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

Danika Gupta
The Harker School
San Jose, CA 95129
dan@gprof.com

Ming Zhao
Arizona State University
Tempe, AZ 85281
mingzhao@asu.edu

Neha Rajendra Vadnere
Arizona State University
Tempe, AZ 85281
nvadnere@asu.edu

Abstract

Climate change is expanding the geography and altering habitats of mosquitoes that transmit Dengue, Zika, etc. For example, China’s first large-scale chikungunya outbreak in August 2025 underscores the urgency for scalable and adaptive mosquito surveillance. Existing approaches often rely on expert entomologists or extensive labeled datasets, making them poorly suited for rapid deployment in low-resource settings. Our approach uses weak labels from time-aligned video recordings to train audio classifiers via Multi Instance Learning (MIL). This approach combines local data, citizen labelers and low power audio classification to enable rapid adaptation to climate driven outbreaks. Early proof points suggest potential for MIL in audio classification from mosquito video labels.

1 Introduction

Mosquito-borne diseases pose a significant public health challenge. In mid-2025, China experienced its first-ever large-scale outbreak of chikungunya — with over 7,000 reported cases [29], primarily where the population had little prior immunity. Authorities attributed the rapid spread to changing regional climate dynamics, underscoring the pressing need for low-cost, lightweight mosquito surveillance systems in at-risk regions. Malaria claims over 600,000 lives annually, while dengue has seen a thirtyfold increase in incidences over the past fifty years [1], with a large variety of mosquitoes spreading pathogens [3]. Climate change expands mosquito habitats, increasing the risk [2]. Mosquito surveillance is crucial for outbreak prediction but is limited by manual, subjective methods such as distribution maps based on expert-validated data [4], and field genetic studies [5]. Datasets like ABUZZ [6] and HumbugDB [7] have captured mosquito wingbeat sounds, and datasets like MosquitoAlert [8] have pictures. While image datasets generate good classification results [31], they are hard to gather in the wild without exposure to being bitten. Audio data is easier to gather with citizen science and unmanned sensors, but tend to require expert support to label.

We target two challenges in practical mosquito identification, labeling local noisy mosquito data, and low-power mosquito detection deployment. We use video labeling implementable by citizen scientists without entomologist training, which is then used to train power-efficient audio classifiers. Video classification in real time is not practical given the cost and power requirements, lighting dependencies and privacy concerns. We exploit the link between video and audio, using video for labeling and audio for classification, leveraging MIL applied to audio classification. We further propose to implement such audio classifiers in low-power off-the-shelf hardware, enabling environment tailored sensors.

2 Limitations of Existing Mosquito Research

Traditional strategies have largely required entomologist experts to monitor, identify, and ultimately eradicate targeted mosquito populations [10]. Machine learning has shown promise in these endeavors [9], from deep learning for identification, monitoring, and control [11], with visual methods in [12,13,14] and acoustic and pseudo-acoustic techniques in [15,16,17,18,19,31]. Audio datasets such as [6,7,20] are labeled via expert identification of caught mosquitoes or expert review of audio data. Classification via such datasets generates high accuracy and precision [30]. Meta-datasets like MACSFeD provide aggregated species-specific wingbeat and behavioral traits, valuable for tuning detection systems [21]. However, these datasets are best for classifying mosquitoes types, while our first task is to identify mosquito presence. Only [7] provides background samples and our experiments have revealed that classifiers trained on it do poorly in other locales. The ideal scenario would gather data from the target locale, label each sound snippet, and create a custom classifier. This however requires trained labelers, impractical for many low resource scenarios.

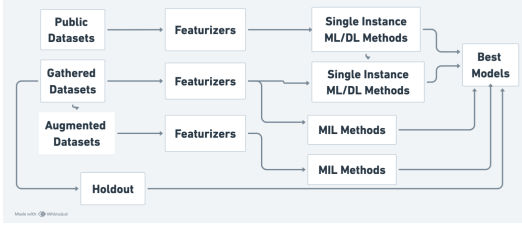


Figure 1: Methodology

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

Figure 2: Sample Results

3 Methodology

Our methodology (Figure 1) gathers local video labeled by citizen scientists trained to recognize mosquito movement (zig-zag etc.), strengthened via inter-rater reliability. The video labels are leveraged for audio classification. The challenge is that seeing a mosquito does not imply hearing it or vice versa, but mosquito sound is likely in temporally close audio. Given this, we explore MIL to extract the signals within audio instances in a bag if the entire bag is labeled mosquito. We evaluate MIL by comparing single instance approaches using public data [6,7] with MIL variants trained both our gathered data and augmented data integrating contextual information.

To our knowledge, we are the first to use MIL in this way, so we have selected MIL variants likely to be useful where a bag contains a time ordered sequence of audio instances whose overall video showed one or more mosquitoes. We chose Attention-based MIL [23] which assigns importance weights to instances, targeting sparse key events like occasional mosquito sounds in a noisy bag, Gated Attention MIL which adds a nonlinear gating mechanism to capture context-dependent audibility, CLAM [24] which integrates attention pooling with clustering constraints to detect temporal or spatial bursts of events, TransMIL [26] which captures long-range dependencies between instances, and Temporal MIL [27] which preserves the chronological order of instances to exploit evolving sound patterns. Other architectures, such as [25, 28] may also be worth evaluating. We have selected featurization options including extracting average spectral, cepstral and rhythmic features, generating overlapping log-mel spectrogram slices that preserve the fine-grained time–frequency structure for each clip (both via [32]), YAMNet featurization [33] and MEL spectrograms. We also plan to augment the dataset with synthetic bags containing known mosquito sounds from [6–8], specific to mosquitoes known to be present in each region. The best models are compared against a holdout set from our testbed. An ablation study can help isolate impact of featurization and model techniques. The methodology can be further improved, for example by helping citizen scientists distinguish between mosquitoes and known regional insects, experimentation with a range of audio/video capture devices and attractant mechanisms (such as those used in [7]), and backfilling with expert validation of sample data points.

We have established testbeds in Florida (United States), Uttar Pradesh (India), Colombo (Sri Lanka), where at least six known mosquito types (*Aedes aegypti*, *Aedes albopictus*, *Culex quinquefascia-*

tus, *Culex nigripalpus*, *Aedes taeniorhynchus*, and *Anopheles stephensi*) spread dengue, Malaria, chikungunya, West Nile, St. Louis encephalitis, lymphatic filariasis, and Japanese encephalitis.

4 Initial Proof Points

We have gathered 2000 10s videos from Florida using Wyze Cam Pan v3 [22], a weather-resistant 1080p HD camera with integrated time-aligned audio recording. To date, 1200 video snippets have been manually labeled (64.8% background and 35.2% mosquito). Each snippet was filtered of sounds that could be caused by the attractant trap [23] (via filter and interpolation techniques [34,35]). Comparing [7] to our weakly labeled dataset reveals a strong distributional mismatch with Kolmogorov–Smirnov showing 100% of features significantly differing ($p < 0.05$), with top discriminating features exhibiting near-zero overlap, confirmed by PCA projections. Figure 2 further shows that while average spectral value featurization and Random Forest generate high AUC on [7], this model has very poor results on the Florida dataset. Classification via our weak labels and a range of ML and YAMNet methods revealed best AUC of 0.67. Expert audio labeling of the Florida data gave AUC of 0.852. Multi-instance audio classification (bags of 1s snippets corresponding to the 10s video label) shows AUC of 0.726 in with Gated Attention MIL and similar results with MIL variants (skipped for space). While not conclusive, the results suggest necessity for local data rather than sole reliance on clean datasets, and suggest that MIL has potential to extract from weak labels what other methods can with strong labels. Sound acquisition, spectral featurization and MIL inference have been implemented on Raspberry PI 4 [36] using Python and ONNX [37] with inference taking less than 2ms, providing supporting evidence of viability in low power settings.

5 Pathway to Impact

There are >2800 mosquito variants, of which >330 are implicated in human disease [41], but only a handful of datasets. Our published dataset (to our knowledge the first multimodal audio/video time/location stamped mosquito surveillance dataset) enables researchers to explore multimodal surveillance and contextual metadata data integration such as [22], enabling exploration of time/season optimized methods such as [42]. We expect to add 6000 more labeled video samples from our testbeds within 3 months. We will also publish our testbed design (\$90) and citizen science labeling guide, making repeatable testbeds easy to set up anywhere for tailored mosquito surveillance with limited entomologist support. MIL enables low power automated audio surveillance in rural areas (our Raspberry Pi prototype is <\$100 with 10x power savings over small desktops, further reduced by 5x by porting to ESP32 (a task in process). Consumers can be health organizations and governments for rapid response. The speed of chikungunya spread in China shows how even a few days can make a difference [29]. We already have interest from Sri Lanka’s parasitology experts for pilot trials. Countries like Somalia and South Sudan, with heavy mosquito burden and limited entomology resources can particularly benefit [39,40]. Furthermore, the approach may benefit other domains critical to climate adaptation, but with limited data. Examples are wildfire, wildlife and undersea species monitoring where bioacoustics have shown promise but background noise is complex and location specific, and where integrating visual labels can provide an advantage.

6 Summary

We propose and demonstrate early evidence that mosquito presence can be detected from weakly-labeled audio using only video-based annotations, enabling low-power, real-world mosquito surveillance without expert labeling — a crucial adaptation strategy as mosquito habitats shift under climate change. Code, datasets and testbed designs will be available on Github at <https://github.com/danikagupta/DeepMosquito> as the project proceeds.

7 Acknowledgements

We gratefully acknowledge the feedback of the anonymous reviewers, as well as the assistance of Michel Roger in gathering the Florida videos.

References

- [1] Raja Danasekaran, MANI Geetha, Kalaivani Annadurai, and Jegadeesh Ramasamy. Small bite, big threat: the burden of vector-borne diseases. *Iranian Journal of Public Health*, 43(7):1014–1015, 2014.
- [2] Felipe J Colón-González, Maquins Odhiambo Sewe, Adrian M Tompkins, Henrik Sjödin, Alejandro Casallas, Joacim Rocklöv, Cyril Caminade, and Rachel Lowe. Projecting the risk of mosquito-borne diseases in a warmer and more populated world: a multi-model, multisenario intercomparison modelling study. *The Lancet Planetary Health*, 5(7):e404–e414, 2021.
- [3] WHO. Vector borne diseases. World Health Organization www.who.int/news-room/fact-sheets/detail/vector-borne-diseases, 2020.
- [4] EUDC. Mosquito maps. <https://www.ecdc.europa.eu/en/diseasevectors/surveillance-and-disease-data/mosquito-maps>.
- [5] Direct pcr of indigenous and invasive mosquito species: a time- and costeffective technique of mosquito barcoding. *Med Vet Intermol*, 2016.
- [6] Haripriya Mukundarajan, Felix Jan Hein Hol, Erica Araceli Castillo, Cooper Newby, and Manu Prakash. Using mobile phones as acoustic sensors for high-throughput mosquito surveillance. *elife*, 6:e27854, 2017.
- [7] Ivan Kiskin, Marianne Sinka, Adam D Cobb, Waqas Rafique, Lawrence Wang, Davide Zilli, Benjamin Gutteridge, Rinita Dam, Theodoros Marinos, Yunpeng Li, et al. Humbugdb: a large-scale acoustic mosquito dataset. *arXiv preprint arXiv:2110.07607*, 2021.
- [8] Mosquito Alert, “Mosquito alert opens its set of 20,000 photographs to train deep learning software.” <https://www.mosquitoalert.com/en/mosquit-alert-obre-la-seva-colB7leccio-de-20-000-fotografies-per-entrenar-programes-de-deep-learning/> (2023).
- [9] Pedro Miguel Rodrigues, João Paulo Madeiro, and João Alexandre Lobo Marques. Enhancing health and public health through machine learning: Decision support for smarter choices, 2023.
- [10] Handi Dahmana and Oleg Mediannikov. Mosquito-borne diseases emergence/resurgence and how to effectively control it biologically. *Pathogens*, 9(4):310, 2020.
- [11] Basudev Nayak, Bonomali Khuntia, Laxman Kumar Murmu, Bijayalaxmi Sahu, Rabi Sankar Pandit, and Tapan Kumar Barik. Artificial intelligence (ai): a new window to revamp the vector-borne disease control. *Parasitology Research*, 122(2):369–379, 2023.
- [12] Kazushige Okayasu, Kota Yoshida, Masataka Fuchida, and Akio Nakamura. Vision-based classification of mosquito species: Comparison of conventional and deep learning methods. *Applied sciences*, 9(18):3935, 2019.
- [13] Daniel Motta, Alex Alisson Bandeira Santos, Ingrid Winkler, Bruna Aparecida Souza Machado, Daniel André Dias Imperial Pereira, Alexandre Morais Cavalcanti, Eduardo Oyama Lins Fonseca, Frank Kirchner, and Roberto Badaró. Application of convolutional neural networks for classification of adult mosquitoes in the field. *PloS one*, 14(1):e0210829, 2019.
- [14] Li-Pang Huang, Ming-Hong Hong, Cyuan-Heng Luo, Sachit Mahajan, and Ling-Jyh Chen. A vector mosquitoes classification system based on edge computing and deep learning. In *2018 Conference on Technologies and Applications of Artificial Intelligence (TAAI)*, pages 24–27. IEEE, 2018.
- [15] Marcelo Schreiber Fernandes, Weverton Cordeiro, and Mariana Recamonde-Mendoza. Detecting aedes aegypti mosquitoes through audio classification with convolutional neural networks. *Computers in Biology and Medicine*, 129:104152, 2021.
- [16] Ayush Jhaveri, KS Sangwan, Vinod Maan, and Dhiraj. Deep learningbased mosquito species detection using wingbeat frequencies. In *Intelligent Data Engineering and Analytics: Proceedings of the 9th International Conference on Frontiers in Intelligent Computing: Theory and Applications (FICTA 2021)*, pages 71–80. Springer, 2022.
- [17] Myat Su Yin, Peter Haddawy, Borvorntat Nirandmongkol, Tup Kongthaworn, Chanaporn Chaisumritchoke, Akara Supratak, Chaitawat Sangamuang, and Patchara Sriwichai. A lightweight deep learning approach to mosquito classification from wingbeat sounds. In *Proceedings of the conference on information technology for social good*, pages 37–42, 2021.
- [18] Xutong Wei, Md Zakir Hossain, and Khandaker Asif Ahmed. A resnet attention model for classifying mosquitoes from wing-beating sounds. *Scientific Reports*, 12(1):10334, 2022.
- [19] Ankur Singh Bist, Mohd Mursleen, Lalit Mohan, Himanshu Pant, and Purushottam Das. Mosquito detection using deep learning based on acoustics. *Journal of Contemporary Issues in Business and Government Vol*, 27(1), 2021.

- [20] Potamitis, Ilyas, and Ioannis Rigakis. "Large aperture optoelectronic devices to record and time-stamp insects' wingbeats." *IEEE Systems Journal*, vol. 12, no. 4, Dec. 2018, pp. 3599–3608. IEEE, doi:10.1109/JSYST.2017.2788438.
- [21] Diab, Kal, et al. "MACSFeD: An Open Database of Mosquito Acoustic Communication and Swarming Features." *Database: The Journal of Biological Databases and Curation*, vol. 2024, 2024, baae086, Oxford University Press, doi:10.1093/database/baae086.
- [22] Wyze Labs. *Wyze Cam Pan v3*. Wyze, <https://www.wyze.com/products/wyze-cam-pan>. Accessed 11 Aug. 2025.
- [22] Maroš Kollár. (n.d.) PyPDF: A Python PDF Library. Retrieved from <https://pypdf.readthedocs.io/>.
- [23] GOOTOP. GOOTOP Bug Zapper Outdoor, Mosquito Zapper, Electric Fly Zapper, Mosquito Killer, Fly Traps, 3-Prong Plug, Flying Insects Zapper Indoor Outdoor 90-130 V, 4200 V, ABS Plastic Outer (Black). Amazon, <https://www.amazon.com/dp/B09PQF39PG>. Accessed 11 Aug. 2025.
- [23] Ilse, Maximilian, Jakub M. Tomczak, and Max Welling. Attention-based Deep Multiple Instance Learning. *arXiv*, 13 Feb. 2018, arxiv.org/abs/1802.04712.
- [24] Lu, Ming Y., et al. Data-Efficient and Weakly Supervised Computational Pathology on Whole-Slide Images. *arXiv*, 20 Apr. 2020, arxiv.org/abs/2004.09666.
- [25] Li, Bo, et al. Dual-Stream Multiple Instance Learning Network for Whole Slide Image Classification with Self-Supervised Contrastive Learning. *arXiv*, 17 Nov. 2020, arxiv.org/abs/2011.08939.
- [26] Shao, Zhi, et al. TransMIL: Transformer based Correlated Multiple Instance Learning for Whole Slide Image Classification. *arXiv*, 2 June 2021, arxiv.org/abs/2106.00908.
- [27] Ma, Jiayao, et al. "Hierarchical Temporal Multi-Instance Learning for Video-based Student Learning Engagement Assessment." *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21)*, 2021, pp. 2782–2789.
- [28] Wang, Qiang, et al. Revisiting Multiple Instance Neural Networks. *arXiv*, 8 Oct. 2016, arxiv.org/abs/1610.02501.
- [29] Associated Press. "China Reports First-Ever Large Chikungunya Outbreak." *AP News*, 9 Aug. 2025, <https://apnews.com/article/china-chikungunya-virus-outbreak-control-measures-3ee97c21152b05ed0912af646594eafd>.
- [30] Gupta, Danika, and Awani Gadre. Sound-based Mosquito Classification via Featurization and Machine Learning. *Proceedings of the 2024 IEEE 7th International Conference on Big Data and Artificial Intelligence (BDAl)*, Beijing, China, 5–7 July 2024, pp. 1–6. IEEE, 2024.
- [31] Gupta, Danika, and Awani Gadre. "Enhancing Mosquito Identification and Tracking with Object Detection and Citizen-Science Smartphone Imagery." *Applications of Digital Image Processing XLVII*, vol. 13137, 131370D, SPIE, 2024, <https://doi.org/10.1117/12.3028328>.
- [32] librosa development team. "Feature extraction." *librosa 0.11.0 Documentation*, librosa.org, 2025, <https://librosa.org/doc/0.11.0/feature.html>.
- [33] "Sound classification with YAMNet." *TensorFlow Hub Tutorials*, TensorFlow, last updated 9 March 2024, www.tensorflow.org/hub/tutorials/yamnet
- [34] Lewicki, Michael S. "A review of methods for spike sorting: the detection and classification of neural action potentials." *Network: Computation in Neural Systems*, vol. 9, no. 4, 1998, pp. R53–R78.
- [35] Mewett, David T., David N. Reynolds, and George B. N. Milos. "Reducing power line interference in digitised electromyogram recordings by spectrum interpolation." *Medical and Biological Engineering and Computing*, vol. 42, no. 4, 2004, pp. 524–531.
- [36] "Raspberry Pi 4 Model B Specifications." *Raspberry Pi*, Raspberry Pi Ltd., <https://www.raspberrypi.com/products/raspberry-pi-4-model-b/specifications/>. Accessed 19 Aug. 2025.
- [37] "ONNX Concepts." *ONNX 1.20.0 Documentation*, ONNX.ai, 2024, onnx.ai/onnx/intro/concepts.html. Accessed 19 Aug. 2025.
- [38] TensorFlow Models Team. YAMNet: Audio event classification model [GitHub repository]. *GitHub*. Retrieved August 19, 2025, from <https://github.com/tensorflow/models/tree/master/research/audioset/yamnet>
- [39] Chanda, Emmanuel, et al. "Integrated Vector Management: A Critical Strategy for Combating Vector-Borne Diseases in South Sudan." *Malaria Journal*, vol. 12, 25 Oct. 2013, Article No. 369, BioMed Central, DOI: 10.1186/1475-2875-12-369. PubMed Central, PMCID: PMC3816579.

[40] Ministry of Health of the Federal Government of Somalia, Ministry of Health of the Republic of Somaliland, and Ministry of Health of Puntland. Somalia Malaria Programme Performance Review Report. March 2014, Federal Government of Somalia, Puntland Somaliland. Somalia MPR Phase Report FINAL_{020914.pdf}. *Federal Ministry of Health and Human Services, Somalia*.

[41] Yee, Donald A., et al. Robust Network Stability of Mosquitoes and Human Pathogens of Medical Importance. *Parasites Vectors*, vol.15, article no.216, 20 June 2022, <https://doi.org/10.1186/s13071-022-05333-4>. Published 20 June 2022.

[42] Yang, Yuanlin, et al. "Meta Multi-Instance Multi-Label learning by heterogeneous network fusion." *Information Fusion*, vol. 95, 2023, pp. 417-429, doi.org/10.1016/j.inffus.2023.02.017.