Low-Power Weakly-Supervised Audio Detection for Real-World Mosquito Surveillance

Danika Gupta (The Harker School) , Ming Zhao (Arizona State University), Neha Rajendra Vadnere (Arizona State University)

NeurIPS Climate Change AI Workshop 2025

Problem

- Mosquito-borne diseases affect >3 billion worldwide.
- More than 600,000 fatalities annually. Example - recent outbreaks in China attributed to climate change.
- Identifying/tracking mosquito species essential for predicting disease, yet uses manual, subjective methods.
- Climate change further aggravates this challenge by altering habitats.

Outbreak News Today

ASIA

China reports more than 16,000 chikungunya cases

ROBERT HERRIMAN
OCT 07, 2025

Share

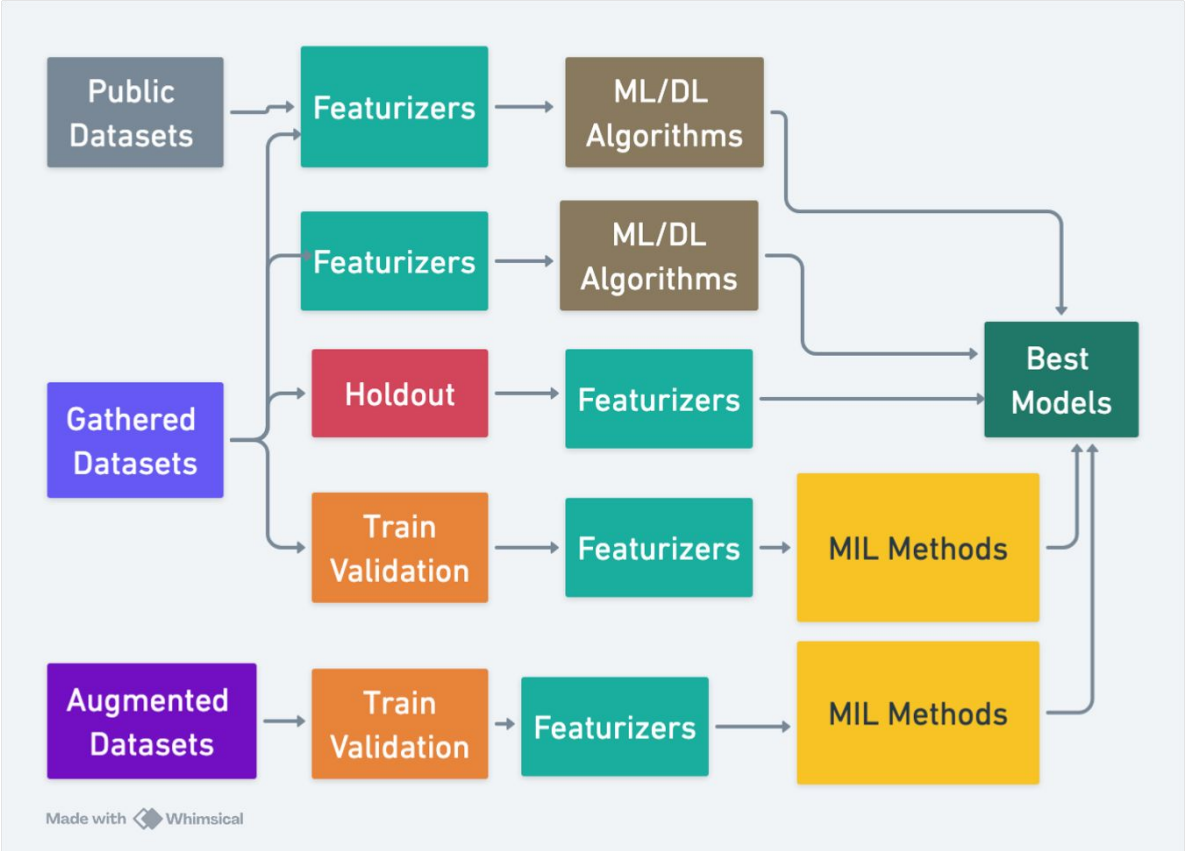
As of the end of September, more than 16,000 locally transmitted cases has been reported in Guangdong Province. All cases were laboratory-confirmed. This represents the largest documented chikungunya outbreak to date in China.

Audio Detection – Benefits and Challenges

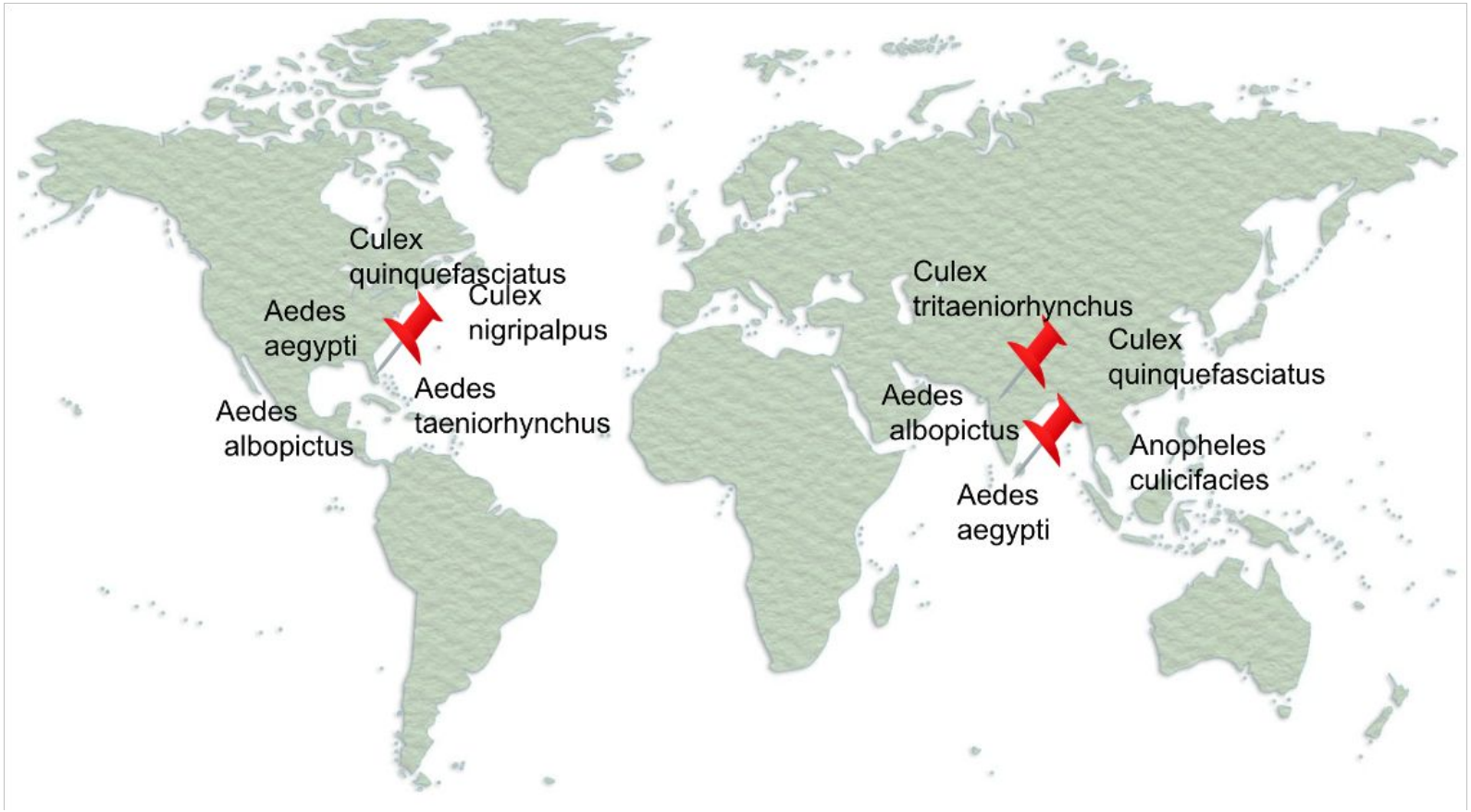
- Audio detectors can be implemented at lower power and cost (compared to video or image detectors, and have less privacy concerns or be impacted by lighting variations.
- Labeling audio requires expert entomologist support.
- Mosquito detection requires training on background data from affected locations. Existing research primarily focused on entomologist-labeled audio or video and mosquito classification

Methodology

- Citizen scientist video labeling coupled with audio classifiers.
- Leverage Multi Instance Learning to adapt to weak video labels for audio data.
- Varied MIL algorithms and librosa/YAMNet featurization, compared to ML and CNN baselines.
- HumBugDB used as supplemental dataset



Dataset Gathering



Preliminary Dataset

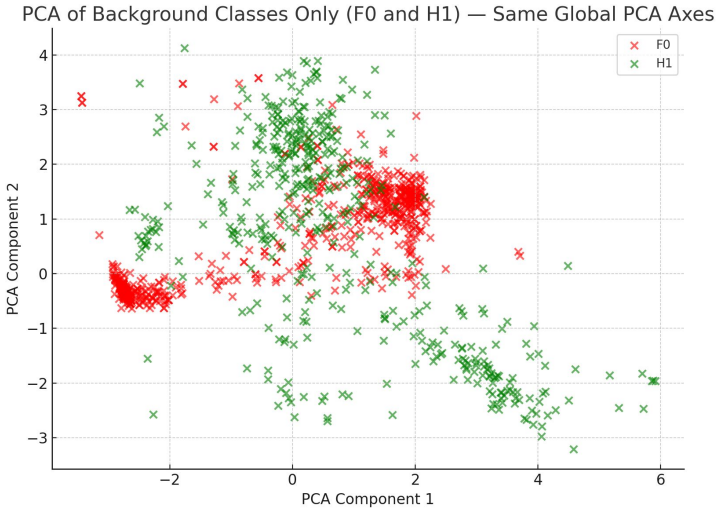
- 2144 videos gathered from florida, with preliminary results on 1200 videos.
- Videos span day/dusk/night – time of mosquito presence.
- Each video is 10s and manually labeled by citizen scientists, with 64.8% background and 35.2% mosquito.
- Audio preprocessed to filter out sounds from the mosquito attractant device.

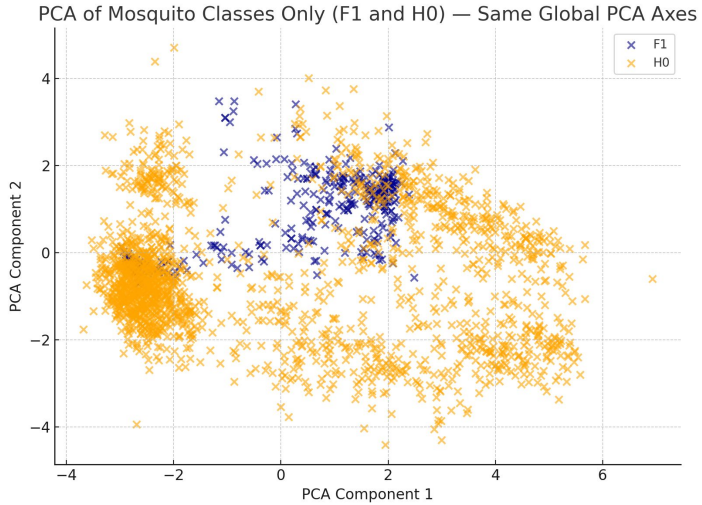
Preliminary Results

Dataset	Labels	Method	Best AUC
HumbugDB	Entomologist	Spectral Features + Random Forest	0.997
Florida	Video	Best Humbug Model	0.553
Florida	Video	Varied algorithms (Spectral Features + Random Forest, YAMNet embeddings)	0.673
Florida	Audio Expert	Spectral Features + Random Forest	0.852
Florida	Video	Spectral Features + MIL Gated Attention	0.726

Preliminary Results

- Comparing HumbugDB to our weakly labeled dataset reveals a strong distributional mismatch- Kolmogorov–Smirnov 100% of features significantly differing ($p < 0.05$),





Pathway To Impact

- >2800 mosquito variants, >330 implicated in human disease, but only a handful of datasets.
- We contribute the first multimodal audio/video time/location stamped mosquito surveillance dataset for researchers to explore multimodal surveillance and contextual metadata data integration.
- With 6000 more labeled video samples within 3 months, published testbed design ($90) and citizen science labeling guide, repeatable testbeds can be set up for tailored mosquito surveillance with limited entomologist support.
- The approach may also benefit other domains critical to climate adaptation, but with limited data.

Code and Dataset will be available at at https://github.com/danikagupta/DeepMosquito as project proceeds

Thank you!