sleep Apnea Detection from Tracheal and Ambient Sound Recordings. Under review.
- E. Zhang^, **T. Xia^**, et al. Towards Open Respiratory Acoustic Foundation Models: Pretraining and Benchmarking. https://arxiv.org/abs/2406.16148

Other publications (^equal contribution):

- T. Li, **T. Xia**, H. Wang, Z. Tu, S. Tarkoma, Z. Han, and P. Hui. Smartphone App Usage Analysis: Datasets, Methods, and Applications. IEEE Communications Surveys & Tutorials, 2022 (Google scholar citation 78)
- J. Han, C. Brown^, J. Chauhan^ A. Grammenos^, A. Hasthanasombat^, D. Spathis^, **T. Xi**a^, P. Cicuta, C. Mascolo. Exploring automatic COVID-19 diagnosis via voice and symptoms from crowdsourced data. **ICASSP** 2021 (**Google scholar citation 190**)
- **T. Xia**, Y. Yue, Y. Li, et al. Understanding Urban Dynamics via State-sharing Hidden Markov Model. IEEE **TKDE** 2021
- **T. Xia**, J. Lin, Y. Li, J. Feng, P. Hui, F. Sun, D. Guo, and D. Jin. 3DGCN: 3-dimensional Dynamic Graph Convolutional Network for Citywide Crowd Flow Prediction. ACM **TKDD**, 2021 (Google scholar citation 33)
- **T. Xia**, Y. Li, J. Feng, D. Jin, Q. Zhang, H. Luo, and Q. Liao. DeepApp: Predicting Personalized Smartphone App Usage via Context-aware Multi-task Learning . ACM TIST, 2020 (Google scholar citation 23)
- Y. Yu^, **T. Xia**^, H. Wang, J. Feng, Y. Li. Semantic-aware Spatio-temporal App Usage Representation via Graph Convolutional Network. **UbiComp** 2020 (Google scholar citation 27)
- **T. Xia**, Y. Li, J. Feng, D. Jin, Q. Zhang, H. Luo, and Q. Liao. Revealing Urban Dynamics by Learning Online and Offline Behaviours Together. **UbiComp** 2019
- Z. Han, **T. Xia**, Y. Xi, and Y. Li . Healthy Cities, A Comprehensive Dataset for Environmental Determinants of Health in England Cities. **Scientific Data.** 2023

UNIVERSITY OF CAMBRIDGE

QUEENS' COLLEGE CAMBRIDGE

THANK YOU

Wechat

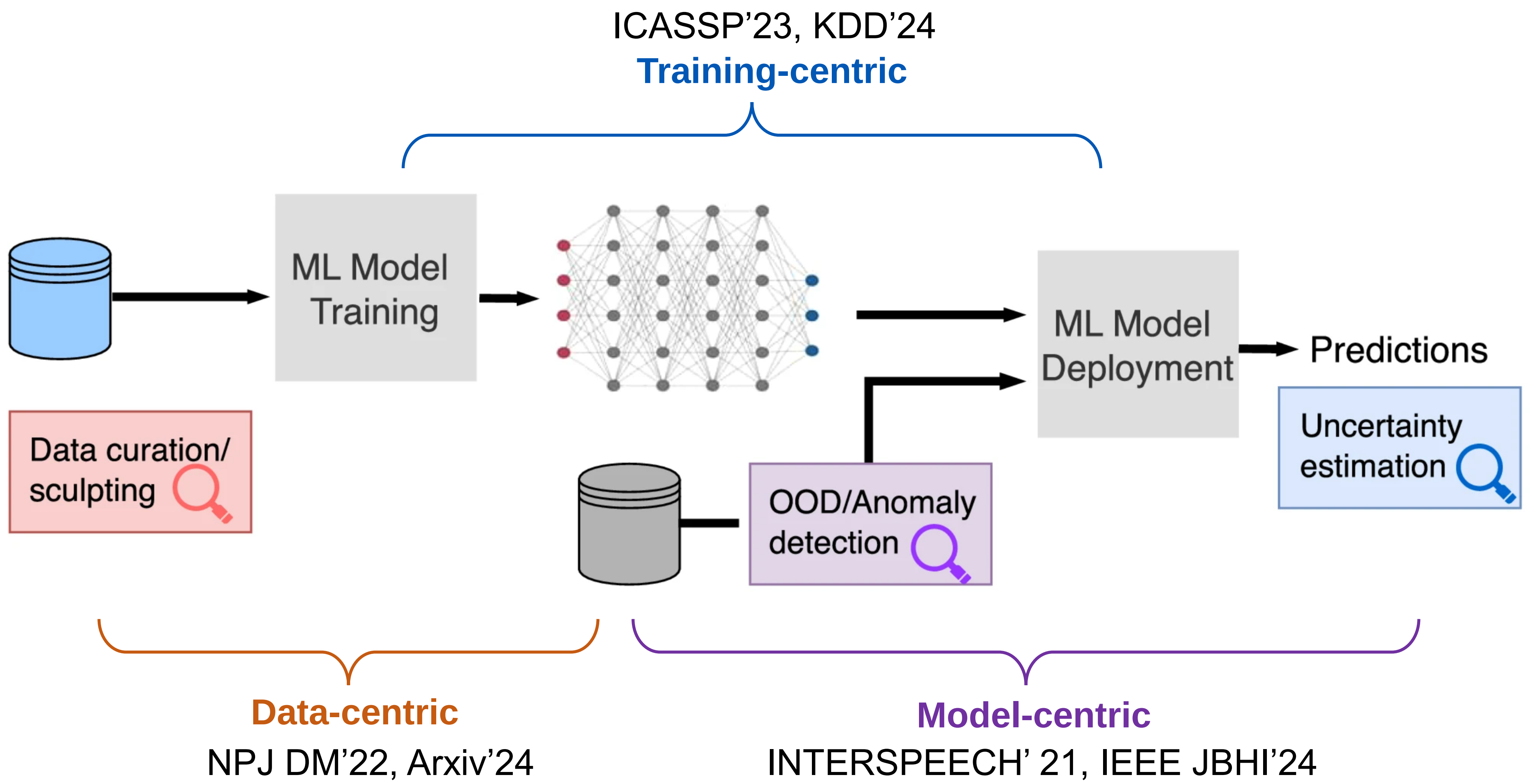
Linkedin

Google scholar

Tong Xia
tx229@cam.ac.uk

Backup

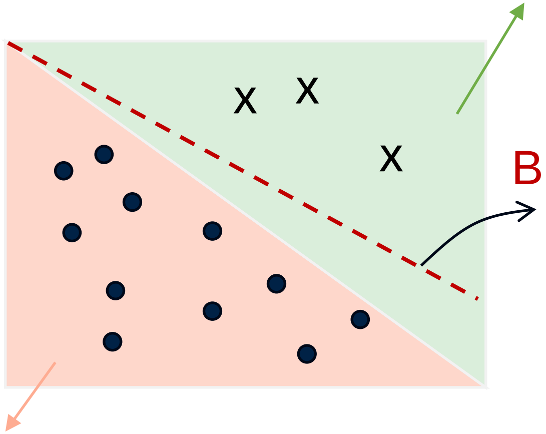
Trustworthy AI for mHealth



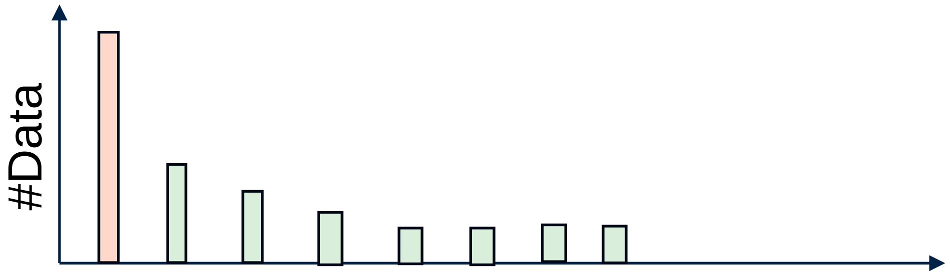
How to enable reliable mHealth in the wild?

- **Challenges:** Class imbalanced and model overconfidence

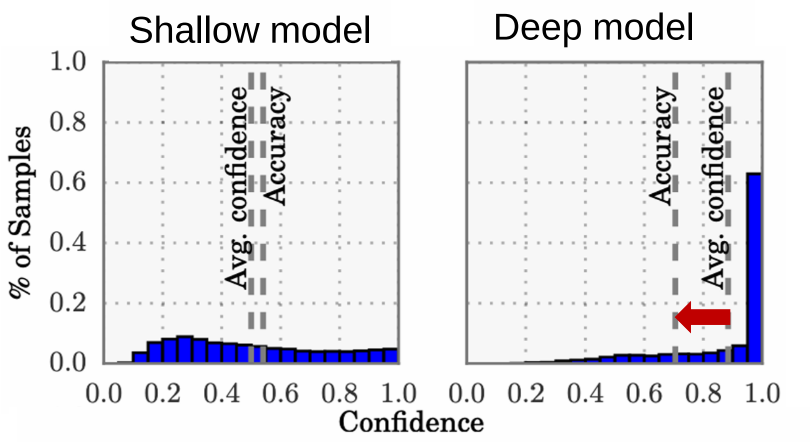
Real distribution for unhealthy samples (minority class)



Real distribution for healthy samples (majority distribution)



Deep learning overconfidence

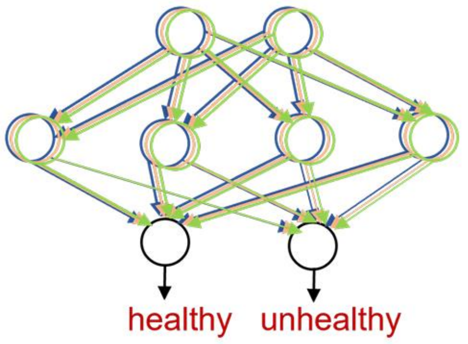


Sensitivity to distributional shift

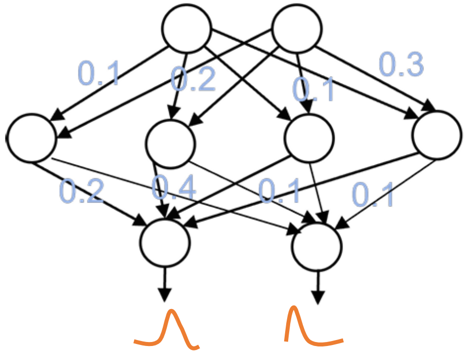
Class-balanced evidential deep learning for uncertainty

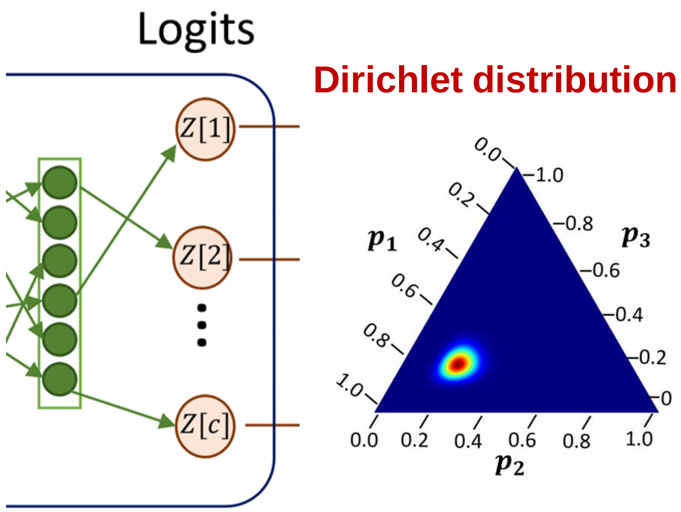
❑ Challenge: on-device efficiency

Ensemble approximation



Evidential deep learning:





$$\min_{\theta} \; \mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}^{(i)},$$

$$\mathcal{L}^{(i)} = \mathbb{E}_{\boldsymbol{p}^{(i)}\sim\boldsymbol{q}^{(i)}}[\mathcal{C}(\boldsymbol{p}^{(i)}, y^{(i)})] + \lambda\cdot\mathcal{L}_r^{(i)},$$

Optimizing the expected cross-entropy

- Instance level:

$$\mathcal{L}_r^{(i)} = KL[(Dir(\boldsymbol{\alpha}^{(i)})||Dir(\mathbf{1})]$$

$$\Downarrow$$

$$\mathcal{L}_r^{\prime(i)} = KL[Dir(\boldsymbol{\alpha}^{\prime(i)})||Dir(\boldsymbol{\beta})]$$

- Dataset level:

$$\min_{\theta} \; \mathcal{L} = \frac{1}{N}\sum_{c=1}^{C}\sum_{y^{(i)}\in c}\mathcal{L}^{(i)}$$

$$\Downarrow$$

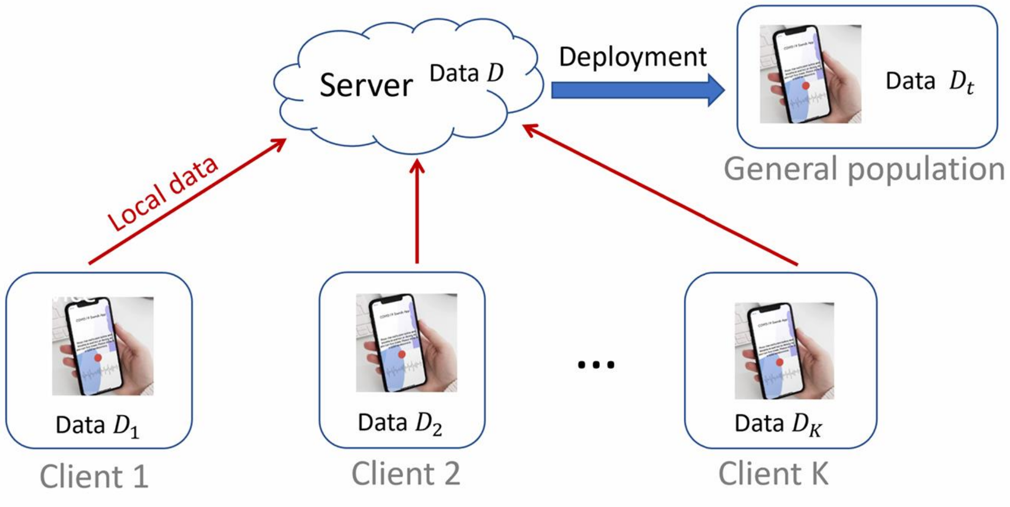
$$\mathcal{L}' = \frac{1}{C}\sum_{c=1}^{C}\frac{1}{N_c}\sum_{y^{(i)}\in c}\mathcal{L}^{(i)},$$

[6] **T. Xia**, T. Dang, J. Han, L. Qendro, and C. Mascolo. Class-balanced Evidential Deep Learning for Health Diagnostics. IEEE **JBHI** 2024

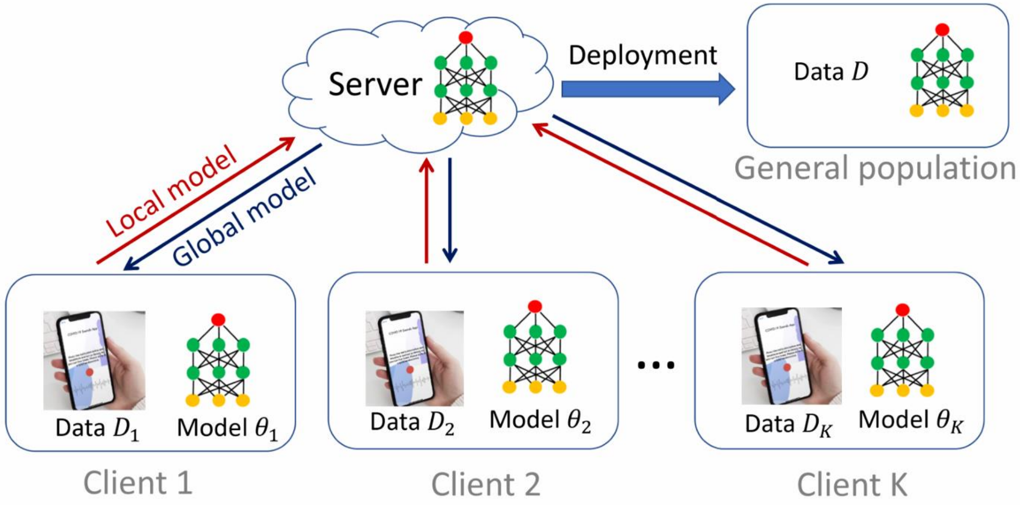
How to protect data privacy for mHealth?

- **Challenge: Data privacy**

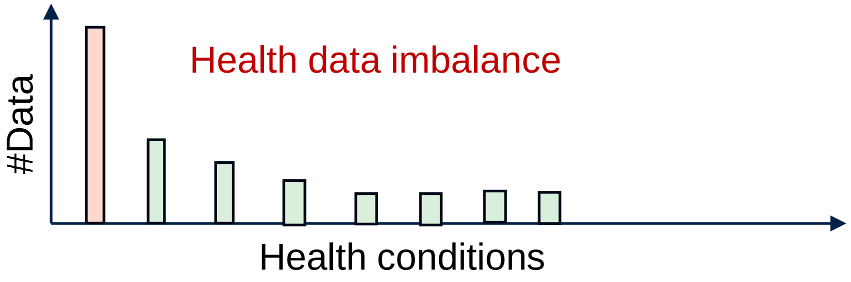
Federated learning (FL):



(a) Model training using centralised data.



(b) Model training using distributed data.



FedAvg:

Local data size

$$\theta^{(t)} = \sum_{k \in \mathcal{K}^{(t)}} \frac{|\mathcal{D}_k|}{\sum_{k \in \mathcal{K}^{(t)}} |\mathcal{D}_k|} \theta_k^{(t)},$$

Feature augmentation based local training

❑ Methodology - FLea

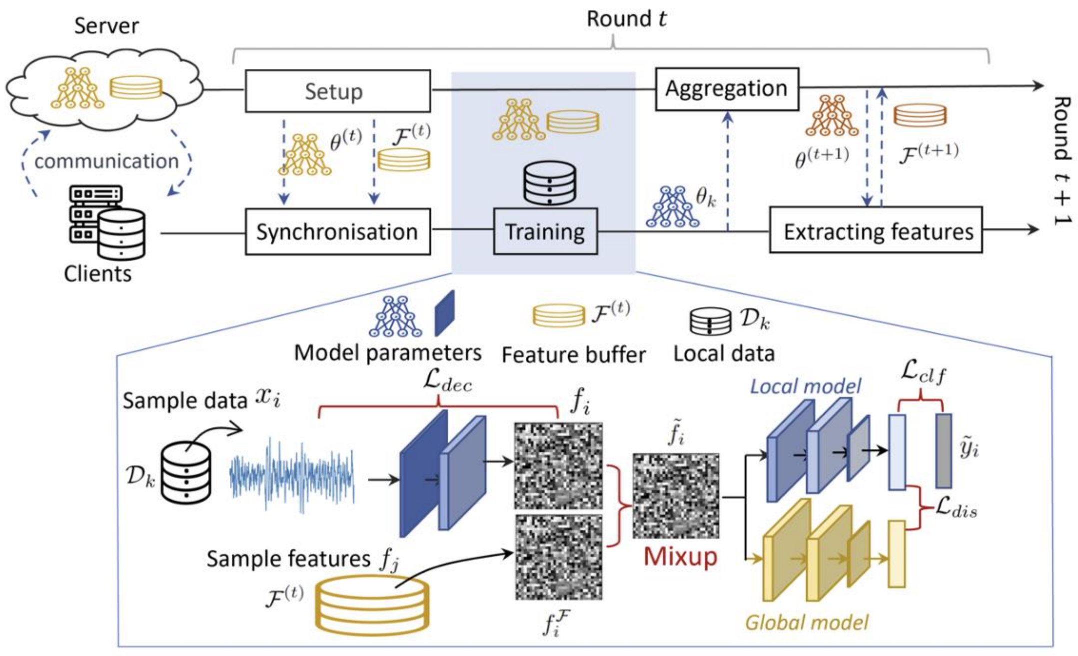


Figure 5: Overview of *FLea* for t-th communication round.

- To address label skew:

$$\mathcal{B}^{\mathcal{F}} = \{(f_i^{\mathcal{F}}, y_i^{\mathcal{F}}) \in \mathcal{F}^{(t)}\}$$

- To address local overfitting:

$$\begin{cases} \tilde{f}_i = \beta_i f_i + (1-\beta_i) f_i^{\mathcal{F}}, \\ \tilde{y}_i = \beta_i y_i + (1-\beta_i) y_i^{\mathcal{F}}, \end{cases}$$

- To protect the privacy of the shared features:

$$\mathcal{L}_{dec}(\mathcal{B}) = \frac{v^2(x,f)}{\sqrt{v^2(x,f)v^2(f,f)}},$$

$$\mathcal{L} = \mathcal{L}_{clf}(\mathcal{B},\mathcal{B}^{\mathcal{F}}) + \lambda_1 \mathcal{L}_{dis}(\mathcal{B},\mathcal{B}^{\mathcal{F}}) + \lambda_2 \mathcal{L}_{dec}(\mathcal{B}),$$

[14] **T. Xia**, A. Ghosh, X. Qiu, and C. Mascolo. FLea: Addressing Data Scarcity and Label Skew in Federated Learning via Privacy-preserving Feature Augmentation. **KDD** 2024

Large-scale unlabeled data pretraining

Table 4: AUROC on health condition inference tasks (higher is better). The best model for each task is highlighted. We report mean and standard deviation from five independent runs. ✓ and * indicates superiority over the opensmile feature set and the other pretrained baselines respectively.

ID	Task Abbr.	Opensmile	VGGish	AudioMAE	CLAP	OPERA-CT	OPERA-CE	OPERA-GT	
T1	Covid (Exhale)	0.550 ± 0.015	0.580 ± 0.001	0.549 ± 0.001	0.565 ± 0.001	0.586 ± 0.008	0.551 ± 0.010	0.605 ± 0.001	✓*
T2	Covid (Cough)	0.649 ± 0.006	0.557 ± 0.005	0.616 ± 0.001	0.648 ± 0.003	0.701 ± 0.002	0.629 ± 0.006	0.677 ± 0.001	✓*
T3	Symptom (Breath)	0.571 ± 0.006	0.571 ± 0.003	0.583 ± 0.003	0.611 ± 0.006	0.603 ± 0.005	0.610 ± 0.004	0.613 ± 0.002	✓*
T4	Symptom (Cough)	0.633 ± 0.012	0.605 ± 0.004	0.659 ± 0.001	0.669 ± 0.002	0.680 ± 0.006	0.665 ± 0.001	0.673 ± 0.001	✓*
T5	Covid (Cough)	0.537 ± 0.011	0.538 ± 0.028	0.554 ± 0.004	0.599 ± 0.007	0.578 ± 0.001	0.566 ± 0.008	0.552 ± 0.003	✓
T6	Gender (Cough)	0.677 ± 0.005	0.600 ± 0.001	0.628 ± 0.001	0.665 ± 0.001	0.795 ± 0.001	0.721 ± 0.001	0.735 ± 0.000	✓*
T7	COPD (Lung)	0.579 ± 0.043	0.605 ± 0.077	0.886 ± 0.017	0.933 ± 0.005	0.855 ± 0.012	0.872 ± 0.011	0.741 ± 0.011	✓
T8	Smoker (Cough)	0.534 ± 0.060	0.507 ± 0.027	0.549 ± 0.022	0.680 ± 0.009	0.685 ± 0.012	0.674 ± 0.013	0.650 ± 0.005	✓*
T9	Gender (Cough)	0.753 ± 0.008	0.606 ± 0.003	0.724 ± 0.001	0.742 ± 0.001	0.874 ± 0.000	0.801 ± 0.002	0.825 ± 0.001	✓*
T10	Obstructive (Lung)	0.636 ± 0.082	0.605 ± 0.036	0.616 ± 0.041	0.697 ± 0.004	0.722 ± 0.016	0.741 ± 0.014	0.703 ± 0.016	✓*
T11	COPD severity (Lung)	0.494 ± 0.054	0.590 ± 0.034	0.510 ± 0.021	0.636 ± 0.045	0.625 ± 0.038	0.683 ± 0.007	0.606 ± 0.015	✓*
T12	Position (Snoring)	0.772 ± 0.005	0.657 ± 0.002	0.649 ± 0.001	0.702 ± 0.001	0.781 ± 0.000	0.769 ± 0.000	0.742 ± 0.001	✓*

Table 5: MAE on lung function estimation tasks (lower is better). Best model per task is highlighted. We report mean and standard deviation across subjects.

ID	Task Abbr.	Opensmile	VGGish	AudioMAE	CLAP	OPERA-CT	OPERA-CE	OPERA-GT	
T13	FVC (Breath)	0.985 ± 0.743	0.904 ± 0.568	0.900 ± 0.551	0.896 ± 0.542	0.924 ± 0.583	0.848 ± 0.607	0.892 ± 0.618	✓*
T14	FEV1 (Breath)	0.756 ± 0.721	0.839 ± 0.563	0.821 ± 0.590	0.840 ± 0.547	0.837 ± 0.563	0.834 ± 0.581	0.825 ± 0.560	
T15	FEV1/FVC (Breath)	0.141 ± 0.185	0.131 ± 0.146	0.129 ± 0.146	0.134 ± 0.146	0.128 ± 0.140	0.132 ± 0.141	0.128 ± 0.141	✓*
T16	FVC (Vowel)	0.850 ± 0.592	0.895 ± 0.559	0.833 ± 0.588	0.883 ± 0.560	0.885 ± 0.553	0.761 ± 0.544	0.878 ± 0.550	✓*
T17	FEV1 (Vowel)	0.730 ± 0.497	0.842 ± 0.559	0.876 ± 0.561	0.859 ± 0.541	0.780 ± 0.542	0.830 ± 0.561	0.774 ± 0.554	*
T18	FEV1/FVC (Vowel)	0.138 ± 0.166	0.130 ± 0.145	0.131 ± 0.141	0.137 ± 0.147	0.132 ± 0.140	0.136 ± 0.150	0.130 ± 0.138	✓*
T19	Breathing Rate	2.714 ± 0.902	2.605 ± 0.759	2.641 ± 0.813	2.650 ± 0.947	2.636 ± 0.858	2.525 ± 0.782	2.416 ± 0.885	✓*

- ✓ Outperform baselines on 16 out of 19 tasks
- ✓ Generalizable to unseen data and new respiratory audio modalities

OPERA

We make everything open for research: https://github.com/evelyn0414/OPERA.git

[12] E. Zhang^, **T. Xia^**, et al. Towards Open Respiratory Acoustic Foundation Models: Pretraining and Benchmarking. https://arxiv.org/abs/2406.16148