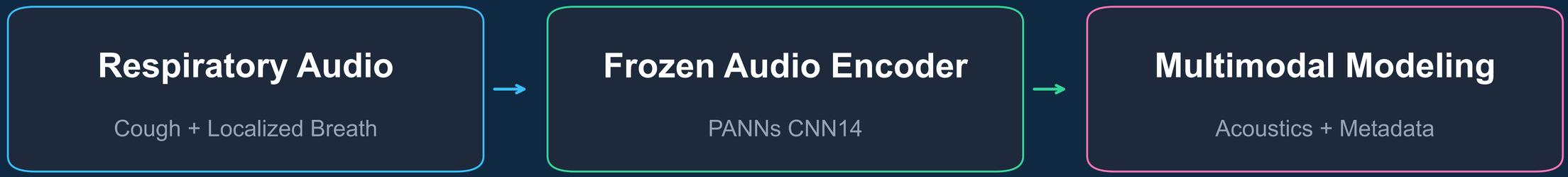
Listening to the Lungs

Acoustic AI & Multimodal Detection of Tuberculosis



Lev Smolsky

MIT PRIMES | Alterovitz Lab (Harvard Medical School & MIT)

Talk Roadmap

01

The Problem

TB screening and respiratory acoustics

02

The Approach

Frozen features, separate experts, simple fusion

03

The Findings

What do the modalities add?

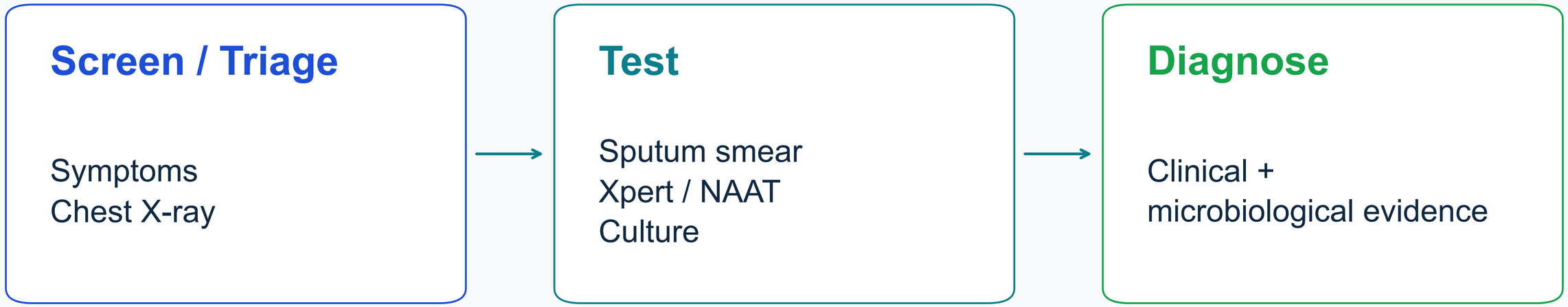
04

The Limitations

Small samples and cross-dataset transfer

Where Could Audio Help?

TB is an airborne bacterial infection, often affecting the lungs



Low-cost screening to help prioritize TB testing

Help people with TB reach diagnosis and treatment sooner

WHO screening guidelines, 2021; Lewinsohn et al., 2017. NAAT: nucleic-acid amplification test.

From Cough Screening to Multimodal Audio

Cough Screening

CODA · TBscreen

Established research direction

Localized Lung Sounds

Chest-wall recordings

Smaller TB evidence base

Our Question

Cough + breath + metadata

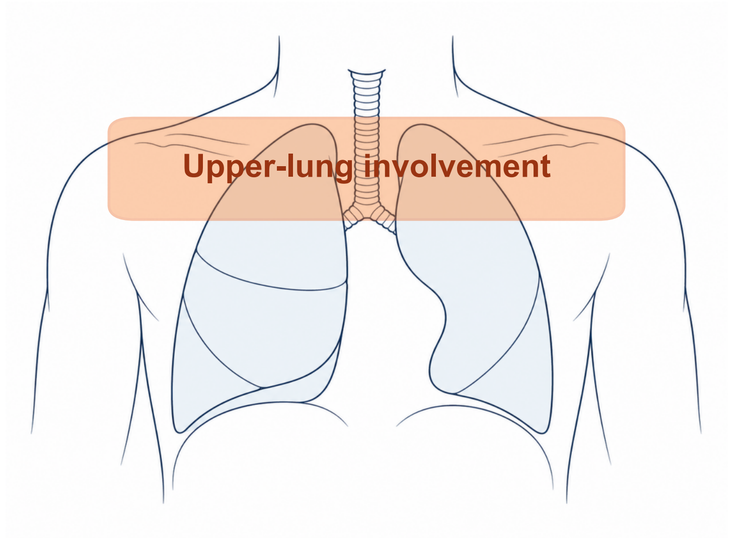
Do synchronized inputs add value?

CODA: Huddart et al., 2024; Jaganath et al., 2025. TBscreen: Sharma et al., 2024. Lung sounds: Becker et al., 2013.

Why Listen at Multiple Lung Sites?

Pulmonary TB can affect different lung regions

Regional Lung Changes



Compare sounds across lung regions

Cough

Mixed airway signal

Localized Breath

10 anatomical recording sites

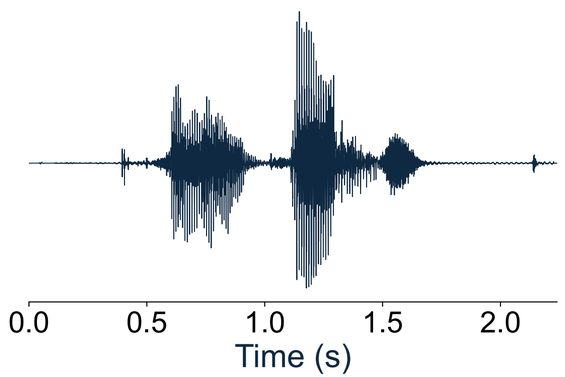
Clinical motivation: Nachiappan et al., 2017; Bohadana et al., 2014

Turning Sound into Features

Real audio example: “Hello PRIMES” voice memo

1. Raw Waveform

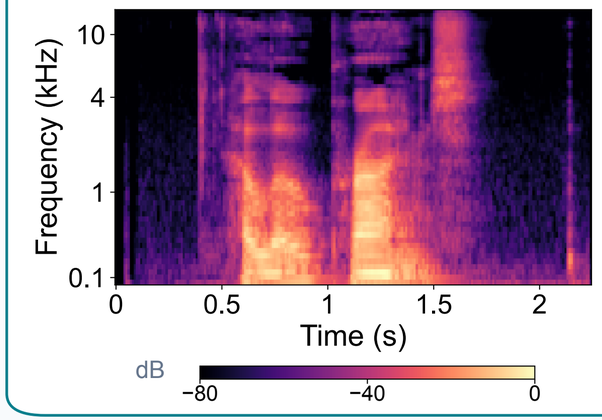
Sound vibrations over time



STFT + mel →

2. Mel-Spectrogram

Color = relative sound energy



CNN14 →

3. Deep Embedding

2,048-dimensional vector

f_1 f_2 … f_{2048}

Learned feature vector

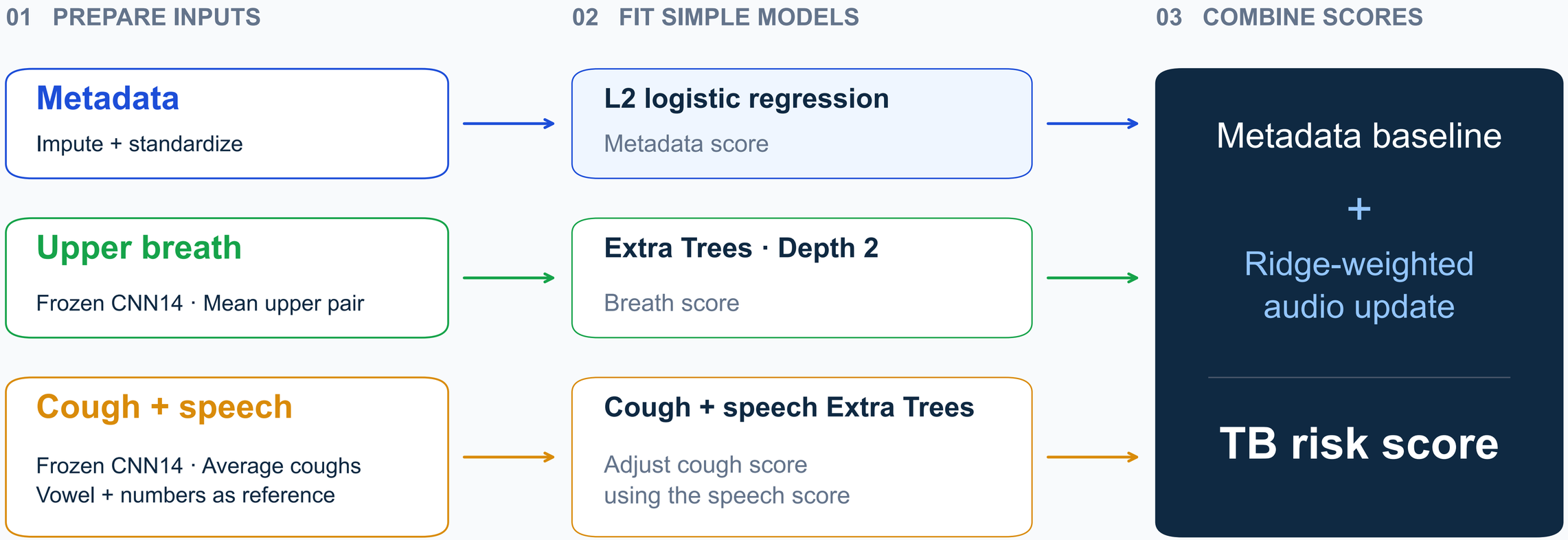
AudioSet: ~5,800 hours of diverse sounds

Speech · Music · Animals · Car horns | CNN14 weights held fixed in our study

Kong et al., 2020; Google AudioSet. Original-duration memo; 64 mel bands; feature-vector diagram is symbolic.

Our Small-Sample Modeling Pipeline

200 participants · Primary analysis · Participant-level cross-validation



SMALL-COHORT DESIGN

Freeze the encoder; limit model depth

INTERPRETABLE STRUCTURE

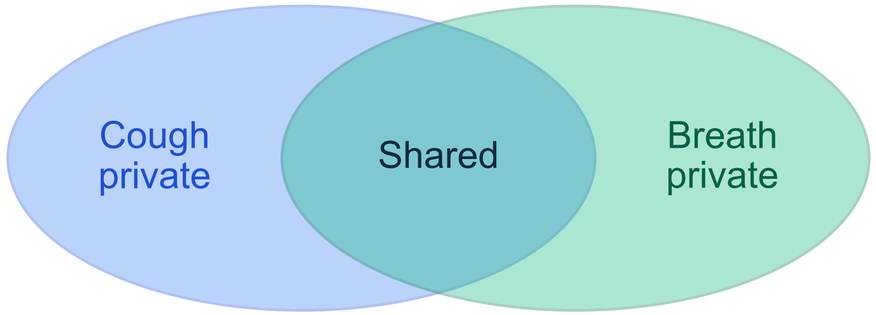
Inspect each modality's contribution

Evaluation: 10 × 5-fold participant CV; supervised fusion inputs cross-fitted within each training fold.

Shared vs. Modality-Specific Information

Exploratory JIVE-style decomposition of shared and modality-specific signals

Each modality = shared + private + residual



Shared factors alone were insufficient

Broader breath pooling also underperformed

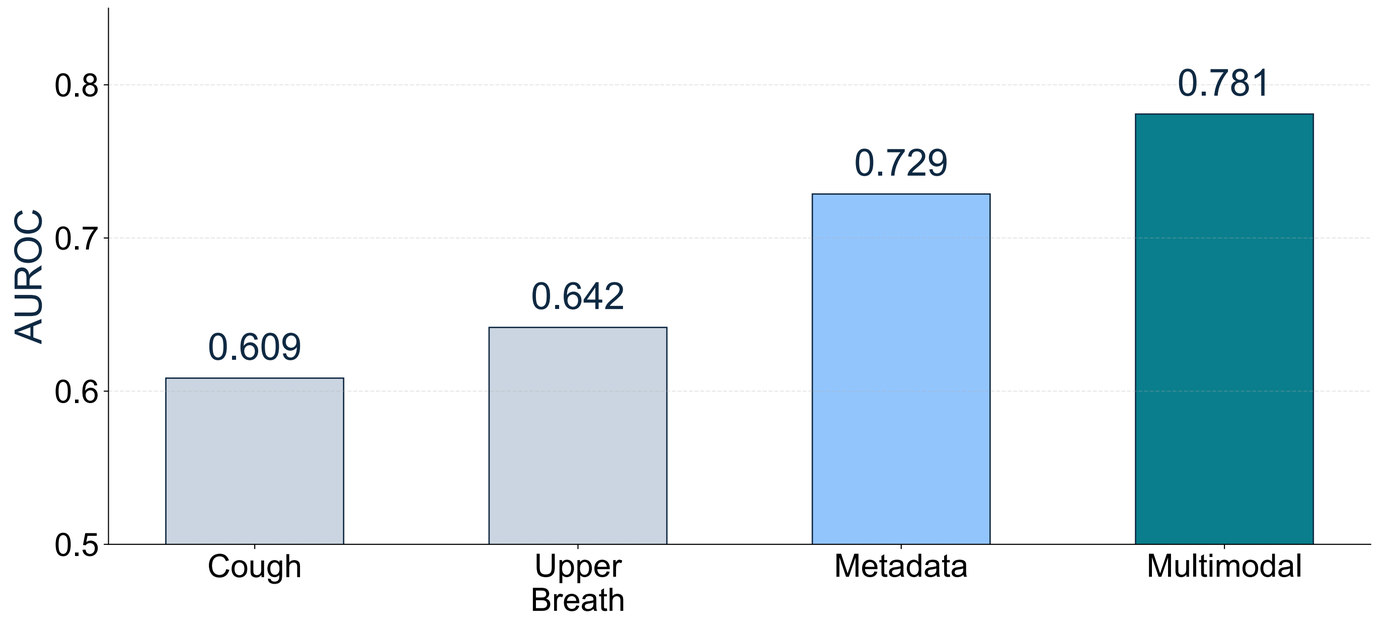
Preserve Separate Experts

Combine scores, rather than collapse all inputs

Decomposition framework: Lock et al., 2013. Schematic overlap illustrates statistical components.

Combining Modalities Adds Information

Internal development comparison · Same 200 participants and splits



Stricter Nested Analysis

Metadata → Multimodal

0.710 → 0.746

Also lower log loss
and Brier score

Development: depth-4 experts + raw cough. Nested: depth-2 experts + speech-adjusted cough.

Development candidate selected in this cohort; independent validation is the next step.

The Challenge: Transfer Across Datasets

External evaluation of the frozen cough branch across new cohorts

External Cough-Branch AUROC

CODA — 0.592

1,082 participants

TBscreen — 0.623

120 participants

Acquisition + Population Shift

Devices · Tasks · Recording protocols

COUGHVID: Similar Group Scores

Similar scores across respiratory groups
New recording conditions limit interpretation

External data: Huddart et al., 2024; Sharma et al., 2024; Orlandic et al., 2021.

Three Takeaways

01

Complementary Modalities

Localized breath + cough add to metadata within this cohort.

02

Small-Sample, Structured Modeling

Frozen features · Shallow experts · Separate modality scores

03

Independent Validation Comes Next

New-site multimodal data + linked radiology

Acknowledgments: Dr. Pei Shaojun · Prof. Gil Alterovitz · MIT PRIMES