Energy Efficient Methods for Mosquito Classification

1st Danika Gupta
The Harker School
San Jose, California USA
dan@gprof.com

2nd Ming Zhao
Arizona State University
Phoenix, Arizona USA

3rd Awani Gadre
Department of Biology
Santa Clara University
Santa Clara, California, USA

***Abstract*—Mosquito-borne diseases impact over 3 billion people globally, causing more than 600,000 deaths each year. Precise mosquito species identification is crucial for outbreak prediction but is limited by manual, subjective methods requiring specialized skills and equipment, hindering scalability. Climate change exacerbates this by altering habitats. Low-cost identification techniques can enable broad deployment in sensors and other inexpensive devices for non-invasive mosquito tracking in remote areas, reducing human disease exposure. Our work investigates low power sound and image methods for non-invasive mosquito type identification. Using two datasets of mosquito sounds, and two datasets of mosquito images, augmented by an additional dataset of insect sounds, we show that it is possible to create a range of sound-based and object-based mosquito detection methods using low-power machine learning techniques.**

***Index Terms*—Mosquito borne diseases, deep learning, machine learning, transfer learning, sound featurization, object detection, spectrograms, mel frequency cepstral coefficients, YOLO, energy efficient YOLO**

I. Introduction

Mosquito-borne diseases, such as malaria and dengue, pose a significant public health challenge. Malaria claims over 600,000 lives annually, while dengue has seen a thirtyfold increase in incidences over the past fifty years [1]. The situation is further exacerbated by climate change, which expands mosquito habitats, increasing the risk of these diseases [2]. Different mosquito species transmit distinct pathogens [3], making species classification crucial. Precise mosquito species identification is crucial for outbreak prediction but is limited by manual, subjective methods requiring specialized skills and equipment, hindering scalability. Countries employ various methods to track mosquitoes, including smartphone apps that allow citizens to report sightings and the creation of detailed distribution maps based on expert-validated data [4]. Field studies also play a crucial role, using techniques such as polymerase chain reaction tests to detect viral genetic material in blood samples [5]. These methods help identify hidden outbreaks and inform targeted control efforts. The integration of low power monitoring is particularly valuable, as it enables broader surveillance efforts, provides real-time data to further mosquito control initiatives.

Leveraging low power methods for mosquito identification and tracking using non-invasive methods has intriguing potential for global mosquito identification and tracking. In particular, using sound creates a further separation between humans and mosquitoes (it is far easier to record sound in a room rather than getting close enough to a mosquito to photograph it), reducing the risk of being bitten. Installing unmanned sensors that can record the sound is also possible, further preventing human infections. We present a comprehensive evaluation of both sound and image based methods for mosquito identification and tracking, focusing on low power machine learning and deep learning techniques.

The paper presents three key contributions: 1) Demonstrating the feasibility of sound-based mosquito classification using energy-efficient featurization techniques with a focus on Mel Frequency Cepstral Coefficients (MFCC); 2) demonstration of the effectiveness of object detection methods in classifying mosquitoes based on real-world imagery with background noise; and 3) Providing six new datasets for future research.

While we offer preliminary results in 7 mosquito variants, the bulk of our work focuses on the classification of three mosquito variants which are described below:

- Anopheles dirus is a mosquito species that is a known vector for several diseases, the most significant of which is malaria. It is considered one of the primary vectors for malaria transmission in Southeast Asia, particularly in forested areas where it is most commonly found. In addition to malaria, Anopheles mosquitoes can potentially carry other diseases such as filariasis, but malaria is the primary concern with Anopheles dirus. The biology, behavior, and distribution of Anopheles dirus, as well as its significance as a primary vector for malaria transmission in the region is available in [6],
- Anopheles farauti is also a known vector for malaria, particularly in the Western Pacific region, including countries like Papua New Guinea, Solomon Islands, and northern Australia. Like Anopheles dirus, it is primarily associated with the transmission of malaria and is considered a major vector in the areas where it is prevalent. More information on this variant can be found in [7] which contributed to the understanding of its implications for malaria transmission.
- Culiseta incidens, commonly known as the cool weather mosquito or fish pond mosquito, is a species of mosquito found in various regions, including western Canada and the United States. While primarily considered a domestic nuisance, Culiseta incidens has the potential for medical

importance as laboratory experiments have shown that it can transmit viruses such as St. Louis Encephalitis, Western Equine Encephalitis, and Japanese B Encephalitis. [8].

- Aedes albopictus, commonly known as the Asian tiger mosquito, is a species characterized by its distinctive black and white striped body and legs. It is a vector for various viral infections, including dengue fever, chikungunya, and Zika virus. The species is highly adaptable and often found in urban environments. It exhibits aggressive daytime biting behavior and is known for its rapid spread across different regions.
- Culex quinquefasciatus, commonly known as the southern house mosquito, is a widespread species that thrives in urban and suburban areas [9]. It is a primary vector for diseases such as West Nile virus, lymphatic filariasis, and encephalitis. The mosquito is known for its preference for breeding in stagnant water with high organic content. It is primarily active during the night, targeting both birds and humans for blood meals.
  All of these mosquitoes were chosen due to their individual impacts on human health and the fact that they co-exist in urban and suburban areas worldwide, making precise classification necessary for human health management [9].

II. Related Work in Mosquito Classification

In recent years, the application of machine learning has shown promise in addressing a variety of challenges across different sectors, including healthcare. One significant area where these algorithms is making an impact is in the management of mosquito-transmitted diseases, which continue to pose a substantial health risk globally [10].

Traditional strategies for controlling these diseases have largely focused on controlling the mosquito vectors. This approach typically requires a considerable number of entomology experts to monitor, identify, and ultimately eradicate targeted mosquito populations. However, this process can be resource-intensive and challenging to manage [11]. Deep learning (DL) based algorithms offers a potential solution to these challenges. Recent studies have highlighted the effectiveness of deep learning algorithms in the rapid detection, identification, monitoring, and control of target mosquito populations using minimal resources [12]. These algorithms can accelerate the pace of operations and facilitate the exploration of data on the ongoing evolutionary status, blood-feeding tendencies, and age grading of mosquitoes. While research has explored many approaches to mosquito mitigation, ranging from outbreak forecasting to intelligent traps, we focus on related work in mosquito identification. Visual methods have been investigated in [13] and [14], while video-based approaches are discussed in [15]. Acoustic and pseudo-acoustic identification techniques are explored in a series of studies by [16], [17], [18], [19], and [20]. Additionally, audio datasets for mosquito research are available in [21] and [22]. We use both of these in our study.

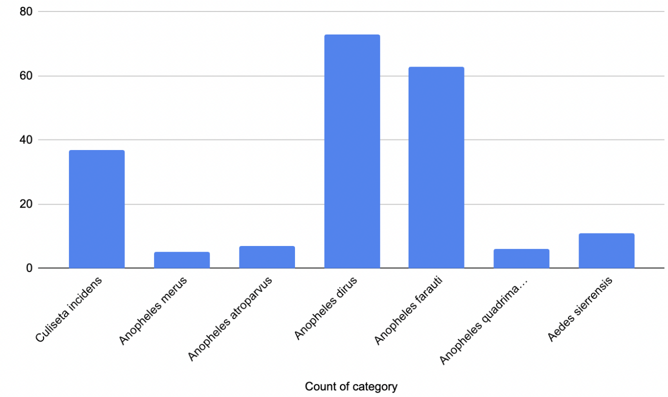


Fig. 1: Dataset Distribution

Our work differs from existing research in that we appear to be the first to (a) use Mel Frequency Cepstral Coefficients (MFCCs)s to featurize the audio and then apply a range of machine learning methods to distinguish mosquito breeds, the first to We are interested in this approach because it is more energy efficient than deep learning approaches, and has shown value in other domains (emotion detection [23], music genre classification [24] and environmental sound classification [25]) (b) the first to demonstrate the efficacy of object detection, and (c) the first to demonstrate the capability of low power variants of machine learning and deep learning algorithms in both contexts.

III. Approach 1: Sound Based Mosquito Classification

We use the dataset in [22] with 202 sound samples of 7 mosquitoes (Culiseta incidens, Anopheles merus, Anopheles atroparvus, Anopheles dirus, Anopheles farauti, Anopheles quadrimaculatus, and Aedes sierrensis). The dataset is heavily skewed, with Culiseta incidens, Anopheles dirus and Anopheles farauti making up the majority of samples (see Figure 1). Given this, in phase one, all 7 categories were included. In phase two the best represented three categories were included. In the third phase, the two Anopheles categories were merged resulting in two categories. As noted above, Anopheles carries malaria, representing significant death, rendering distinction between Anopheles and Culiseta valuable [26].

Initially, the dataset undergoes a transformation into numerical features through three distinct approaches.

- **MFCC Sound featurization**: In the first approach, audio files are processed using the Librosa library [27], resulting in feature vectors that include Mel Frequency Cepstral Coefficients (MFCCs) [28], tempo, beats, Chroma Short-Time Fourier Transform (STFT), Root Mean Square Energy (RMSE, which measures the signal's power or loudness), spectral centroid, bandwidth, rolloff, and zero-crossing rate. These features capture various characteristics of the audio signal, such as its rhythm, timbre, and pitch.

- **Spectrogram Conversion for Visual Analysis**: The second approach involves converting audio signals into spectrograms, which visually represent the frequency spectrum and intensities over time [29].
- **Spectrogram Featurization with CNN**: The third approach involves converting the spectrograms into vectors through image featurization using a pre-trained Convolutional Neural Network (CNN), specifically MobileNetV2.

Detailed descriptions of each featurization method are provided below, outlining the process of transforming audio signals into informative numerical features for subsequent analysis.

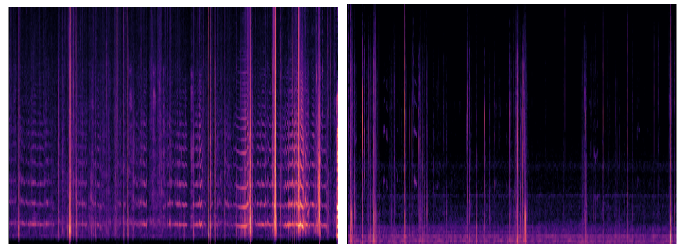
A. MFCC Sound Featurization

MFCC feature extraction (using [27]), generated the following features

- Tempo: The speed or pace of a piece of music, usually measured in beats per minute (BPM).
- Beats: Time points in the audio signal where a beat occurs, typically corresponding to the rhythmic pulse of the music.
- Chroma_STFT: A representation of the energy distribution across 12 different pitch classes (chroma) within an octave, using the Short-Time Fourier Transform (STFT).
- RMSE (Root Mean Square Energy): A measure of the energy of the audio signal, calculated as the square root of the average squared amplitude values.
- Spectral Centroid: Indicates the "center of mass" of the spectrum, often associated with the perceived brightness of a sound.
- Spectral Bandwidth: Describes the width of the spectral band, reflecting the spread of frequencies present in the sound.
- Rolloff: The frequency below which a certain percentage (e.g., 85%) of the total spectral energy is contained.
- Zero Crossing Rate: The rate at which the audio signal changes sign, often used to measure the noisiness or the percussiveness of a sound.
- MFCC1 - MFCC20: Mel Frequency Cepstral Coefficients, a set of features that represent the short-term power spectrum of a sound, often used in speech and audio processing. Each MFCC (from 1 to 20) captures different characteristics of the sound, with lower coefficients typically representing basic spectral shape and higher coefficients capturing finer details.

B. Spectrograms

1) Visual Spectrograms: A spectrogram is a time-frequency representation of a signal, displaying the amplitude of frequencies over time. It is depicted as a 2D plot with time on the horizontal axis, frequency on the vertical axis, and amplitude represented by color intensity. This visualization enables the analysis of spectral content, identification of dominant frequencies, and detection of temporal frequency changes and patterns. In our case, each spectrogram image was generated from a single audio file (using [27]) and classified via standard convolutional neural networks (MobileNetV2 pretrained with ImageNet). Example spectrograms for each category are shown in 2.



(a) Culex Incidens (b) Anopheles

Fig. 2: Example spectrogram of sounds from the two categories of mosquitoes.

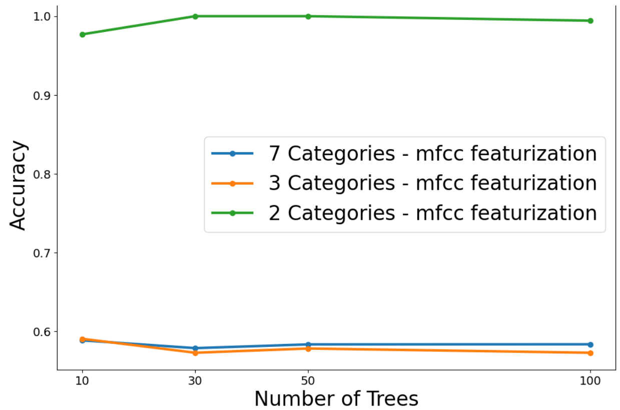
2) Featurized Spectrograms: In a further step - the spectrogram images generated above were featurized using a Convolutional Neural Network (CNN) and the subsequent feature vectors used to train traditional machine learning classifiers. For consistency, MobileNetV2 was used for teh featurization process as well. The MobileNetV2 was pretrained on ImageNet to learn a rich set of features for image representation. When used for feature extraction, the final classification layer was removed and the output from the penultimate layer is taken as the feature vector for each input image. This feature vector captures the essential characteristics of the image and can be used as input to conventional machine learning algorithms for classification tasks. This approach leverages the power of deep learning for feature extraction while utilizing traditional machine learning techniques for the final classification, offering a balance between performance and computational efficiency.

C. Algorithms, Hyper-Parameters, and Metrics

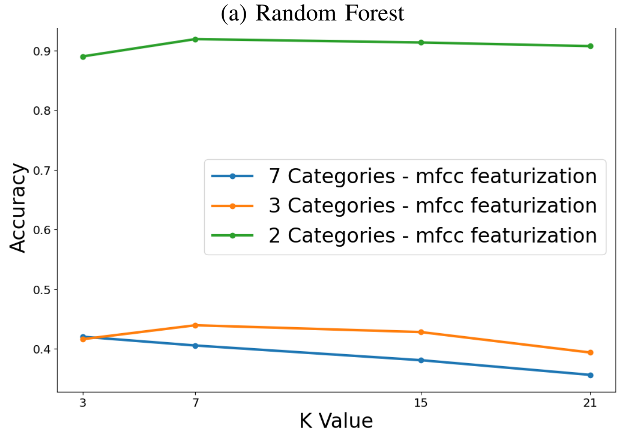
Random Forest and K-Nearest Neighbors (KNN) were used with the numerical feature vectors (with first and third featurization approach). Number of estimators hyper-paramter was varied between 10-100 for Random Forest and K was varied between 3-21 for KNN. In the second approach, a MobileNetV2 model is trained using the spectrogram dataset, tuned by varying learning rate between 0.05-0.0001 in two different scenarios where the values of epochs were 20 and 50. MobileNetV2 was chosen because of its lightweight nature - matching our goals for low-cost solutions. All results are measured via 20% of the dataset reserved for validation, with 5-fold cross validation. All models are assessed via Accuracy and Confusion Matrix, with the Confusion Matrix providing insight into the classification success of each category.

D. Results

Figures 3a and 3b show averaged accuracies for MFCC featurization, indicating that Random Forest outperforms KNN, and that the two category experiment notably outperforms the others. The 7 and 3 category experiments perform similarly because of the heavy dataset skew - the additional categories

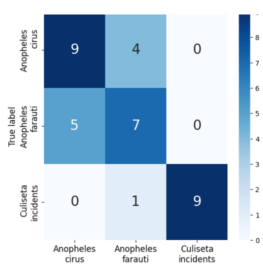


(a) Random Forest

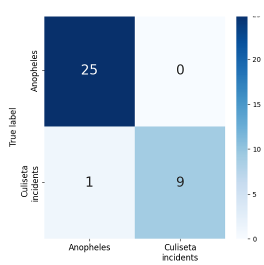


(b) KNN

Fig. 3: Results - Random Forest and KNN for MFCC Featurization.



(a) 3 Category Confusion Matrix - MFCC Featurization with Random Forest



(b) 2 Category Confusion Matrix - MFCC Featurization with Random Forest

Fig. 4: Confusion matrix with 3 and 2 category experiments.

in the 7 category experiment do not contribute substantially to the accuracy given their limited representation. Further detail shown in 4a and 4b show that the three category model struggles to distinguish between Anopheles Dirus and Anopheles Farauti, while distinction with Culiseta Incidens is well achieved. As a result, when the Anopheles breeds are merged into one category, classification performance improves dramatically.

Figure 5a shows results for spectrogram classification (with learning rate and epochs hyper-parameter tuning) for the 3 category and 2 category problem. Figures 5c and 5b show the results for the featured spectrograms under Random Forest and KNN. The confusion matrices for these two approaches, omitted for brevity, indicated the same issue with the 3 category problem - notably the struggle to distinguish between Anopheles breeds.

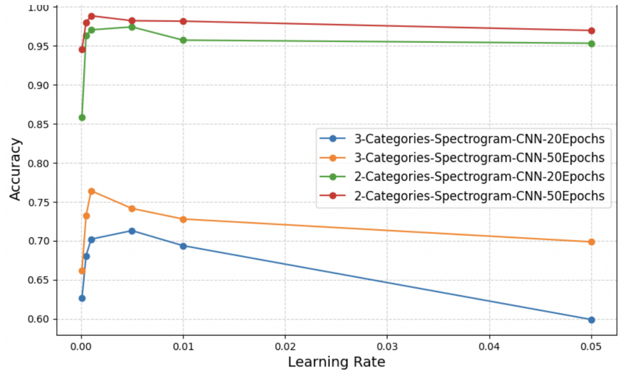
E. Comparison

Comparisons of the best results of all featurization experiments (Figure 6) shows that while spectrograms with CNN outperform MFCC featurization with Random Forest in 3 categories, the MFCC featurization outperfoms all others in 2 categories by 1-2%. All three approaches encountered difficulties in accurately distinguishing between the two Anopheles breeds, as evidenced by the confusion matrices presented in Figures 4a and 4b. This challenge may be attributed to the limited size of the dataset. However, the results indicate that distinguishing between Anopheles and Culiseta is feasible with the methods employed.

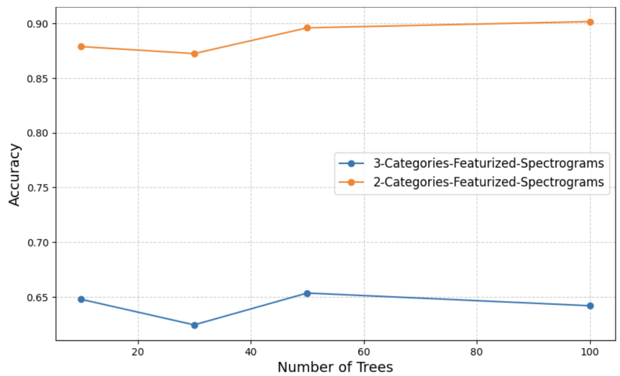
IV. Approach 2: Sound Based Mosquito vs Background Classification

Base on our observations in Approach 2, we continue our work with MFCC based featurization, and next attempt to distinguish between mosquitoes and background sounds. For this we use the dataset in [21] and create two categories, one with mosquitoes and one with background only. Figure 7 shows the results for Random Forest. When the whole files are used. Even though the dataset is heavily skewed - the accuracies reported by all three algorithms was very good (Random forest classifier - Accuracy across both classes - 98.44% MLP Classifier - 97.32%, KNN Classifier - 96.35%). Next we examined whether the same effectiveness would hold if only a small sample of the sound was used (rather than the whole file which in many cases was several seconds long). The rationale for this test is that we may want to use an audio detector to detect the presence of mosquitos and turn on more power hungry monitoring methods (like camera) if there are mosquitoes present. To enable such a strategy, we would need to detect mosquitoes from a very small audio sample. We tried 0.1s, 0.5s, and 1.0s. To handle the imbalance, we took random samples from different start points in the background files and combined them with the start of any mosquito file (since our goal is to detect mosquitoes quickly). This led to a balanced dataset between background and mosquito. The results were 1s snippets, accuracy 93.32%, 0.5s snippets Accuracy 93.55% and 0.5s snippets Accuracy 92.71%, indicating that it is likely feasible to use this detector to dynamically control other measurement devices with higher power requirements.

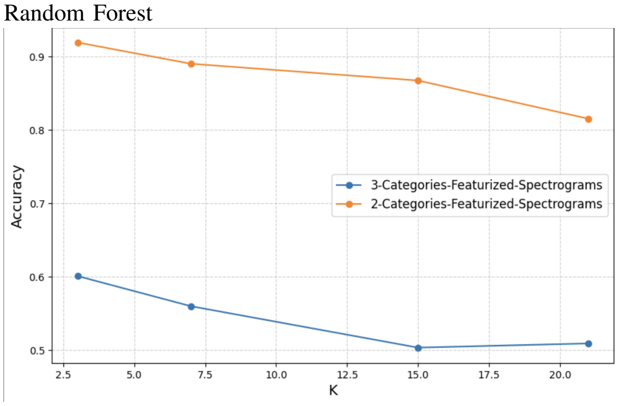
To further reinforce the ability to distinguish mosquito sounds from other sounds, We tried the Insect1000 dataset. We downloaded about 6000 sounds of insects from the dataset in [30] (all 12 represented categories) and merged it with the background sounds from the Humbug DB dataset - to see if the model can distinguish mosquitoes from background when



(a) Results from Mel-Spectrogram Classification via MobileNetV2



(b) Results from Featurized Mel-Spectrogram Classification via Random Forest



(c) Results from Featurized Mel-Spectrogram Classification via KNN

Fig. 5: Results from Featurized Mel-Spectrogram Classification for different algorithms

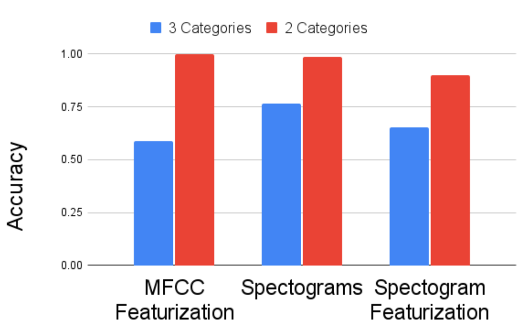


Fig. 6: Test Results for best model performance from each method

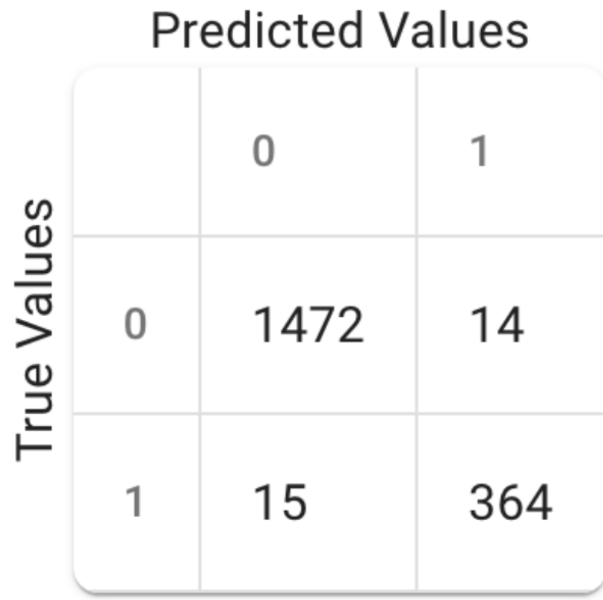


Fig. 7: Mosquito vs Background Detection

there were insect noises in the background. It could. We were able to demonstrate accuracy of almost 99%.

V. Approach 3: Object Based Visual Mosquito Detection

We begin with two datasets: one large dataset of mosquito images gathered via a range of citizen science efforts and the second with a lab-acquired dataset. In both cases, we use DINO [31] to label the images for mosquitoes and manually hand-checked to correct issues. Next, we explored YOLOv8-based object classification for both datasets via several hyper-parameter tuning options. Finally, we tested the best version of models trained on each dataset on the test data for the other dataset. To provide a baseline to compare these results, we repeat the same process with ResNet101 based image classification of the original images, coupled with the same cross-testing of the models trained on one dataset with the test of the other dataset.

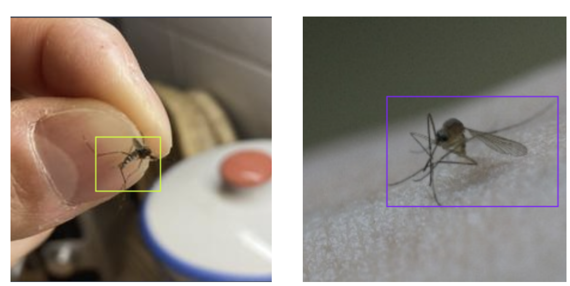


Fig. 8: Images from the Mosquito Alert Dataset (first) and the Mosquitoes on Human Skin Dataset (second)

The first dataset, referred to as the Mosquito Alert dataset, consists of approximately 10,000 images of mosquitoes gathered through various citizen science initiatives [32]. Mosquito Alert engages volunteers to capture and submit mosquito photographs. These contributors come from a diverse geographical span, including Spain, the Netherlands, Italy, and Hungary, and have contributed to the dataset over a period from 2014 to 2022. These images were captured in natural environments using different devices via the Mosquito Alert app, encouraging community participation in mosquito monitoring efforts. This dataset is particularly valuable for training models operating in diverse real-world conditions. We selected two mosquito species: Aedes albopictus and Culex quinquefasciatus for our experiments.

The second dataset, Mosquito on Human Skin (MOHS) [33] was created in a controlled laboratory setting and contains 1,500 images of three mosquito species—Aedes aegypti, Aedes albopictus, and Culex quinquefasciatus—in both normal and smashed states. However, for the purposes of this study, only images of Aedes albopictus and Culex quinquefasciatus in their normal state were used to maintain consistency with the first dataset. Figure 8 shows sample images, all of which are of mosquitoes either held between human fingers or on a human hand.

Both datasets were initially without bounding boxes, which are essential for training object detection models. To address this, the DINO (Self-Distillation with No Labels) method was employed to automatically generate bounding boxes around the mosquitoes. DINO, a self-supervised learning technique, allows models to learn visual representations without needing human-provided labels, making it particularly effective in scenarios with limited labeled data [31].

Following the automated generation of bounding boxes, each image was manually inspected to ensure annotation accuracy, providing high-quality training data for subsequent experiments. The data was split into train, validation, and test to conduct experiments moving forward.

A. *Algorithms, Tuning and Metrics*

For the task of mosquito identification and classification, we employed the YOLOv8 (You Only Look Once) object detection model. YOLO is a real-time object detection framework that treats object detection as a single regression problem. It simultaneously predicts bounding boxes and class probabilities for objects within an image during a single evaluation, offering a unified and efficient approach to object detection [34]. YOLOv8 further refines this architecture with enhancements in feature extraction, improved backbone networks, and optimized anchor-free detection heads, making it particularly effective for tasks requiring high precision and speed [35].

During training, we fine-tuned the YOLOv8 model by experimenting with different learning rates (ranging from 0.0001 to 0.05) and varying the number of epochs (5 and 10). The performance of each configuration was evaluated using a validation set, comprising 20% of the total data.

As a baseline comparison, we also utilized ResNet101, a state-of-the-art convolutional neural network (CNN) known for its strong performance in image classification tasks [36]. Similar to YOLOv8, ResNet101 was subjected to hyperparameter tuning, exploring learning rates between 0.0001 and 0.05, and trained over 10 epochs to ensure robust comparison across models.

The object detection results were measured via mAP50 and confusion matrix. mAP50 (mean Average Precision at 50% Intersection over Union) is a performance metric for object detection models that measures the average precision when predicted bounding boxes have at least a 50% overlap with ground truth boxes [37]. The precision/recall curves for the best-performing models under mAP50 are also reported in the paper.

The baseline CNN results are reported via accuracy. Cross-testing is also performed across the datasets where the best-performing model trained on each dataset is tested against the other dataset.

B. *Results*

Table I shows the mAP50 results for each experiment. For both datasets, high mAP50 values can be achieved, particularly at low learning rates such as 0.0001. Figure 9 shows the precision-recall curves for the best-performing models, indicating that both classes in the validation set are performing well. Additional experiments were run with 10 epochs, which delivered similar results.

TABLE I: mAP50 values for MA and MOHS dataset experiments

Learning Rate	MA mAP50	MOHS mAP50
0.0001	0.993	0.978
0.0005	0.987	0.979
0.001	0.99	0.981
0.005	0.877	0.381
0.01	0.601	0
0.05	0.505	0.051

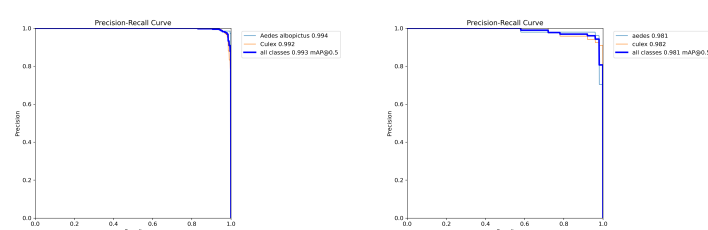


Fig. 9: Precision-Recall curves for best performance models (a) MosquitoAlert dataset (b) Mosquito on Human Skin dataset

Next, we cross-tested the models, validating the best model trained on Mosquito Alert with the validation dataset from MOHS and vice versa. The validation dataset from MOHS

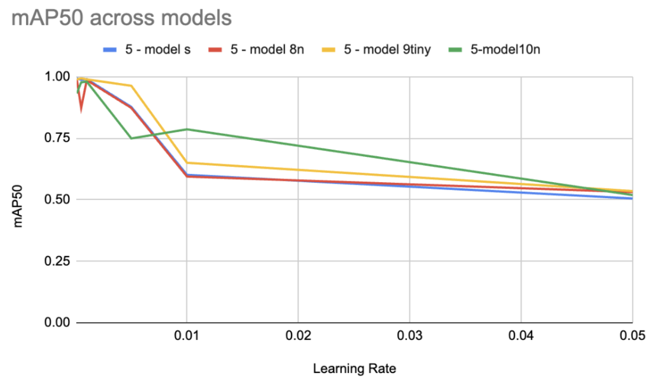


Fig. 10: Mosquito vs Background Detection

delivered a mAP50 of 0.977 on the best model from Mosquito Alert, and the validation dataset from Mosquito Alert delivered a mAP50 of 0.985 on the best model trained on MOHS. The latter result is significant because lab-generated datasets are much easier to gather and train from than datasets from the wild. This result suggests that effective mosquito classification for real-world use cases can be created from lab data when combined with effective object detection techniques.

The baseline results (without object detection) showed that the Mosquito Alert (MA) dataset achieved a best-case accuracy of 87.1% at a learning rate of 0.05. In contrast, the MOHS dataset achieved perfect validation accuracy at several learning rate settings. While mAP50 and accuracy cannot be directly compared, it is worth noting that the performance of the MosquitoAlert model relative to the MOHS model is lower for CNN than it was for object detection - suggesting that the background noise in the real-world images may be presenting a challenge. Furthermore, the ResNet101 model trained on MOHS was not able to perform well for the Mosquito Alert dataset, achieving a best-case accuracy of only 68.9%. The CNN model trained on Mosquito Alert was able to perform well on the MOHS dataset, achieving an accuracy of 98.2%.

VI. Approach 4: Object Based Visual Mosquito Detection

Once the viability of the object detection method was established, we explored low power object detection methods using the same two datasets created in approach 3. Figure 10 shows the results for the larger Mosquito Alert dataset. The results of the power efficient YOLO models is nearly identical to that of the standard models.

VII. Discussion

Overall, the study demonstrates that sound featurization coupled with machine learning can effectively distinguish some mosquito breeds, with MFCC featurization showing promise in both mosquito classification and distinguishing between mosquitoes and background noise. The results also demonstrate the efficacy of the object detection approach. In particular, the cross-testing across datasets revealed that the YOLOv8 model trained on MOHS (a much smaller dataset acquired under controlled circumstances, was capable of classifying mosquito images gathered in much more varied circumstances to be expected in real-world scenarios. Comparison with the CNN baselines shows that, while the easier task of classifying lab-captured mosquito pictures can be done with either approach, the object detection approach does better in scenarios with real-world complexities. Low power variants of YOLO also showed significant promise.

VIII. Future Work

Our work indicates that sound featurization coupled with machine learning can distinguish some mosquito breeds and, as importantly, that energy efficient low cost techniques can perform mosquito detection and classification very effectively. Further work is needed to determine whether sub-breeds can be classified effectively in a similar manner. To this end, our future work is to evaluate more algorithms, dataset augmentation, and add modalities. We would also like to explore whether explainability can assist our understanding of mosquito classification. All our code, and the featurized datasets, will be open sourced to assist other researchers in building upon our work (4 datasets total). Similarly, low-power object detection shows promise. Potential issues in our work include the need to manually inspect the bounding boxes generated by DINO, which can be error-prone in large datasets. It would be ideal to automate this in some fashion, particularly to check mis-classification for evidence of poor bounding box labeling. This is an area for our future work. We are also creating a physical implementation of our algorithms in Raspberry Pi 4 and Raspberry Pi Zero-W and conducting detailed power consumption experiments. Our goal is to combine our findings into a low power sensor that can be easily installed in remote areas for accurate and non-invasive mosquito pattern assessment.

IX. Acknowledgements

We would like to thank Professor Renu Wickramasinghe, of the University of Sri Jayewardenapura, for her assistance during our project.

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, multi-scenario 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/disease-vectors/surveillance-and-disease-data/mosquito-maps*.

[5] Direct pcr of indigenous and invasive mosquito species: a time- and cost-effective technique of mosquito barcoding. *Med Vet Intermol*, 2016.

[6] Thin Thin Oo, Volker Storch, and Norbert Becker. Anopheles dirus and its role in malaria transmission in myanmar. *Journal of Vector Ecology*, 2003.

[7] J.H. Bryan and M. Coluzzi. Cytogenetic observations on anopheles farauti laveran. *Bulletin of the World Health Organization*, 1971.

[8] Fish pond mosquito (culiseta incidens). *Napa county mosquito abatement district Fish Pond Mosquito (Culiseta incidens)*, Accessed Feb 25, 2024.

[9] et al. Lopez-Solis, Alma D. Aedes aegypti, ae. albopictus and culex quinquefasciatus adults found coexisting in urban and semiurban dwellings of southern chiapas, mexico. *Insects*, 14:565, 2023.

[10] 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.

[11] Handi Dahmana and Oleg Mediannikov. Mosquito-borne diseases emergence/resurgence and how to effectively control it biologically. *Pathogens*, 9(4):310, 2020.

[12] 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.

[13] 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.

[14] Daniel Motta, Alex Álisson 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.

[15] 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.

[16] 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.

[17] Ayush Jhaveri, KS Sangwan, Vinod Maan, and Dhiraj. Deep learning-based 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] Nikhil Kopparaju, Drutesh Kattubadi, Mayank Manchuri, Abhishekand Muthyala, and Hari Kiran Bikki. Speech emotion recognition using librosa. *International Journal of Scientific Development and Research*, 2023.

[24] Aryan Raj Chhetri, Kshitiz Kumar, Mayank Prashanth Muthyala, MR Shreyas, and Raghu A Bangalore. Carnatic music identification of melakarta ragas through machine and deep learning using audio signal processing. In *2023 4th International Conference for Emerging Technology (INCET)*, pages 1–5. IEEE, 2023.

[25] Wenjie Mu, Bo Yin, Xianqing Huang, Jiali Xu, and Zehua Du. Environmental sound classification using temporal-frequency attention based convolutional neural network. *Scientific Reports*, 11(1):21552, 2021.

[26] Baylor. Mosquito-borne diseases. *Baylor College of Medicine*, 2017.

[27] Librosa. librosa — librosa 0.8.0 documentation. (n.d.). *Librosa.org*.

[28] Mel Frequency Cepstral Coefficients. Mel frequency cepstral coefficients (n.d.). *https://medium.com/@derutycsl/intuitive-understanding-of-mfccs-836d36a1f779*.

[29] audiospectrograms. Audio spectrograms (n.d.). *https://www.izotope.com/en/learn/understanding-spectrograms.html*.

[30] J. Branding, D. von Hörsten, E. Böckmann, J. K. Wegener, and E. Hartung. Insectsound1000: An insect sound dataset for deep learning based acoustic insect recognition. *Scientific Data*, 11:475, 2024.

[31] et al. Caron, Mathilde. Emerging properties in self-supervised vision transformers. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

[32] 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-col

[33] Song-Quan Ong and Hamdan Ahmad. An annotated image dataset for training mosquito species recognition system on human skin. *Scientific Data*, 9(413), 2022.

[34] et al.. Redmon, Joseph. You only look once: Unified, real-time object detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[35] Ultralytics. Ultralytics YOLO Docs. docs.ultralytics.com/. . (Accessed: July 2024).

[36] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 770–778, 2016.

[37] KDNuggets. Metrics to Use to Evaluate Deep Learning Object Detectors. www.kdnuggets.com/2020/08/metrics-evaluate-deep-learning-object-detectors.html. (Accessed: July 2024).