AI-empowered mHealth:

Pioneering Applications and Overcoming Key Challenges

Tong Xia

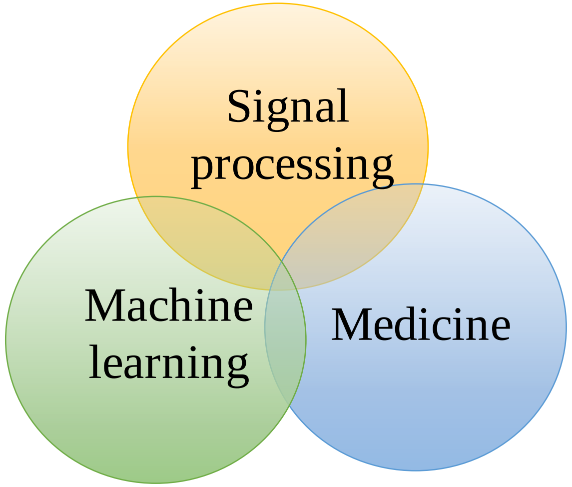
June, 2024

About me

- 2013-2017 ***Bachelor's***
  Wuhan University
  Electronic Information

- 2017-2020 ***Master's***
  Tsinghua University
  Electronics and Communication

- 2020-2024 ***Ph.D.*** (*Huawei studentship*)
  University of Cambridge
  Computer Science
- Now ***Postdoctoral researcher associate***



Mobile computing
Machine learning
Application to health

University of Cambridge

Mobile Systems Research Lab

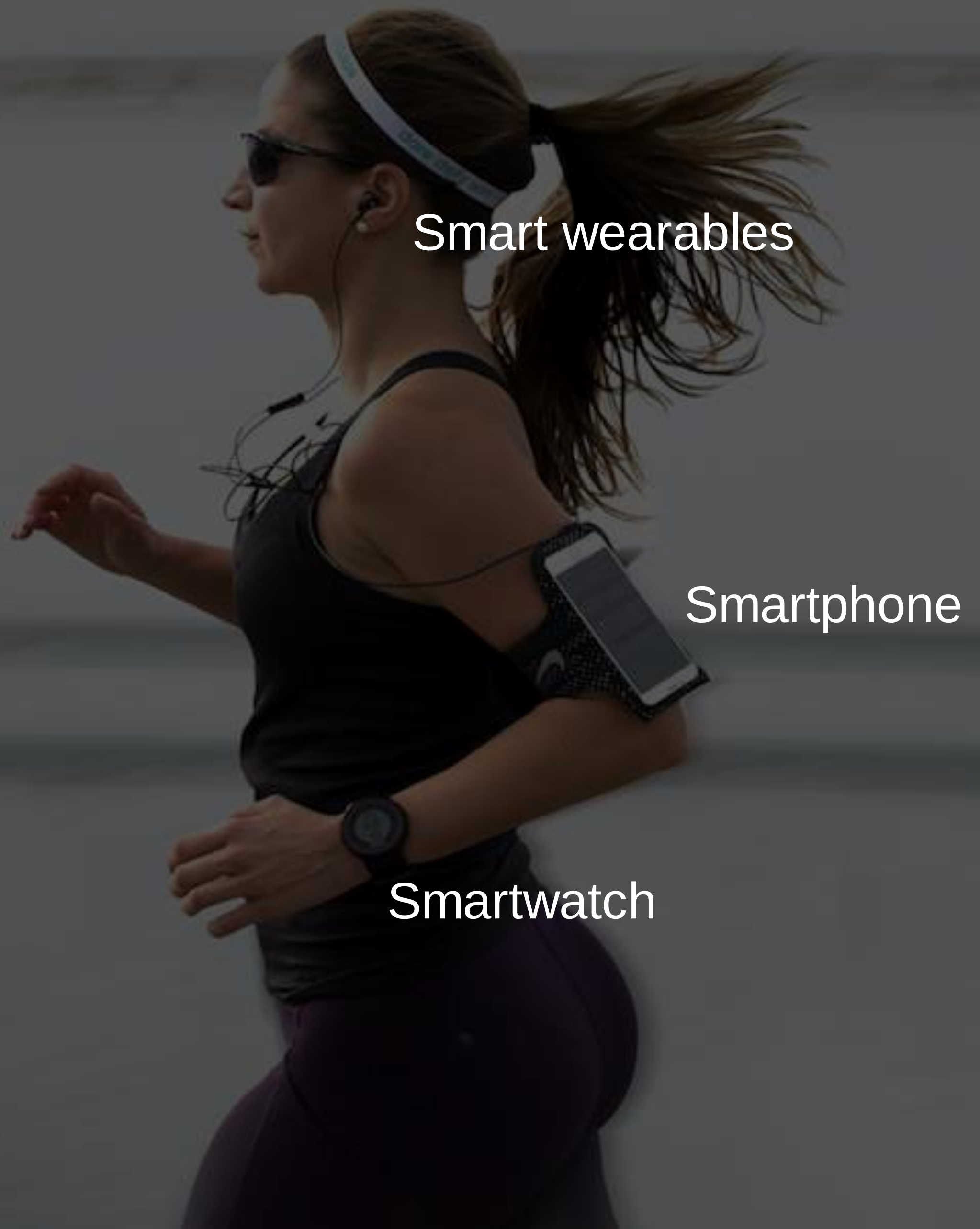
Prof. Cecilia Mascolo

AI-empowered mHealth:

Pioneering Applications and Overcoming Key Challenges

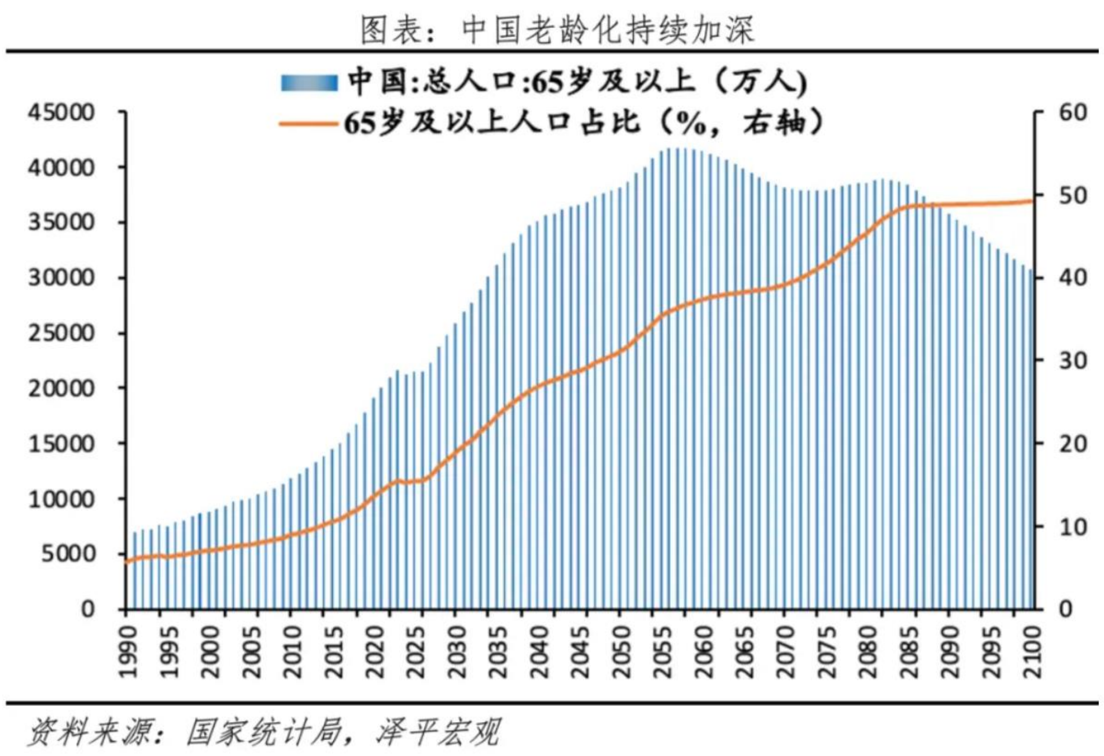
mHealth

- What?
- Why?



- Shortage of medical resources globally

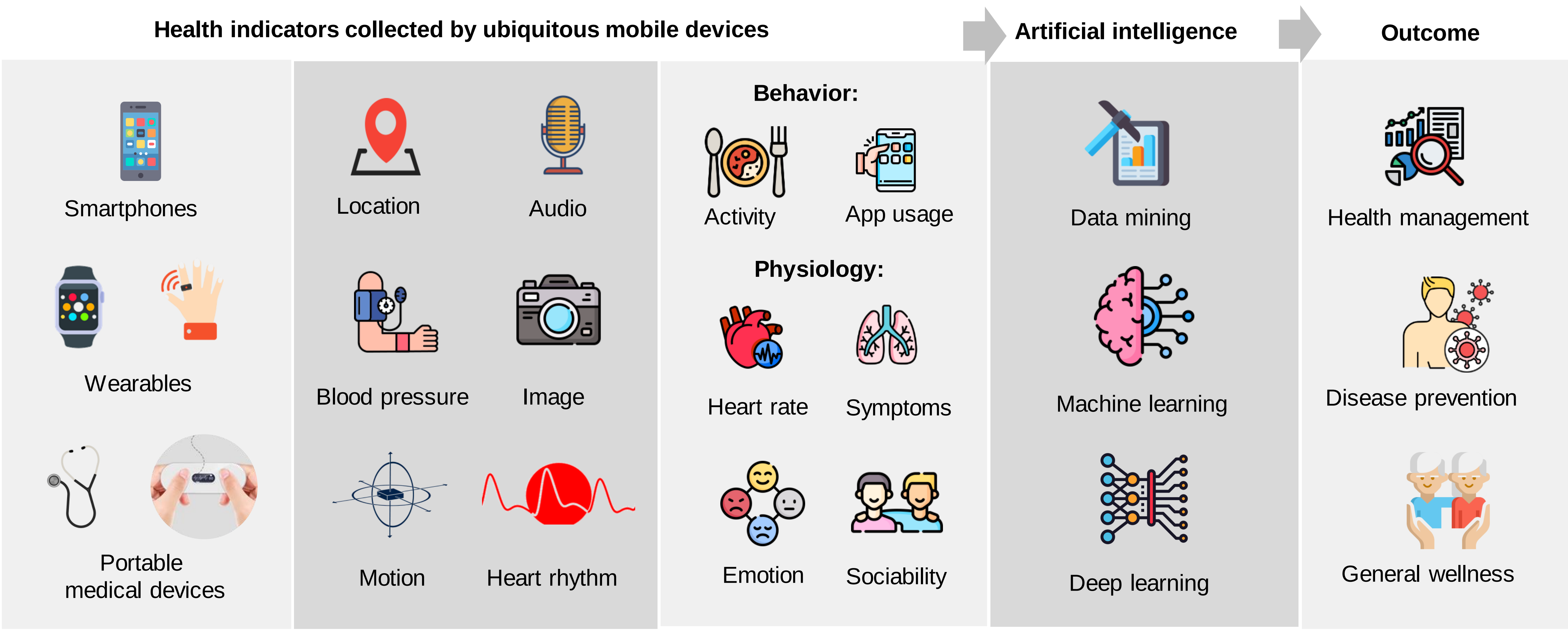
- About **47%** of the global population lacks access to adequate diagnostic services
- China is facing the problem of **aging population**



99.9% of people in China own a smartphone with internet access. The smartwatch market is rising rapidly [2]

- Proliferation of mobile health devices

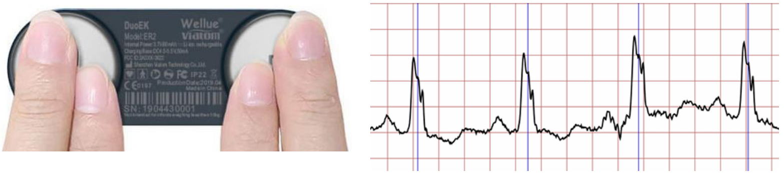
[1] 中国老龄化报告2024_腾讯新闻 (qq.com)
[2] CNNIC：第53次中国互联网络发展状况统计报告 | 互联网数据资讯网-199IT | 中文互联网数据研究资讯中心-199IT



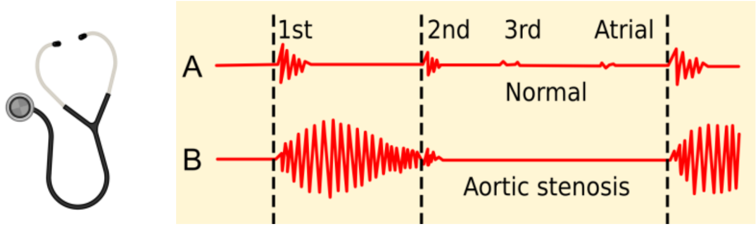
AI-empowered mHealth:

Pioneering Applications and Overcoming Key Challenges

Applications

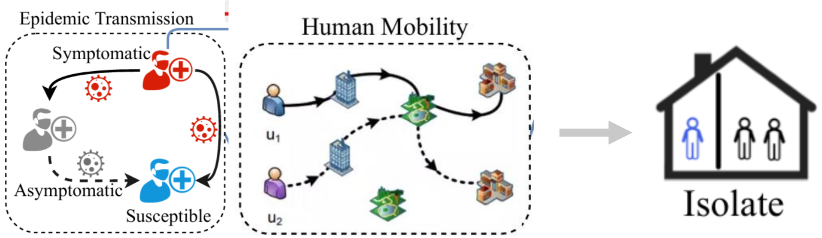


ECG(electrocardiogram)-based heart arrythmia detection
(WHI’22, IEEE JBHI’24)



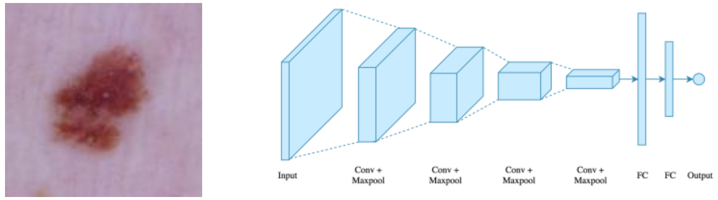
Predicting murmur from heart sound
(CinC’22)

Predicting hospital visits from individual mobility
(AAAI’20, UbiComp’21)



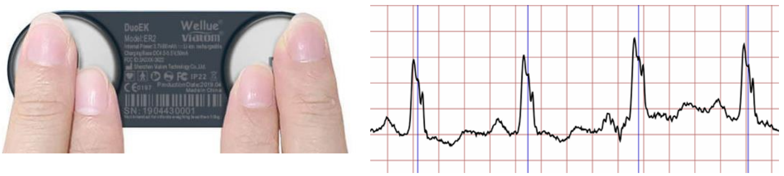
Precise mobility intervention for epidemic control
(BigData’21, KDD’22)

Audio-driven respiratory health screening
(KDD’20,NPJ DM’21,NeurIPS’21, KDD’22)

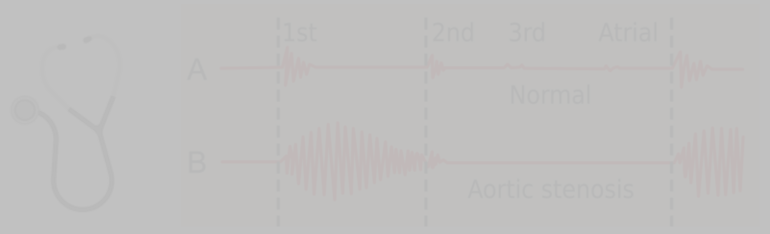


Dermoscopic image-based skin lesion prediction
(KDD FL4Data’23,IEEE JBHI’24)

Applications

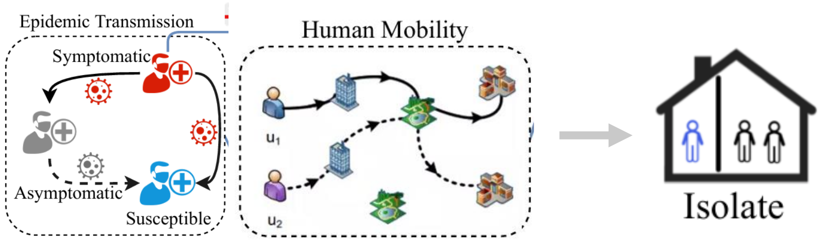


ECG(electrocardiogram)-based heart arrythmia detection
(WHI’22, IEEE J-BHI’24)



Predicting murmur from heart sound
(CinC’22)

Predicting hospital visits from individual mobility
(AAAI’20, UbiComp’21)



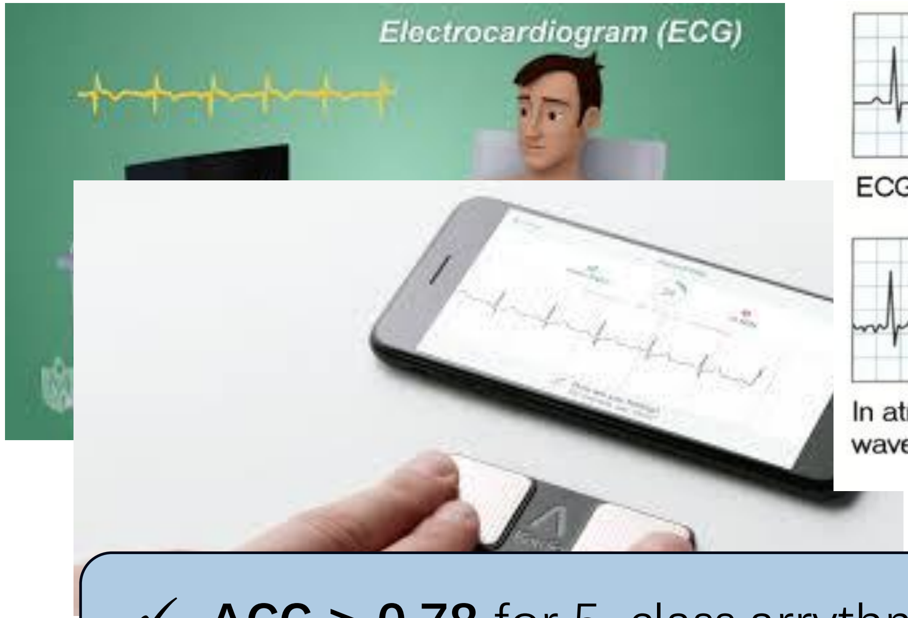
Precise mobility intervention for epidemic control
(BigData’21, KDD’22)

Audio-driven respiratory health screening
(KDD’20,NPJ DM’21,NeurIPS’21, KDD’22)

Dermoscopic image-based skin lesion prediction
(KDD FL4Data’23,IEEE J-BHI’24

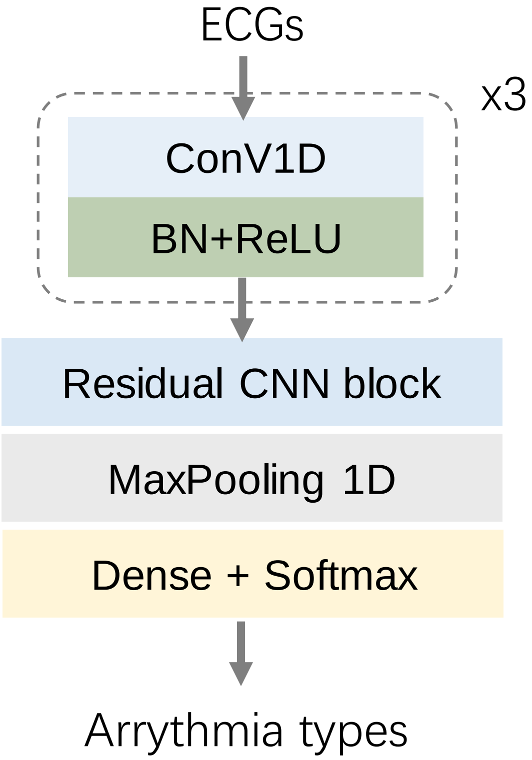
Cardiovascular mHealth

- **ECG (electrocardiogram)-based heart arrythmia detection**



ECG tracing of a normal heart rhythm.

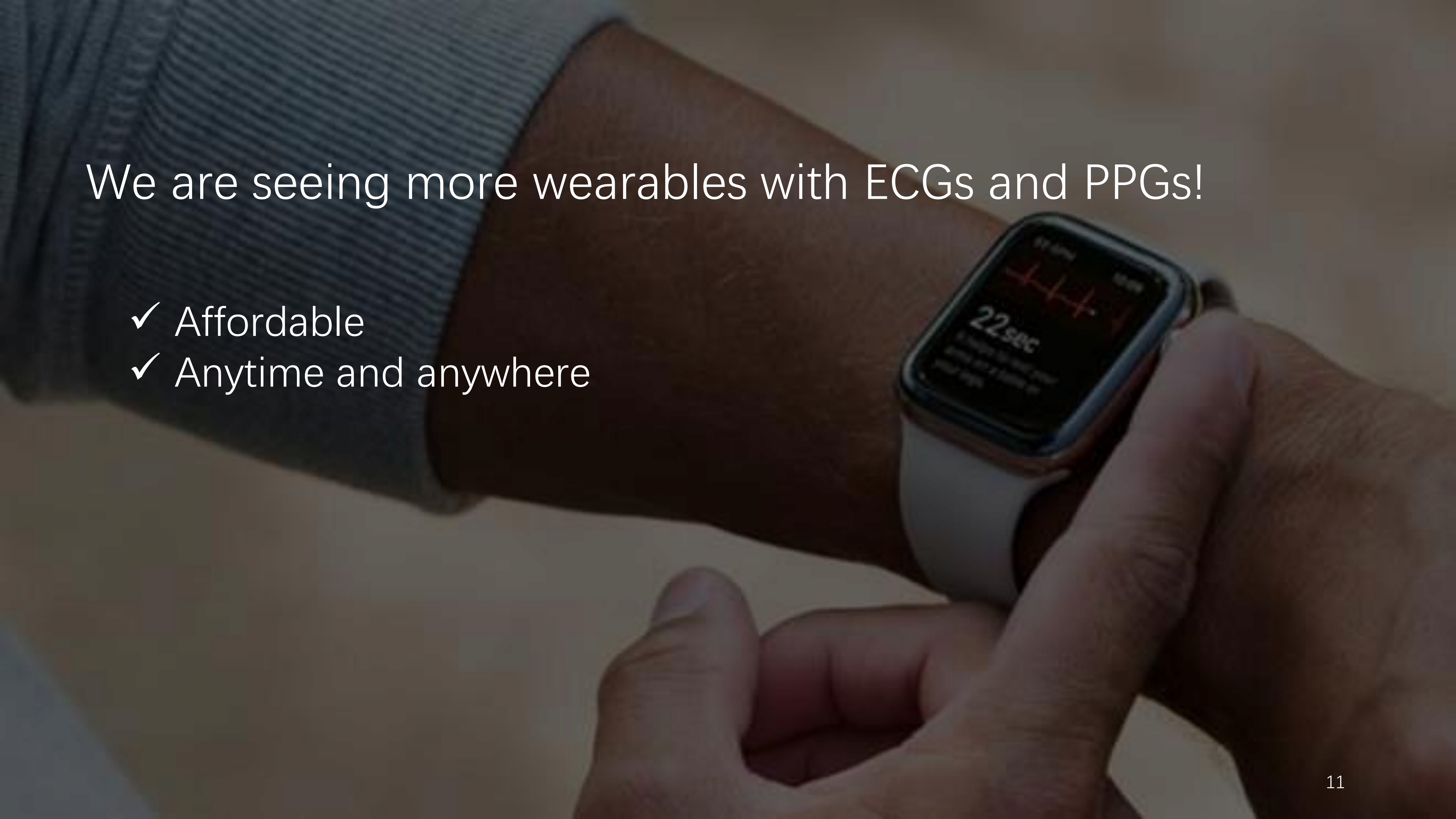
In atrial fibrillation, the tracing shows tiny, irregular "fibrillation" waves between heartbeats. The rhythm is irregular and erratic.



- ✓ **ACC > 0.78** for 5-class arrythmia (心率不齐) classification
- ✓ **Sensitivity > 0.88** for AF (房颤) detection

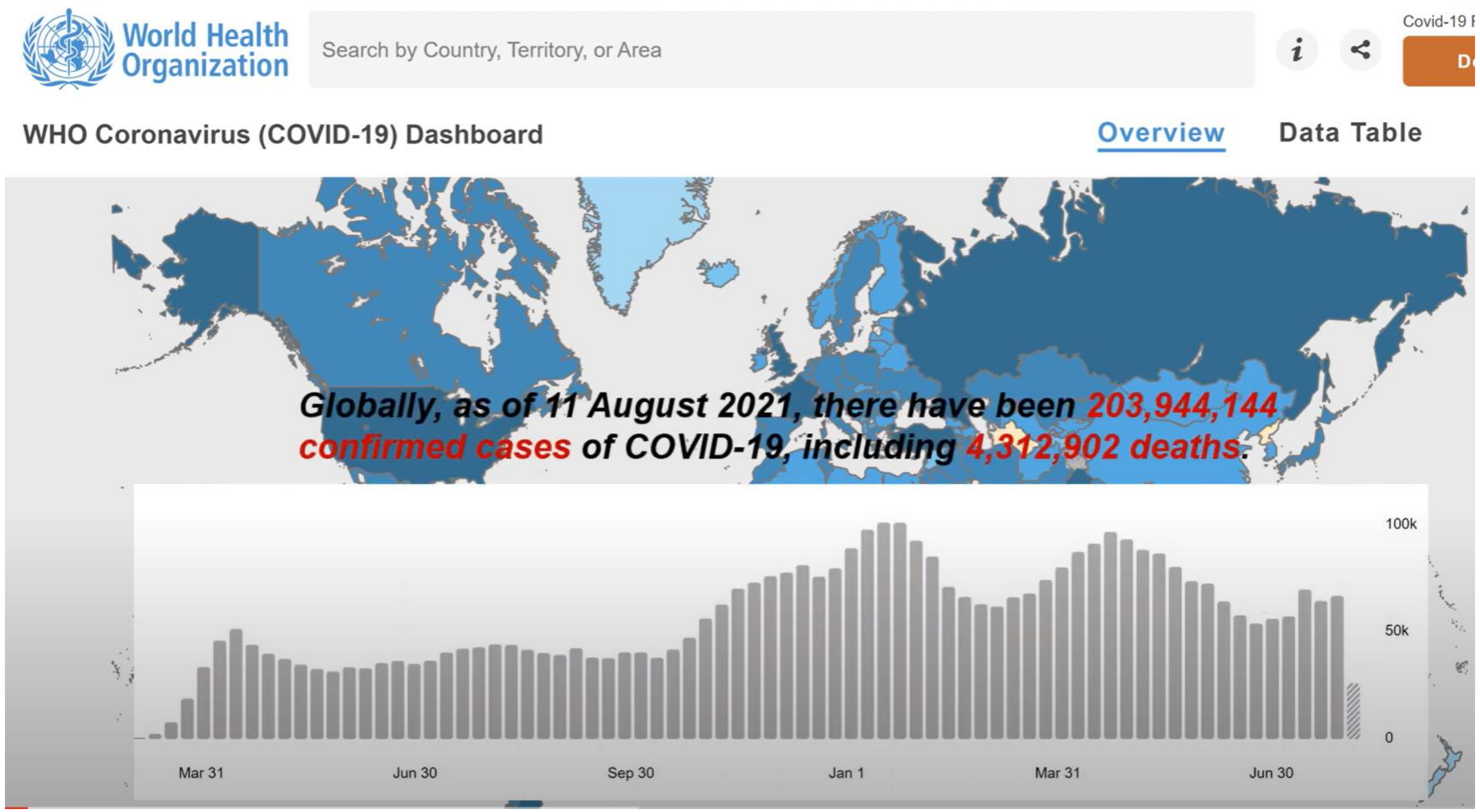
[3] **T. Xia.** Reliable and decentralised deep learning for physiological data. PhD Thesis 2024.
[4] **T. Xia**, J. Han, C. Mascolo. Benchmarking Uncertainty Quantification on Biosignal Classification Tasks under Dataset Shift. Workshop on Health Intelligence, AAAI 2022
[5] **T. Xia**, T. Dang, J. Han, L. Qendro, and C. Mascolo. Class-balanced Evidential Deep Learning for Health Diagnostics. IEEE **JBHI** 2024

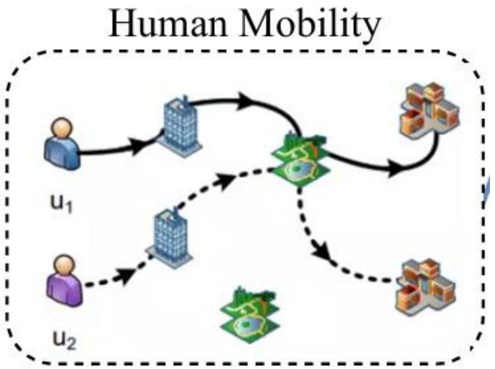
We are seeing more wearables with ECGs and PPGs!

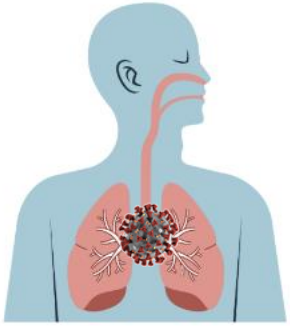


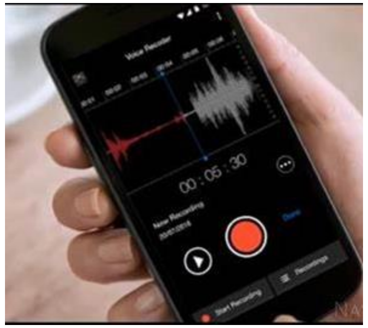
- ✓ Affordable
- ✓ Anytime and anywhere

mHealth in containing COVID-19

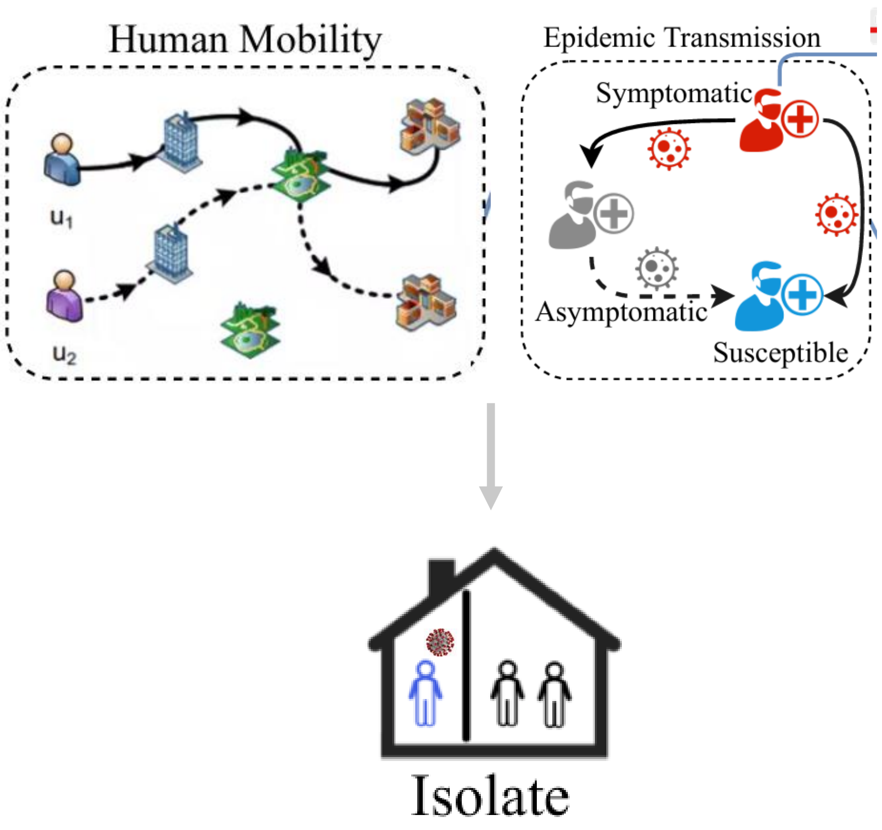


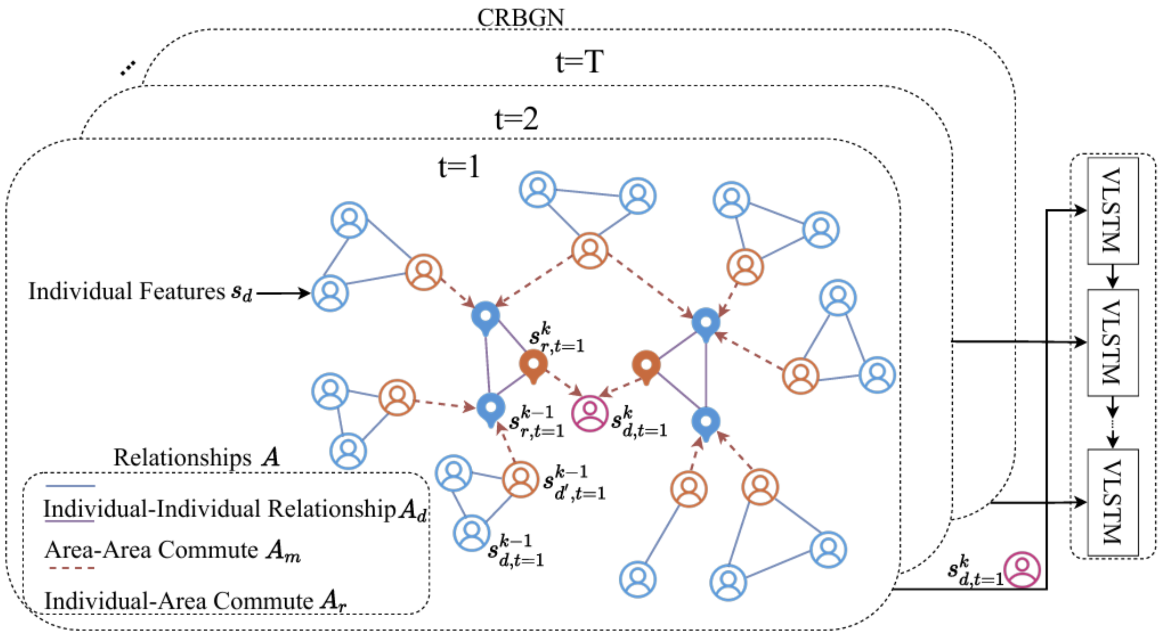






Mobility-based mHealth in containing COVID-19



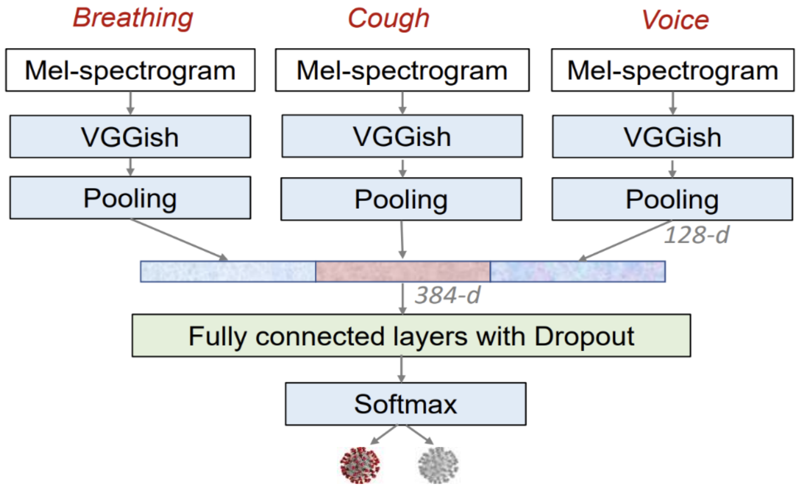


	Quarantine %	Exposure %
No intervention	100	100
Lockdown	100	<5%
Expert policy	20%	<20%
Ours	<5%	<1%

[6] **T. Xia^** and A. Ghosh^. Mobility-based Individual POI Recommendation to Control the COVID-19 Spread. IEEE **Big Data** 2021.
[7]T. Feng, **T. Xia**, et al. Precise Mobility Intervention for Epidemic Control Using Unobservable Information via Deep Reinforcement Learning. **KDD** 2022

Audio-based mHealth in containing COVID-19

Can our voices recorded by smartphones be used for respiratory infections detection?



covid-19-sounds.org
30K+ participants

\>>> **Audio:** AUROC of 0.71, Sensitivity of 0.65, Specificity of 0.69
Flow test: Sensitivity ranges from 0.37 to 0.99

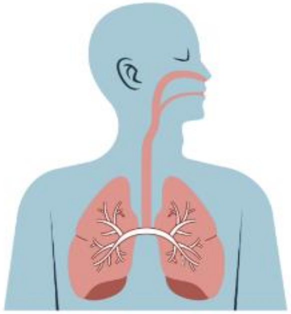
[8] J. Han^, **T. Xia**^, et al. Sounds of COVID-19: Exploring Realistic Performance of Audio-based Digital Testing. **Nature NPJ Digital Medicine** 2022
[9] **T. Xia**^, D. Spathis^, C. Mascolo, et al. COVID-19 Sounds: A Large-Scale Audio Dataset for Digital Respiratory Screening. **NeurIPS** Datasets and Benchmarks Track 2021

The promise of audio-based mHealth

- ✓ Scalable
- ✓ Non-invasive
- ✓ Sustainable

Audio-based health screening

Respiratory audio-driven mHealth application



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

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

- ✓ Crackle prediction
- ✓ Wheeze prediction
- ✓ Infection localization

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

[10] **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
[11] J. Han, **T. Xia**, C. Mascolo. Audio-based Sleep Apnea Detection from Tracheal and Ambient Sound Recordings. Under review.
[12] E. Zhang^, **T. Xia^**, et al. Towards Open Respiratory Acoustic Foundation Models: Pretraining and Benchmarking. https://arxiv.org/abs/2406.16148

mHealth Applications

Are we ready yet?

Key Challenges

Biases and fairness

Explainability and uncertainty

On-device efficiency

Data privacy

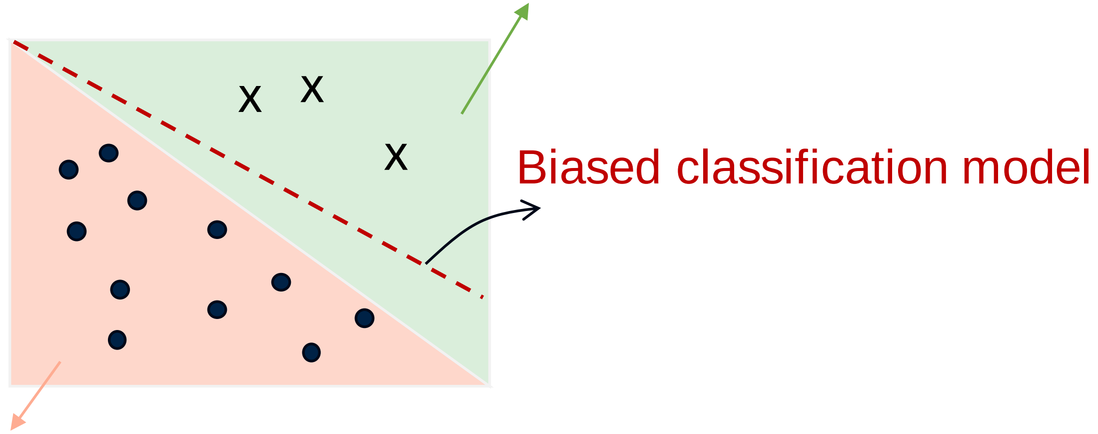
Generalizability

…

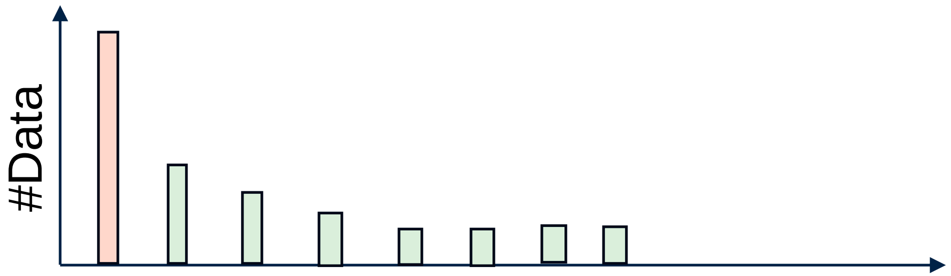
How to enable reliable mHealth in the wild?

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

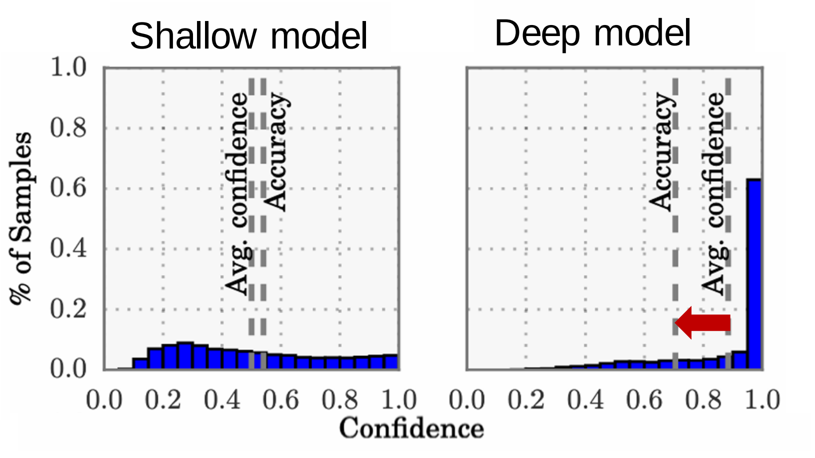
Real distribution for unhealthy samples (minority class)



Real distribution for healthy samples (majority distribution)



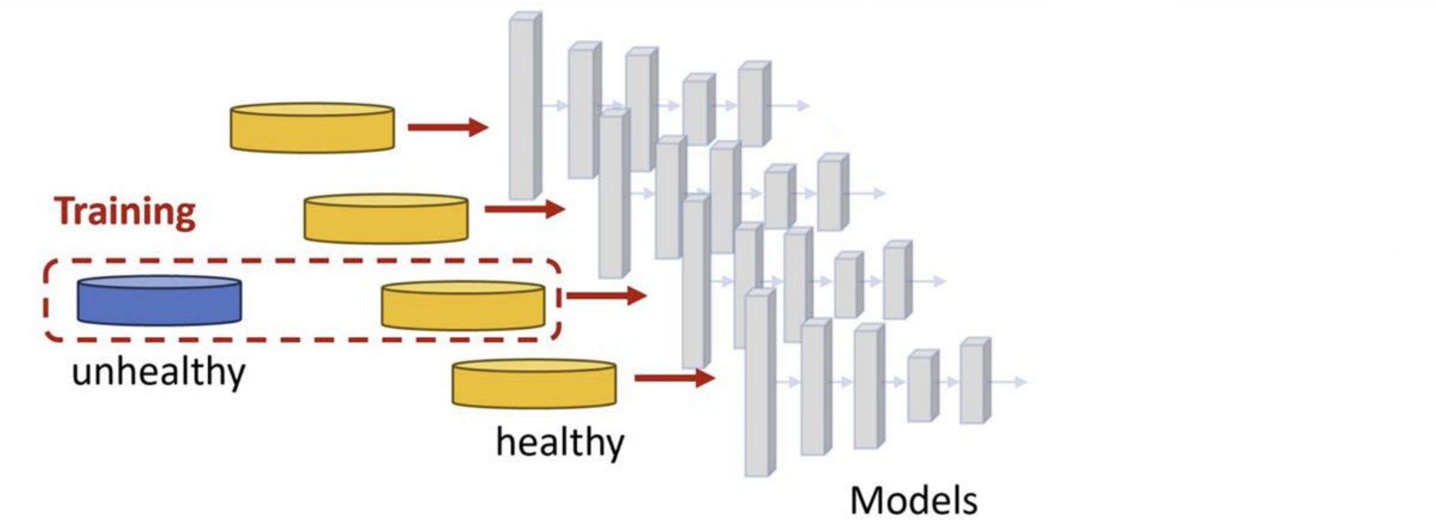
Deep learning overconfidence



Sensitivity to distributional shift

Solution 1: Data-balanced ensemble learning for uncertainty

❑ Methodology



- Model fusion:

$$\boldsymbol{p}^{(i)} = \frac{1}{N_m} \sum_{n=1}^{N_m} \boldsymbol{p}_n^{(i)},$$

- Model uncertainty estimation:

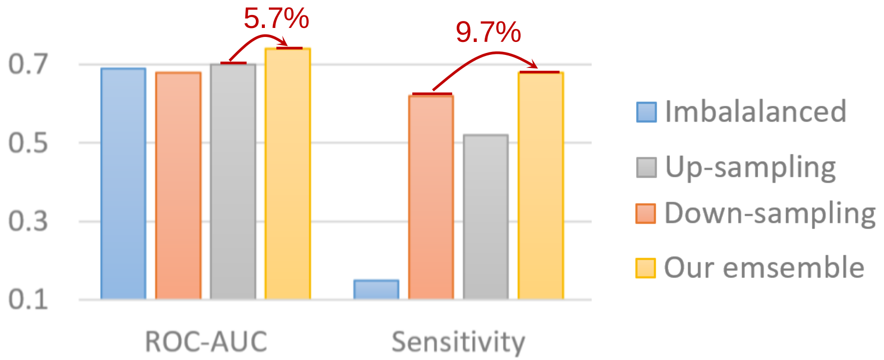
$$\mu^{(i)} = \frac{1}{N_m} \sum_{n=1}^{N_m} (\boldsymbol{p}_n^{(i)}[1]),$$

$$\sigma^{(i)} = \sqrt{\frac{1}{N_m} \sum_{n=1}^{N_m} (\boldsymbol{p}_n^{(i)}[1] - \mu^{(i)})^2}.$$

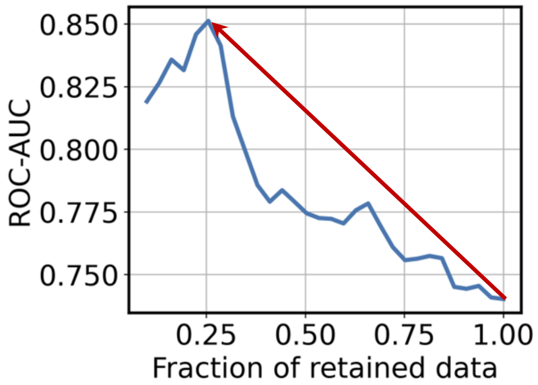
[12] **T. Xia**, J. Han, L. Qendro, T. Dang, and C. Mascolo. Uncertainty-aware COVID-19 Detection from Imbalanced Sound Data. **INTERSPEECH** 2021

Solution 1: Data-balanced ensemble learning for uncertainty

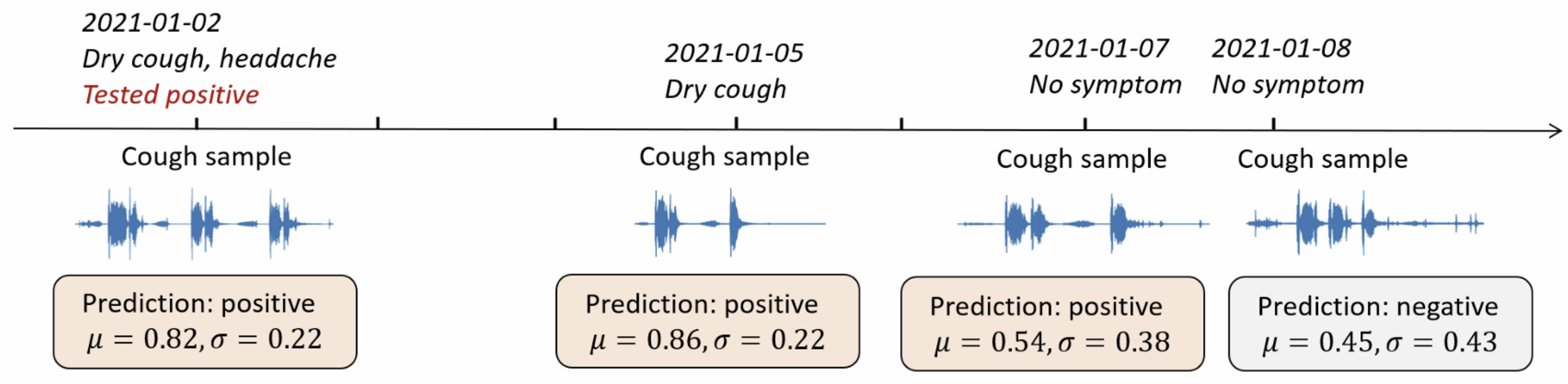
❑ Results on audio-based COVID-19 detection



Less biased model



Detection with referral

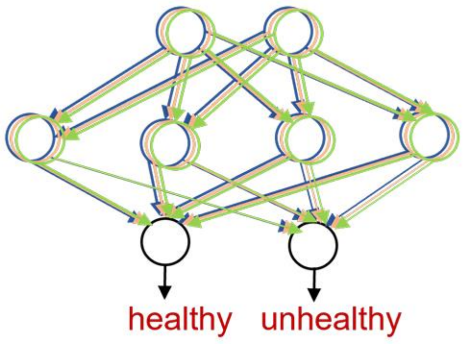


A case study of uncertainty estimates

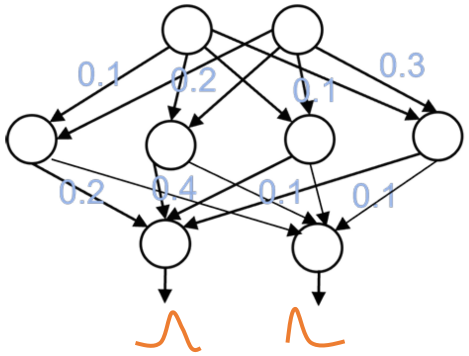
Solution 2: Class-balanced evidential deep learning for uncertainty

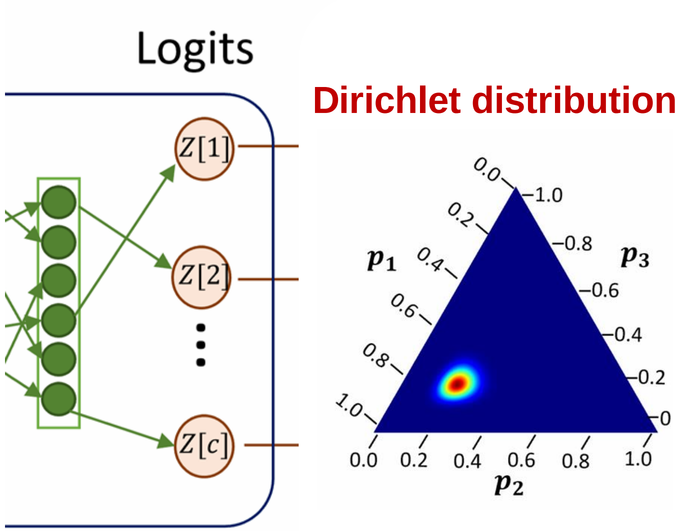
- **Challenge: on-device efficiency**

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:

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

$$\Downarrow$$

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

- Dataset level:

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

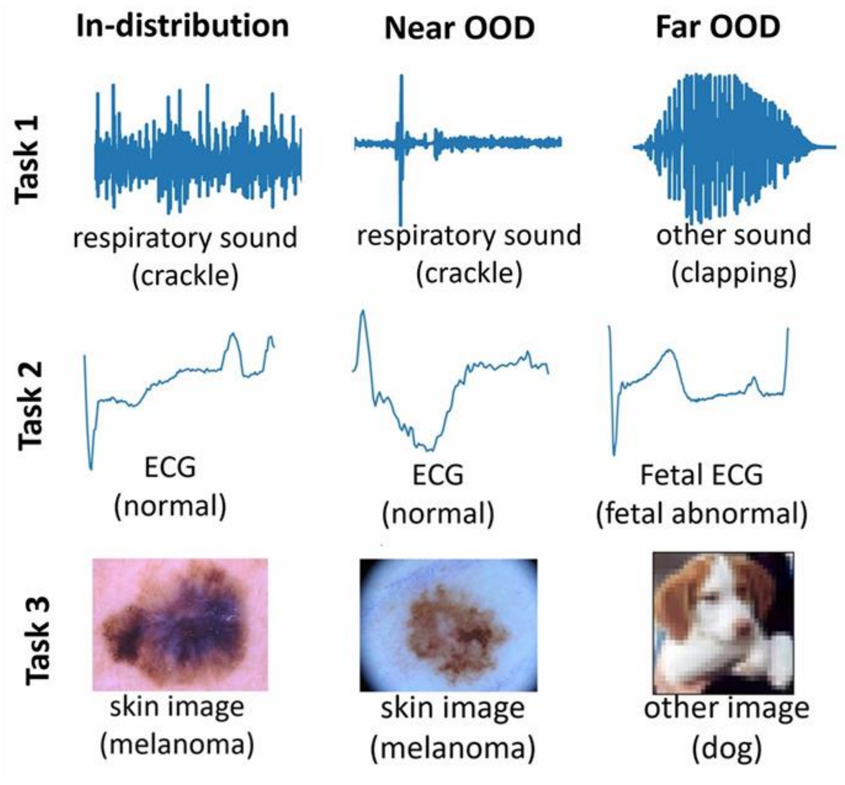
$$\Downarrow$$

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

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

Solution 2: Class-balanced evidential deep learning for uncertainty

- Experiments on three mHealth applications



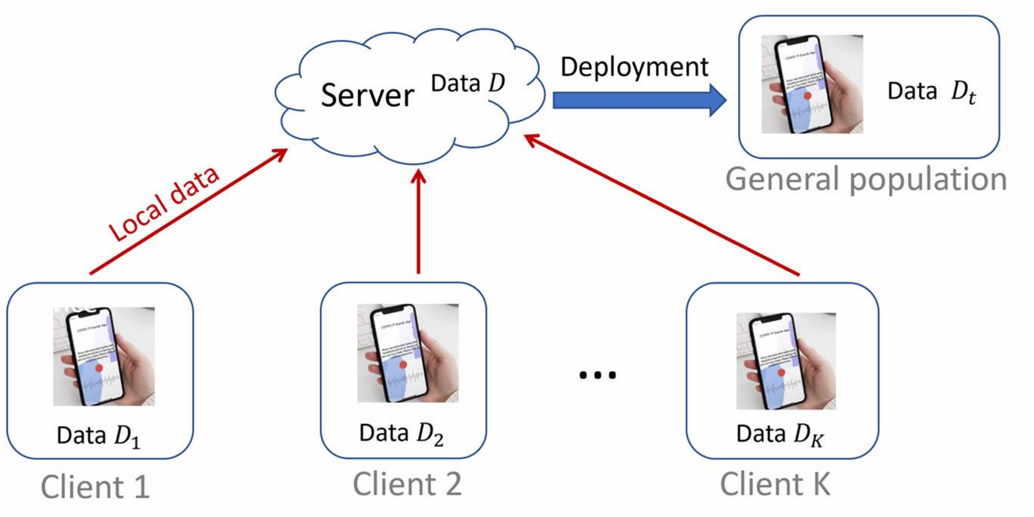
- Conclusions:
  - Competitive accuracy *vs.* ensemble method
  - Reduce overconfident predictions by up to 43%
  - Improve OOD detection by up to 16.1%
  - Require almost no additional memory and computation for uncertainty estimations

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

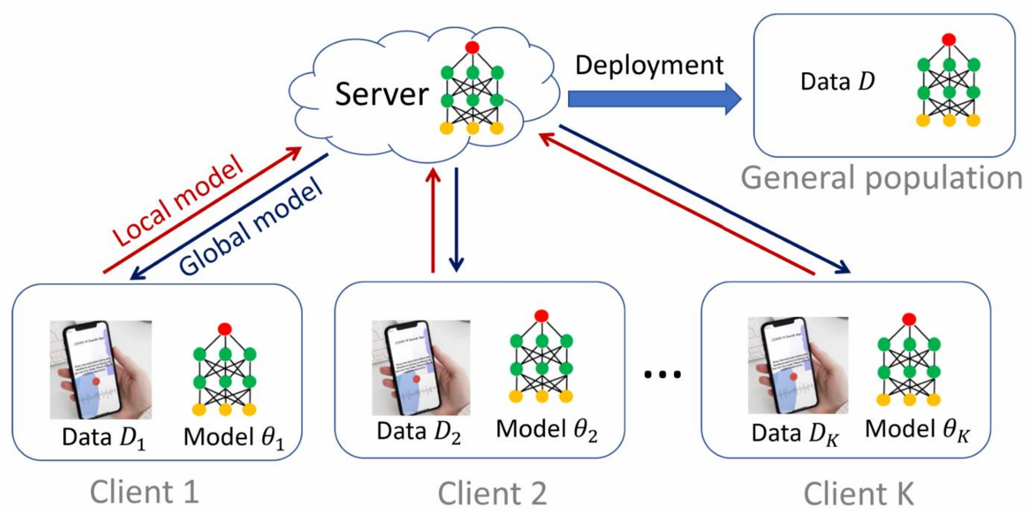
How to ensure trustworthy AI for mHealth?

- Challenge: Data privacy

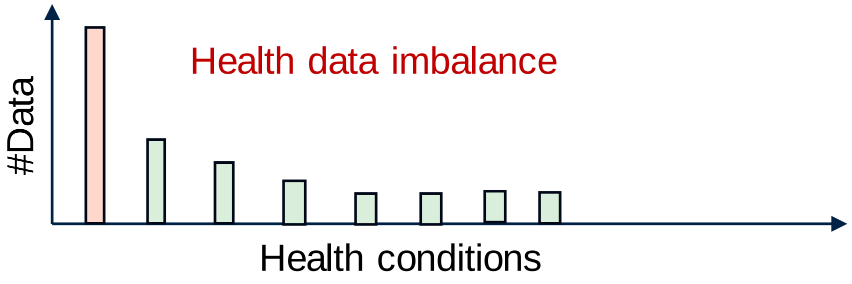
Federated learning (FL):



(a) Model training using centralised data.



(b) Model training using distributed data.

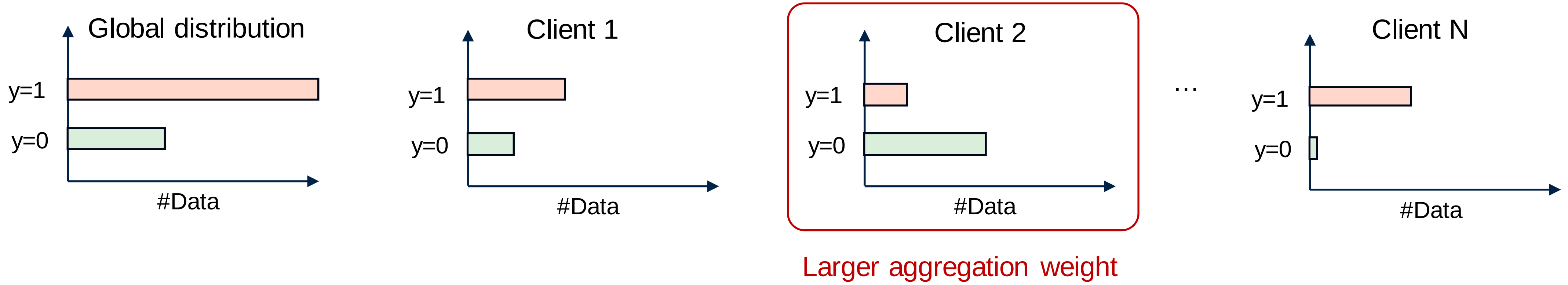


FedAvg:

Local data size

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

Solution 1: Weighted federated aggregation



FedLoss (Proposed):

$$w_1^{(t)}, ..., w_M^{(t)} = Softmax(l_1^{(t)}, ..., l_M^{(t)}),$$

$$\theta^{(t)} = \sum_{k=1}^{M} w_k^{(t)} \theta_k^{(t)},$$

Predictive loss of the global model on the local data size

[13] **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

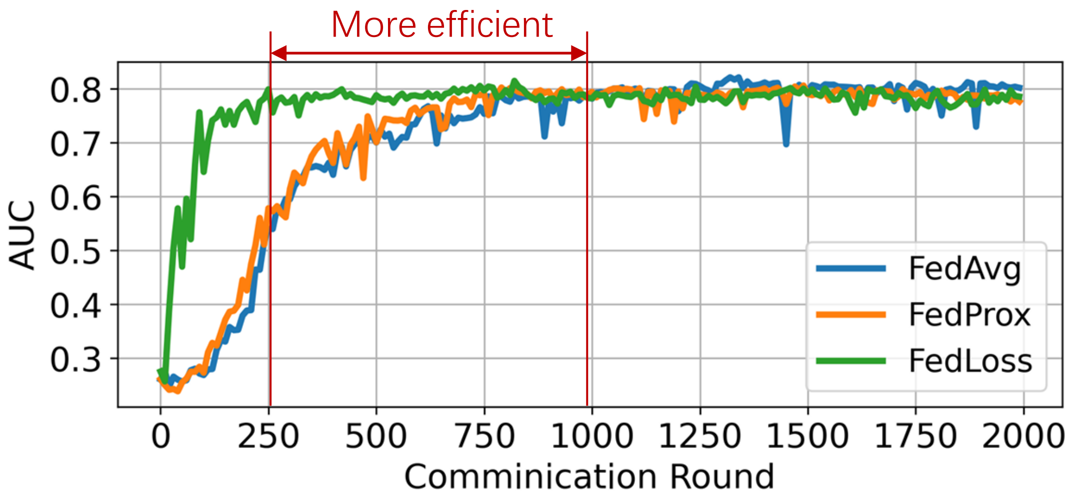
Solution 1: Weighted federated aggregation

❑ Experimental setup

- Data set: COVID-19 physiological audio
- Clients: 2,368 participants (majority always tested COVID-19 negative)
- Training: 30 clients update local models and contrite to the global model per round. Totally 2000 rounds

❑ Results

	ROC-AUC	Sensitivity	Specificity	Youden's index
Centralised	0.79 (0.74-0.84)	0.46 (0.36-0.56)	0.93 (0.91-0.94)	0.40 (0.29-0.50)
FedAvg	0.80 (0.75-0.85)	0.11 (0.06-0.17)	1.00 (1.00-1.00)	0.11 (0.06-0.16)
FedProx	0.75 (0.69-0.80)	0.19 (0.12-0.27)	0.99 (0.99-1.00)	0.18 (0.12-0.26)
FedLoss (Proposed)	0.79 (0.73-0.83)	0.50 (0.40-0.59)	0.90 (0.88-0.92)	0.40 (0.28-0.50)



[13] **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

Solution 2: Feature augmentation based local training

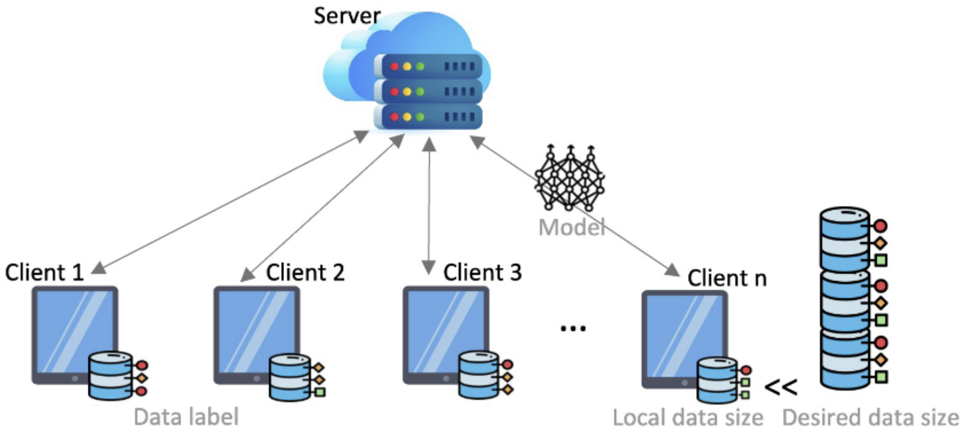


Figure 1: Edge devices as clients in federated learning, where local data exhibits label skew (presented by different markers) and scarcity (usually very small in size).

Cross-device FL for mHealth:

- Multiple classes

  Simple weighted aggregation doesn't work

- Small local data size

  Local model overfitting

- Label skew/class imbalance

  Local model drift → Global model suboptimal

Solution 2: Feature augmentation based local training

❑ Methodology - FLea

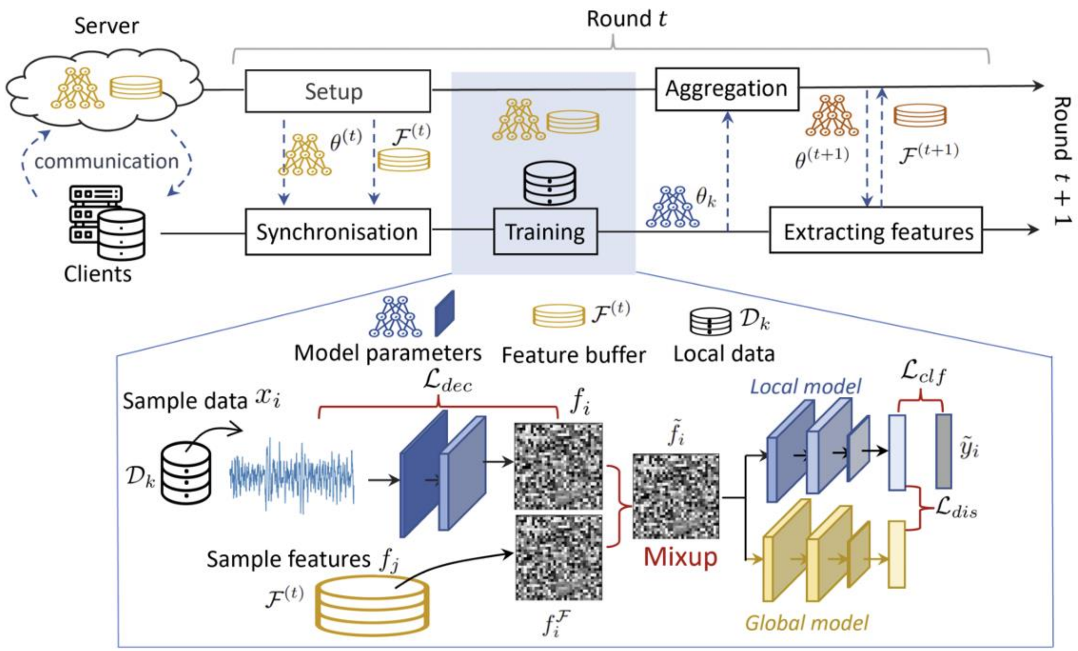


Figure 5: Overview of *FLea* for t-th communication round.

- 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

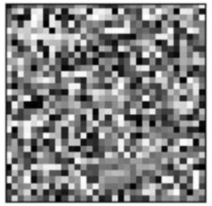
Solution 2: Feature augmentation based local training

❑ Experiments

Table 1: Overall accuracy comparison. Accuracy is reported as *mean* ± *std* across five runs. The best performance under each setting is highlighted in red and the SOTA baseline (*excluding *FedData*) is in grey. ↑ indicates a relative improvement of our method compared to the SOTA over 5% and ↑↑ indicates a relative improvement over 10%.

		CIFAR10			UrbanSound8K			UCI-HAR		
	Accuracy %	*Qua*(3)	*Dir*(0.5)	*Dir*(0.1)	*Qua*(3)	*Dir*(0.5)	*Dir*(0.1)	*Qua*(2)	*Dir*(0.3)	*Dir*(0.1)
$\lvert\tilde{\mathcal{D}}_k\rvert = 100$	FedAvg	30.25±1.33	32.58±1.09	20.46±2.15	43.69±0.56	46.77±0.87	34.59±2.64	66.99±0.87	65.78±0.34	48.43±0.70
	FedProx	31.92±1.45	32.01±1.25	20.86±1.97	38.45±0.48	39.58±1.02	34.81±0.46	68.32±0.50	67.75±0.41	58.35±0.52
	FedDecorr	31.12±1.57	33.57±1.22	21.34±1.59	45.01±0.57	46.77±0.65	35.87±1.03	69.12±0.63	66.68±0.43	57.05±0.38
	FedLC	32.05±1.60	30.17±1.18	18.82±2.01	50.98±0.49	50.11±0.83	37.05±0.87	71.69±0.52	70.57±0.38	62.57±0.42
	FedNTD	39.98±0.97	39.82±0.86	26.78±2.34	49.80±0.45	51.09±0.97	36.53±0.99	68.33±0.72	70.32±0.49	60.13±0.51
	FedBR	31.66±1.07	33.08±1.12	20.98±2.54	44.05±0.63	47.58±0.90	36.15±1.17	67.54±0.68	69.15±0.40	59.87±0.46
	CCVR	35.95±1.63	35.02±1.43	24.21±2.67	47.12±0.72	49.26±0.92	39.62±1.20	70.17±0.49	68.87±0.51	60.28±0.36
	FedGen	32.32±1.21	34.27±1.56	22.56±2.89	45.20±0.89	48.33±1.12	38.27±1.44	70.58±0.61	69.32±0.60	60.07±0.63
	FedMix	44.04±1.53	45.50±1.88	38.13±2.06	51.56±0.59	54.18±0.62	43.35±0.72	68.59±0.54	69.34±0.49	65.63±0.47
	FedData*	54.64±1.02	56.47±1.22	55.35±1.46	62.83±1.25	64.45±0.76	61.11±0.98	78.13±0.46	78.24±0.51	75.93±0.34
	FLea	47.03±1.01↑	48.86±1.43↑	44.40±1.23↑↑	57.73±0.51↑↑	59.22±0.78↑	45.94±0.77↑	75.17±0.42	73.02±0.49	71.68±0.51↑
$\lvert\tilde{\mathcal{D}}_k\rvert = 50$	FedAvg	27.72±1.26	26.92±1.31	21.88±1.87	39.35±0.60	43.98±0.89	31.21±1.62	65.77±0.42	67.10±0.40	46.95±0.62
	FedProx	22.88±2.54	24.47±2.17	21.01±2.46	39.05±0.56	42.21±0.76	32.85±1.22	69.18±0.41	68.28±0.45	59.97±0.46
	FedDecorr	26.45±1.58	25.57±1.84	22.03±1.98	39.67±0.58	44.23±0.95	33.67±1.34	65.77±0.39	68.57±0.51	55.54±0.49
	FedLC	28.64±1.52	26.36±1.47	20.24±1.68	44.33±0.79	45.15±0.80	39.87±1.04	70.63±0.49	71.34±0.45	63.67±0.52
	FedNTD	32.92±1.43	34.64±1.52	30.13±1.67	42.21±0.63	48.63±0.78	40.15±1.22	65.64±0.38	67.16±0.43	59.93±0.46
	FedBR	30.25±1.45	30.32±1.32	28.52±1.56	41.15±0.70	44.37±0.82	34.89±1.36	66.98±0.43	68.23±0.49	57.25±0.52
	CCVR	34.01±1.89	35.12±1.34	33.26±1.56	44.05±0.87	46.68±0.83	36.80±1.37	65.24±0.50	70.15±0.46	60.26±0.57
	FedGen	33.12±1.61	31.89±1.59	29.90±1.76	40.89±0.72	44.54±0.81	35.78±1.40	68.27±0.64	69.82±0.41	59.13±0.45
	FedMix	38.14±1.12	39.87±1.55	36.87±1.38	46.55±0.81	50.00±0.92	42.27±1.15	68.06±0.44	70.80±0.45	61.39±0.46
	FedData*	53.59±1.32	53.02±1.18	53.56±1.64	60.31±0.82	60.48±0.91	59.67±1.55	76.42±0.38	76.45±0.47	75.46±0.47
	FLea	41.98±1.26↑↑	42.01±1.13↑	37.69±1.65	54.35±0.80↑↑	55.68±0.87↑↑	45.05±1.32↑	74.25±0.44↑	73.98±0.46	66.57 ± 0.45

Original image



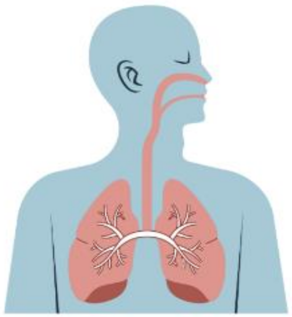
Activation

Reconstruction attack

[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

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

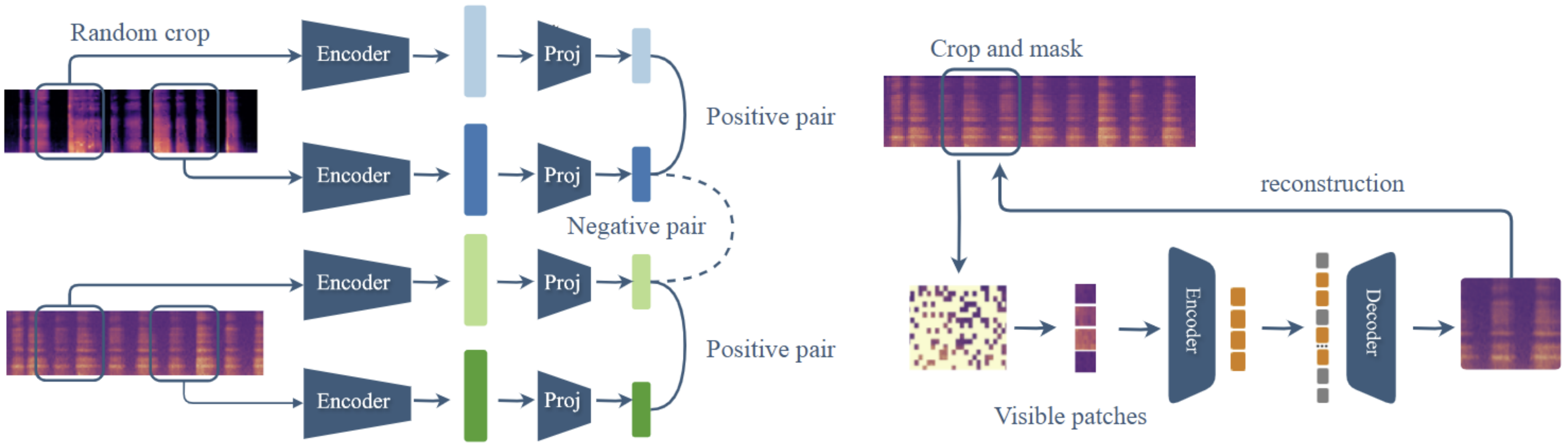
Task specific model v.s. One-for-all ?

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

Solution: Large-scale unlabeled data pretraining

- 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



(a) Contrastive (OPERA-CT, OPERA-CE) (b) Generative (OPERA-GT)

Solution: 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

Addressing challenges of AI

Uncertainty quantification

INTERSPEECH’ 21, IEEE JBHI’24

Data privacy protection

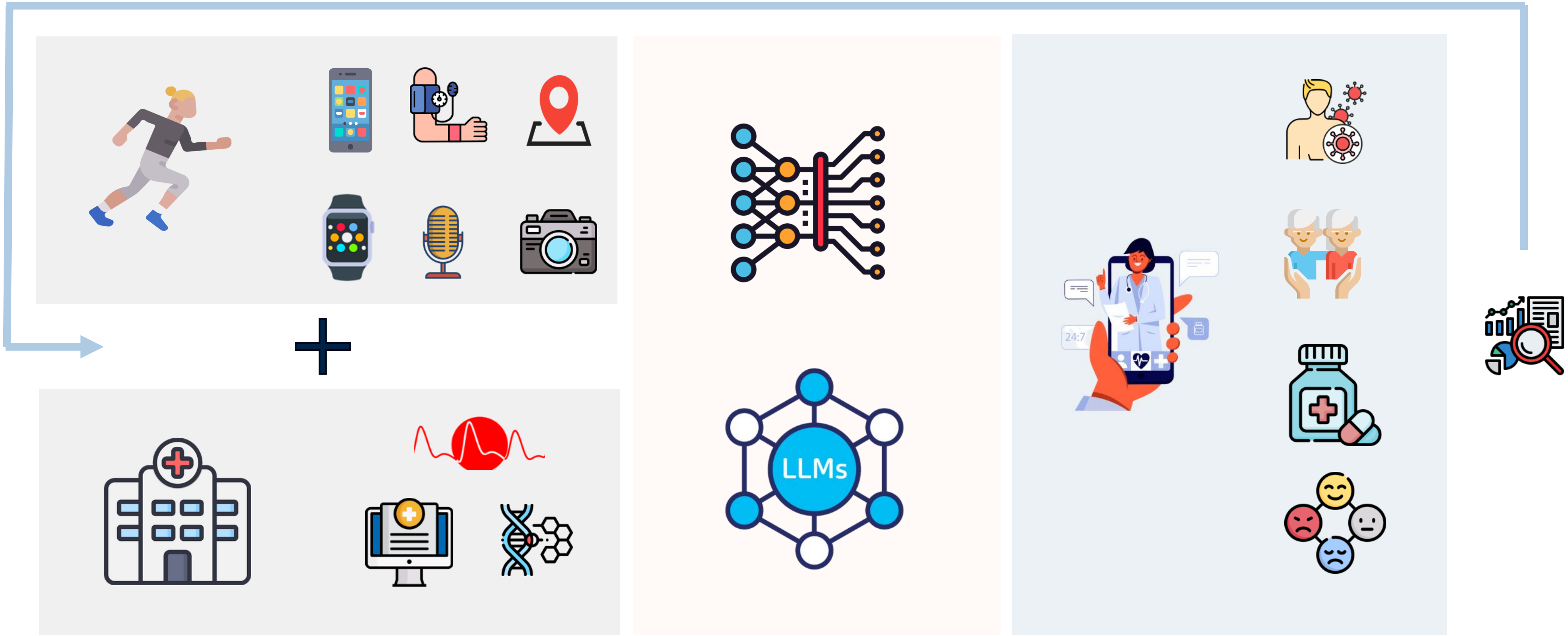
ICASSP’23, KDD’24

Model generalizability

Arxiv’24

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

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

erc European Research Council
Established by the European Commission

腾讯 Tencent

中国移动 China Mobile

HUAWEI

OKIA BELL LABS

Flower Labs

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**)
- 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 Drant 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
- **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. Xia**^, 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

Mobile and Wearable Health Seminar Series

Add to your list(s) Send you e-mail reminders Further detail
Subscribe using ical/vcal (Help)

Guy Fagherazzi: Vocal biomarkers for minimally disruptive clinical research: how to get there?

Februray 2024

Alex Mariakakis, University of Toronto: Embracing Ubiquitous Technology to Complement, Scale, and Extend Traditional Healthcare

April 2024

https://mobile-systems.cl.cam.ac.uk/seminars.html

UNIVERSITY OF CAMBRIDGE

QUEENS' COLLEGE CAMBRIDGE

THANK YOU

Wechat

Linkedin

Google scholar

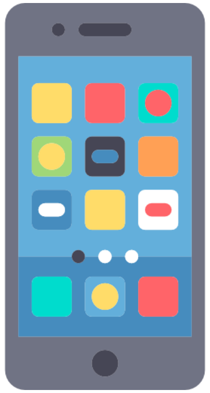
Tong Xia
tx229@cam.ac.uk

AI-empowered mHealth brings a better future!

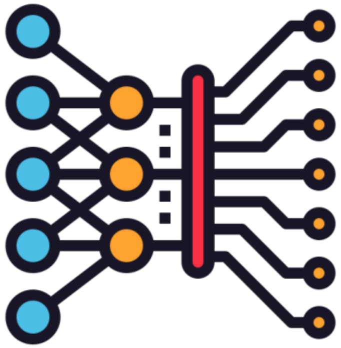
Tong Xia

June, 2024

My research interests



Mobile computing



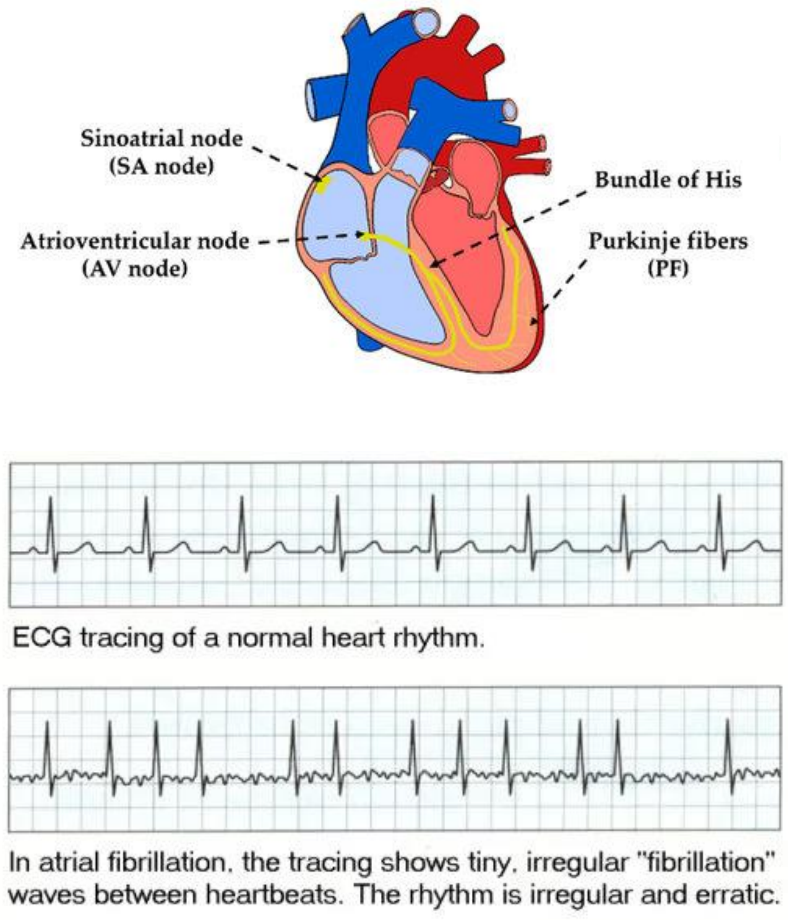
Machine learning

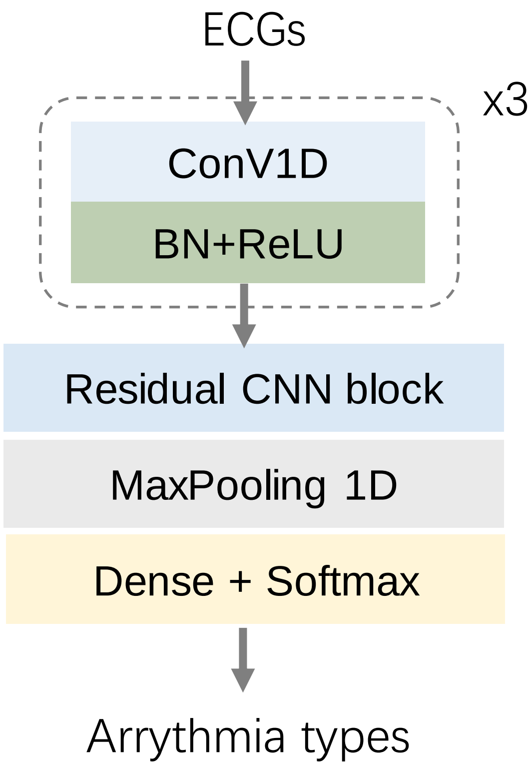
Healthcare

Backup

Tong Xia (tx229)

Electrocardiogram (ECG) -based heart arrythmia detection





- ✓ **ACC > 0.78** for 5-class arrythmia classification
- ✓ **Sensitivity > 0.88** for AF detection

⇩

- ✓ The promise of detecting heart anomality using mobile/wearable devices

[3] **T. Xia.** Reliable and decentralised deep learning for physiological data. PhD Thesis 2024.

[4] **T. Xia**, J. Han, C. Mascolo. Benchmarking Uncertainty Quantification on Biosignal Classification Tasks under Dataset Shift. Workshop on Health Intelligence, AAAI 2022

Can ‘**our footprints’** tell how healthy we are?

- ✓ Ubiquitous
- ✓ Passive

Mobility-based health condition inference

Using sparse GPS records to recover daily mobility patterns

Physical mobility features:

- Radius of gyration
- Standard deviation of displacements
- Distribution entropy of places visited

Contextual mobility features:

- Visit willingness to restaurant, entertainment, sport, scenic spot, fast food, and tobacco/liquor shop

- A health survey conducted in **13 major hospitals covering 2 months** in Beijing
  - **1056** outpatients paid at least one visit to the hospital
  - **1056** healthy hospital staff
  - ✓ Binary prediction **AORUC of 0.8**

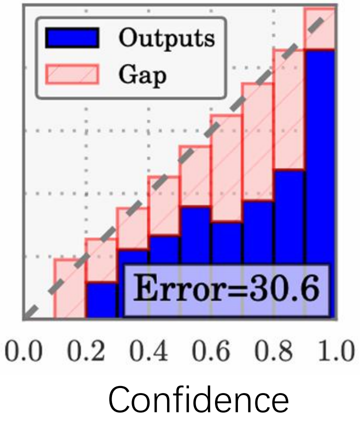
[5] **T. Xia**, et al. Attnmove: History Enhanced Trajectory Recovery via Attentional Network. **AAAI** 2021.
[6] Y. Zhang, F. Xu, **T. Xia**, and Y. Li. Quantifying the Causal Effect of Individual Mobility on Health Status in Urban Space. **UbiComp** 2021.

2.3 Evaluation

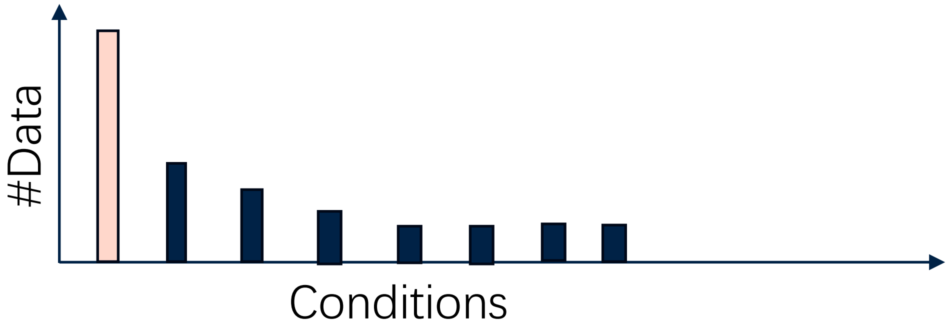
- **Diagnosis (classification) performance**:
  - Sensitivity (recall) = $\frac{TP}{TP+FP}$
  - Specificity = $\frac{TN}{TN+FN}$
  - Accuracy (Acc) = $\frac{TN+TP}{TN+FN+FN+FP}$
  - ROC-AUC

		Actual Values	
		Positive	Negative
Predicted Values	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

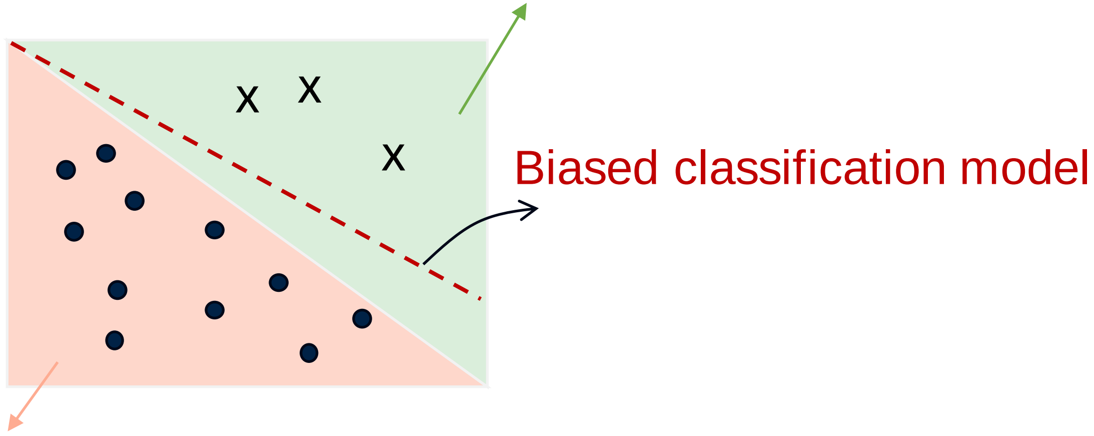
- **Confidence-related performance**:
  - ECE (Expected Calibration Error)
  - Brier score (accuracy of predicted probabilities)
  - ROC-AUC for distributional shift detection



2.1 Class imbalance and long-tailed learning



Real distribution for unhealthy samples (minority class)

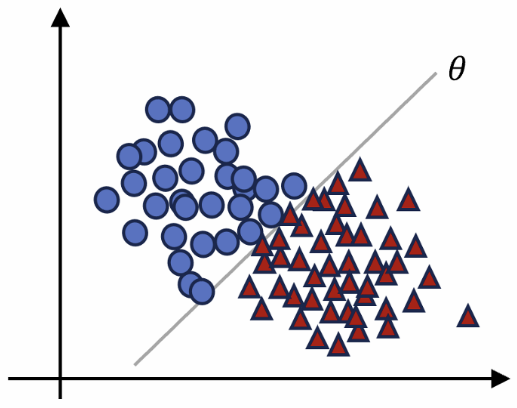


Real distribution for healthy samples (majority distribution)

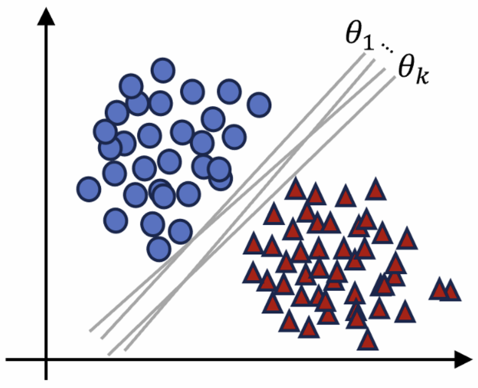
Related work - long-tailed learning:

- Data-level method
  - Up-sampling
  - Down-sampling
  - Augmentations
- Algorithm-level methods
  - Cost-sensitive loss
  - Scaling thresholds
  - Weighted loss

2.2 Uncertainty quantification



(a) Aleatoric uncertainty.

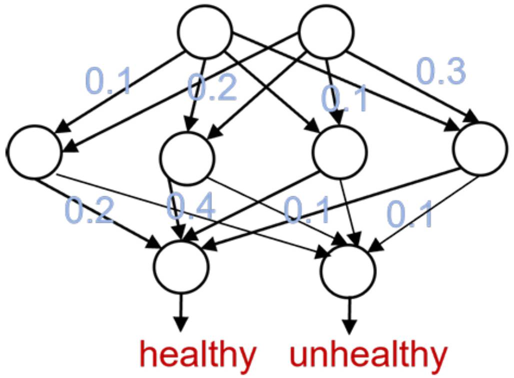


(b) Epistemic uncertainty.

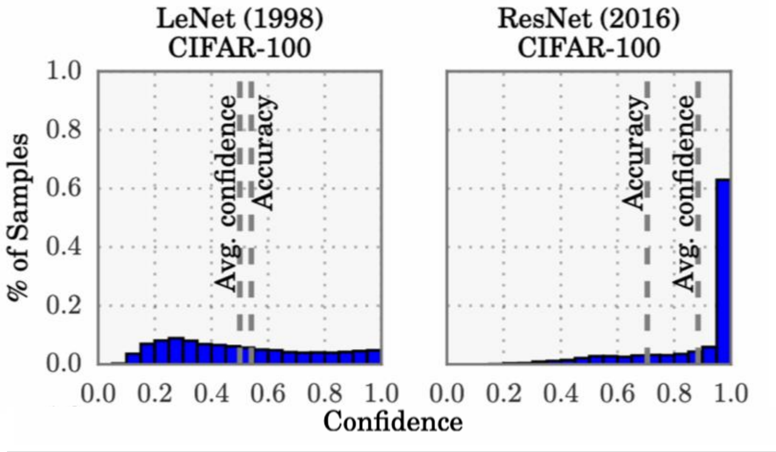
Figure 2.5: **An illustration of uncertainty**. In deep learning, two types of uncertainties are commonly recognised. The first type, called *aleatoric uncertainty* arises from noise, perturbations, and biases present in the data. When the data is noisy or unrepresentative, it can introduce variability in both the input and output. For instance, in (a), after fitting the model θ, data samples located in the overlapping region exhibit high aleatoric certainty. The second type of uncertainty is known as *epistemic uncertainty*, which stems from a lack of sufficient knowledge about the optimal model. (b) illustrates high epistemic uncertainty, as multiple models can fit the training data equally well. This kind of uncertainty can be reduced by adjusting the model or supplementing it with additional data.

2.2 Model overconfidence and uncertainty quantification

Standard neural network

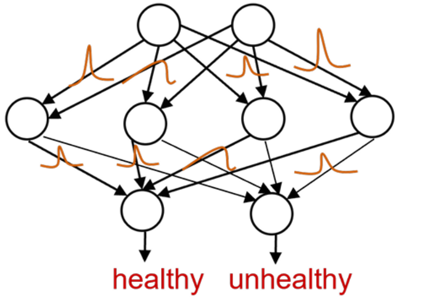


DL overconfident

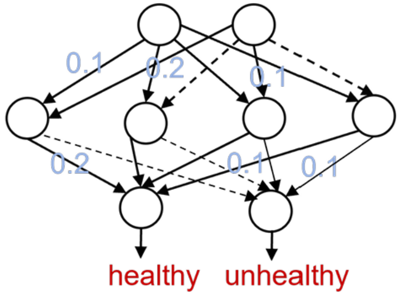


Related work – Uncertainty quantification:

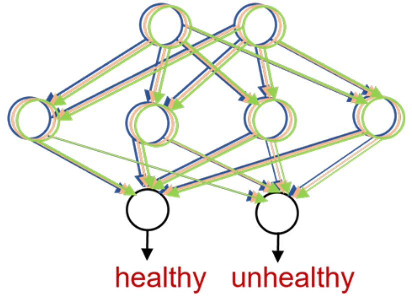
Bayesian neural network



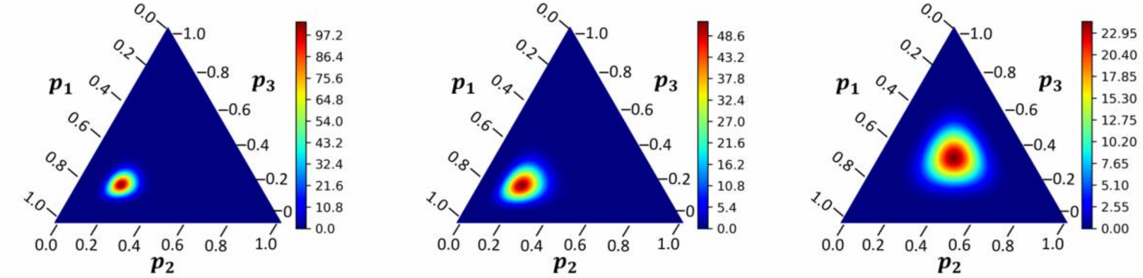
Dropout approximation



Ensemble approximation



Evidential deep learning:



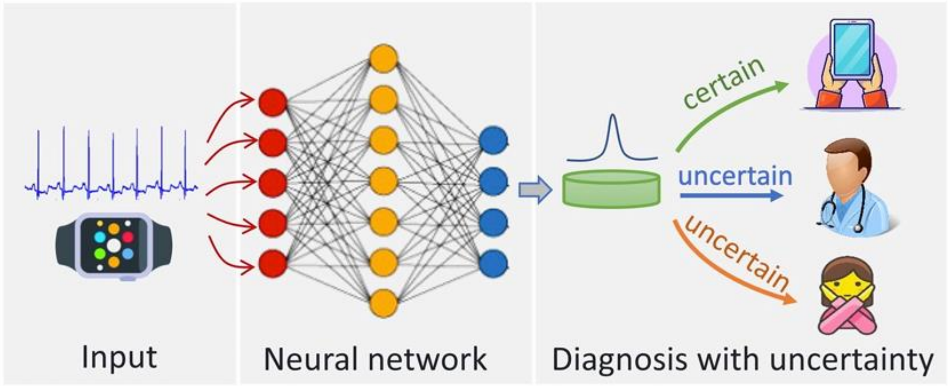
(a) $\boldsymbol{\alpha} = [60,20,20]$
$\mathbb{E}[p] = [0.6,0.2,0.2]$

(b) $\boldsymbol{\alpha} = [30,10,10]$
$\mathbb{E}[p] = [0.6,0.2,0.2]$

(c) $\boldsymbol{\alpha} = [10,10,10]$
$\mathbb{E}[p] = [0.33,0.33,0.33]$

3.2 Uncertainty-aware deep learning for multi-class physiological data

- **Summary**
  - Contributions:
    - Introduced a novel and efficient class-balanced EDL for multi-class physiological data
    - Extensive experiments demonstrate its superiority
    - Provide a systematic understanding for a reliable automated system for health diagnostics

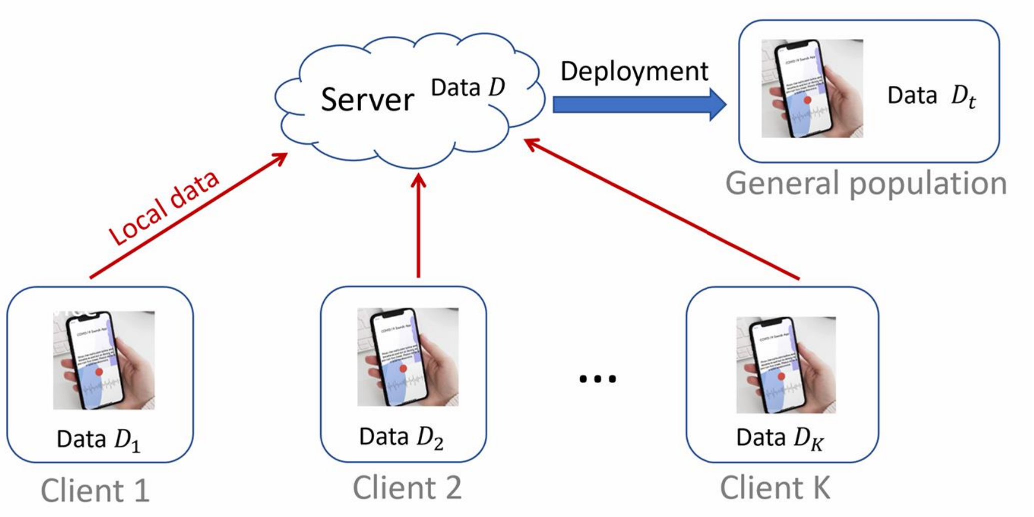


An uncertainty-aware deep learning driven health diagnostics system.

  - Publications:
    1) Xia, T., et al. *Hybrid-EDL: Improving evidential deep learning for uncertainty quantification on imbalanced data*. In Workshop on Trustworthy and Socially Responsible Deep Learning, NeurIPS 2022 (Primary study)
    2) Xia, T., et al. *Uncertainty-aware health diagnostics via class-balanced evidential deep learning*. IEEE Journal of Biomedical and Health Informatics J-BHI 2024 (Full study)

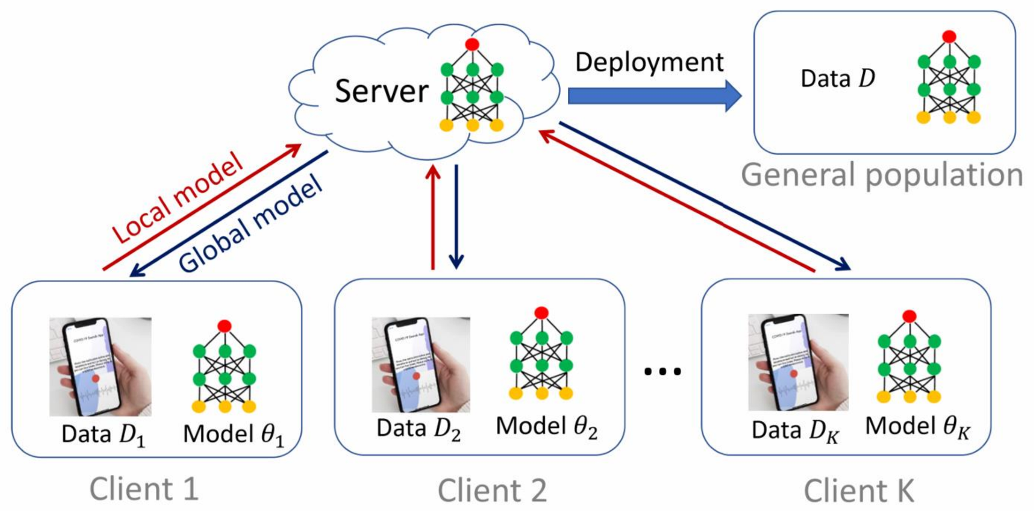
2.3 Data privacy and federated learning

Gathering health data for ML research can face privacy problem given the sensitivity of personal information



(a) Model training using centralised data.

Related work – Federated learning:



(b) Model training using distributed data.

FedAvg:

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