Acoustic Identification of Aedes aegypti: Experimental Assessment of Different Machine Learning Techniques

Kayuã Oleques Paim[1], Thaís Marcelle Dihl da Silva[1], Jaqueline Dilly[1], Anderson Rocha Tavares[1], Marielton dos Passos Cunha[4], Onilda Santos da Silva[2], Mariana Recamonde-Mendoza[1], Rodrigo Brandão Mansilha[3], Weverton Cordeiro[1]

[1] Instituto de Informática, Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre-RS, Brazil

[2]Instituto de Ciências Básicas da Saúde, Departamento de Microbiologia,
Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre-RS, Brazil

[3]Programa de Pós-Graduação em Engenharia de Software (PPGES),
Universidade Federal do Pampa (UNIPAMPA), Alegrete-RS, Brazil

[4]Instituto de Computação, Universidade Estadual de Campinas (UNICAMP),
Campinas-SP, Brazil

Abstract. *Mosquito-borne diseases affect over 700 million people annually, demanding effective and scalable monitoring solutions. Traditional traps, while reliable, are costly and have limited coverage. Recent studies have shown that mosquito wingbeat audio can enable species identification using convolutional neural networks, but alternative machine learning (ML) models remain underexplored. This study compares five ML architectures – Residual Network, Multilayer Perceptron (MLP), Long Short-Term Memory (LSTM), Conformer, and Audio Spectrogram Transformer (AST) – for classifying Aedes aegypti from audio data. Using public datasets and standardized preprocessing, we evaluate performance through cross-validation and multiple classification metrics. All architectures analyzed achieved accuracy and F1-scores above 88%, with particular emphasis on MLP and LSTM which, despite their simpler structures, showed competitive performance compared to other networks. These findings support the development of low-cost, audio-based mosquito monitoring systems.*

1. Introduction

Mosquito-borne diseases represent one of the greatest global health challenges, affecting approximately 700 million people and causing over one million deaths annually [Qureshi 2018]. These issues lead to estimated global economic losses in the billions of dollars each year [Ahmed et al. 2022]. *Aedes aegypti*, a key vector, transmits pathogens responsible for diseases such as Dengue, Zika, Chikungunya, and Yellow Fever. Effective disease control requires managing mosquito populations. Strategies such as source elimination [Forsyth et al. 2020] and mosquito traps [Leandro et al. 2022] have proven effective at a local scale but face scalability challenges due to socioeconomic constraints. Monitoring mosquito populations enables targeted resource allocation and optimization of high-cost strategies, such as the release of genetically modified sterile mosquitoes [Waltz et al. 2021]. Thus, accurate and comprehensive monitoring is essential for effective mosquito control and potential regional eradication.

Monitoring mosquito proliferation is challenging, regardless of the technology used [Fernandes et al. 2021]. Many cities often adopt strategies based on mosquito trap deployment within a given area [Leandro et al. 2022], for example the Mosquito Trap Sites initiative in Washington, D.C.[1]. In this context, each trap is visited weekly to identify how many mosquitoes were captured in that period, which are then taken to the laboratory for PCR exams to identity the mosquito species, genre, etc. Although extremely effective, such strategy pose a high acquisition and maintenance costs [Townson et al. 2005]. Complementary monitoring methods based on crowdsourcing, such as using smartphones, offer a viable alternative [Mukundarajan et al. 2017, Fernandes et al. 2021].

In previous work, [Fernandes et al. 2021] proposed using smartphone applications in combination with convolutional neural networks (CNNs) to classify mosquito wingbeat audio. Their results provided evidence that CNNs can achieve high accuracy without relying on additional metadata, though classification performance is still constrained by the complexity of the task. In spite of the advances achieved, the effectiveness of mosquito audio classification using other Machine Learning (ML) techniques remains to be explored in the literature. For example, advancements in audio detection models, like the Audio Spectrogram Transformer [Gong et al. 2021], have shown promising results in capturing complex acoustic features, which may also enhance mosquito audio classification. Therefore, the research question we pose in this paper is: "*What are the potentialities and limitations of other popular ML techniques in identifying Ae. aegypti mosquitoes based on mosquito wingbeat audio recordings?*" An answer to this question might provide valuable research directions to practitioners in the field in devising smartphone apps or intelligent mosquito traps for real-time detection of *Ae. aegypti* mosquitoes.

To fill in this gap in the literature, we propose in this paper a systematic analysis of different ML techniques for mosquito audio classification, presenting a comparative evaluation of prominent approaches. We carry out our analysis considering raw labeled audio datasets publicly available in the literature and the following algorithms: Audio Spectogram Transformer (AST), Long Short-Term Memory (LSTM), Multilayer Perceptron (MLP), Residual Model, and Conformer.

The remainder of this paper is organized as follows. In Section 2 we review the technical background and related literature. In Section 3 we describe the dataset used in our research and methodology for pre-processing the dataset. In Section 4 we outline the experiments and results achieved. We then close the paper in Section 5 with concluding remarks and prospective directions for future research.

2. Background and Related Work

Research on automated mosquito identification has long explored features like acoustic signals and visual imagery. Early studies date back to 1945 [Kahn et al. 1945], using wingbeat frequency recordings to identify species in West Africa. However, technological limitations confined data collection to controlled environments, reducing the generalizability of the models [Chen et al. 2014]. To overcome these constraints, researchers have explored alternatives such as photosensors, optimized trap designs, and computer vision techniques [Li et al. 2005, Ouyang et al. 2015,

[1] `https://opendata.dc.gov/datasets/DCGIS::mosquito-trap-sites/about`

Kim et al. 2021]. For instance, [Li et al. 2005] achieved 72.67% accuracy in classifying five species using photosensors and neural networks, while [Ouyang et al. 2015] reported 85.4% accuracy for three species with infrared sensors and Gaussian mixture models. Other studies focused on enhancing recording conditions using acoustic and visual stimuli [Johnson and Ritchie 2016, Balestrino et al. 2016]. [Johnson and Ritchie 2016] attracted female mosquitoes with male flight tones and GAT traps, and [Balestrino et al. 2016] showed that *Aedes albopictus* responded strongly to sound stimuli and preferred dark-colored traps.

In the domain of computer vision, CNNs have been widely adopted for mosquito recognition and tracking [Motta et al. 2019, Adhane et al. 2022]. For example, [Motta et al. 2019] employed CNNs to identify species such as *Aedes aegypti* and *Aedes albopictus*. Although recent technological developments have revitalized interest in integrating acoustic sensors into mosquito traps[Vasconcelos et al. 2019], the high costs associated with large-scale deployment remain a significant barrier. As a more accessible alternative, smartphones have been shown to be effective tools for capturing wingbeat sounds [Mukundarajan et al. 2017]. In parallel, progress has been made in the field of audio-based ML. Finally, neural network-based models have also been applied to mosquito sound classification using smartphone recordings, with [Fernandes et al. 2021] reporting classification accuracies of 78.12% for multiclass and 94.5% for binary classification tasks. Subsequent research focused on further enhancing classification performance through the use of deep neural networks [Kiskin et al. 2020, Wei et al. 2022], as well as by investigating methods for model compression and computational efficiency [Yin et al. 2021, Su Yin et al. 2022].

In summary, there has been substantial research on the use of neural networks for mosquito identification. However, to the best of our knowledge, no publicly available tool currently exists that can perform real-time mosquito identification "on the fly" based on wingbeat sounds, analogous to how applications such as Shazam)[2]) identify music. In this paper, we posit that a key challenge yet to be fully addressed by the research community is the systematic evaluation of alternative neural network architectures for this task. In the following sections we explore this research problem in greater depth.

3. Materials and Methods

This section provides a detailed description of the datasets utilized in this study, the corresponding pre-processing procedures, and the models selected for experimental evaluation.

3.1. Datasets

The audio data employed in this study are categorized into three primary datasets, as summarized in Table 1 and discussed next:

No Mosquitoes (RD1): This dataset consists of audio recordings captured in environments devoid of mosquito activity. It serves to train the neural network to distinguish mosquito sounds from background noise and to mitigate potential bias. RD1 includes 2,000 samples selected from the ESC-50 dataset [Piczak 2015], which comprises a wide range of environmental audio events.

[2]Shazam: Name songs in seconds: `https://www.shazam.com/`

Table 1. Raw audio datasets. NA: *Not Available*, F: *Female*, M: *Male*. Dataset sources: [Piczak 2015, Mukundarajan et al. 2017, Paim et al. 2024]

Subset	ID	Species	Gender	Number of Mosquitoes per recording	Number of Recordings	Recording Total Length(s)
Dataset from Piczak (2015) [Piczak 2015]	RD1	No mosquitoes	NA	0	2,000	10,000
Dataset from Mukundarajan et al. (2017) [Mukundarajan et al. 2017]	RD2	Ae. albopictus	F/M	1	22	1,736
	RD3	Non-Aedes	F/M	1	7	966
	RD4		NA	1	871	13,237
Dataset from Paim et al. (2024) [Paim et al. 2024]	RD5		F	1	192	4,607
	RD6	Ae. albopictus	M	1	345	8,279
	RD7	Ae. albopictus	F	1	19	570
	RD8	Non-Aedes	M	1	17	510
	RD9		NA	0	158	4,740

Various Mosquitoes (RD2 to RD4): These datasets are sourced from the Abuzz project [Mukundarajan et al. 2017], comprising 900 smartphone-recorded audio samples from 23 distinct mosquito species. The recordings span a variety of conditions, including both field and laboratory environments, and capture mosquitoes in free flight and under controlled experimental settings.

Specific Mosquitoes (RD5 to RD9): These recordings were specifically collected by [Paim et al. 2024] and encompass various genera, species, and sexes of mosquitoes. Data were acquired using an LG K10 smartphone, according to the authors. Recordings took place within enclosures of 5–10 m^3 under controlled lighting and environmental conditions to minimize background interference.

In our experiments, we considered a dataset comprising four distinct classes: female *Aedes* mosquitoes, male *Aedes* mosquitoes, non-*Aedes* mosquito species, and environmental noise. For analytical purposes, the first two categories—female and male *Aedes* mosquitoes—were grouped as positive samples, while the latter two—non-*Aedes* mosquito species and environmental noise—were treated as negative samples. This dataset, derived from prior studies across various research groups, includes 537 positive samples (12.886 seconds) and 2,158 negative samples (14.740 seconds), offering a challenging and diverse testbed for evaluating the generalization capabilities of the proposed models. The inclusion of heterogeneous acoustic conditions in Dataset D6 {RD1-RD5-RD6-RD9} facilitates a rigorous assessment of model robustness in real-world audio classification scenarios.

3.2. Machine Learning Algorithms

In our research, we considered the following algorithms: Residual Model, MLP, LSTM, AST, and Conformer Model. We discuss each one next, including the reference architecture we used in our evaluation.

The **Residual Model**, shown in Figure 1, has been proposed for audio classification tasks, in which input data—such as audio frames or spectrogram images—is first segmented using a sliding time window and subsequently normalized [Paim et al. 2024]. The model begins with a two-dimensional convolutional layer that extracts local features, followed by a series of residual blocks. These residual blocks deepen the network and help mitigate the vanishing gradient problem, thereby improving training efficiency in deeper architectures. After feature extraction, a max pooling layer reduces the spatial dimensionality, and a dense (fully connected) layer performs the final classification, producing

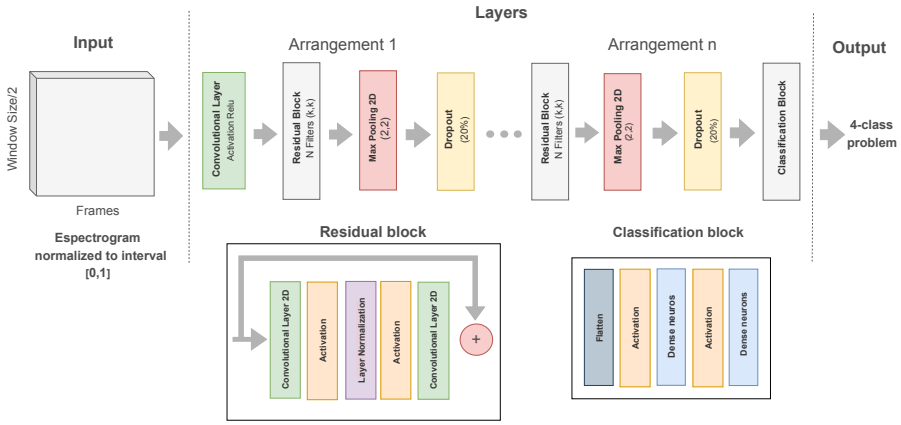


Figure 1. Reference architecture considered for the Residual Model.

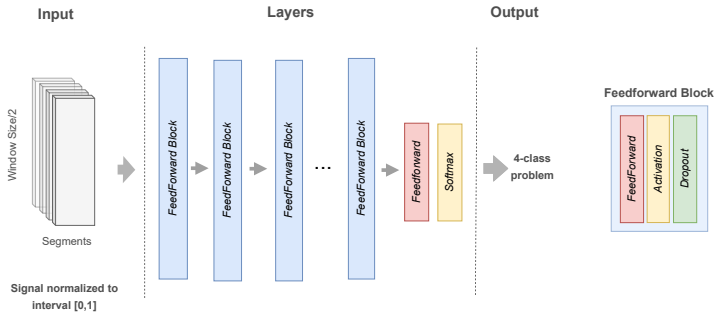


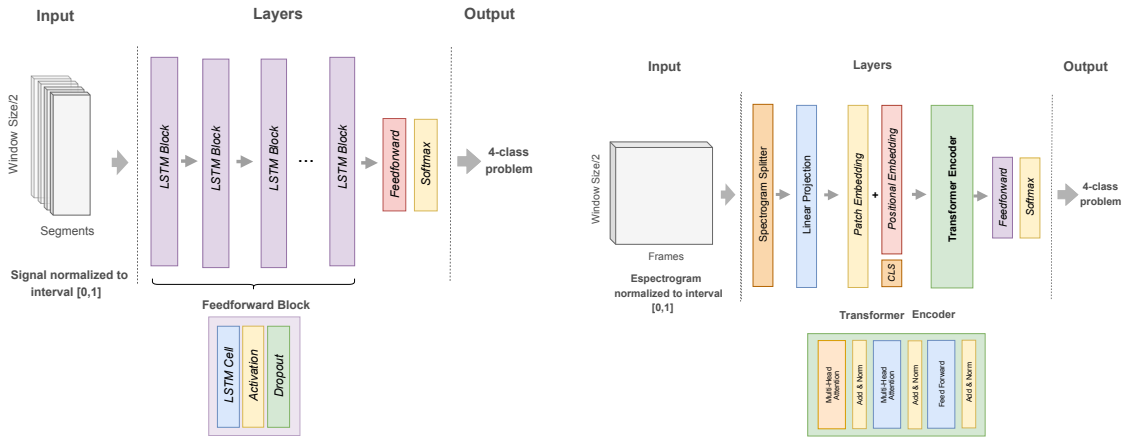
Figure 2. Reference architecture considered for the Multilayer Perceptron model.

a probability distribution over the target classes.

Multilayer Perceptron (MLP) represents a foundational architecture in ML. It is composed of an input layer, one or more hidden layers, and an output layer, forming a fully connected network. Each unit, or neuron, receives input signals, applies a weighted sum and non-linear transformation, and transmits the result forward through the network. The hidden layers are where most of the computation takes place, enabling the MLP to model complex, non-linear relationships in the data. The connections between neurons are characterized by weights, which are adjusted during training through backpropagation to minimize prediction error. Figure 2 presents an overview of the model.

Long Short-Term Memory (LSTM) is widely employed for sequential data processing, including tasks such as language modeling and time series analysis [Hochreiter and Schmidhuber 1997]. LSTM cells are designed with internal gating mechanisms—input, forget, and output gates—that regulate information flow and allow the network to capture long-term dependencies. As input sequences are processed, each cell updates its internal state and transmits output to subsequent cells, maintaining a dynamic memory across time steps. This structure enables LSTMs to retain relevant information while filtering out noise or less significant data. Figure 3a provides a view of the model.

The **Audio Spectrogram Transformer (AST)**, introduced by [Gong et al. 2021], adapts the transformer architecture for audio classification tasks. It operates by segmenting the input spectrogram into overlapping patches, which are linearly projected and enriched with positional and class embeddings. These patches are then fed into a Trans-



(a) Model architecture for LSTM. **(b) Model architecture for AST.**

Figure 3. Architectures of the LSTM and AST models.

former encoder, which models the sequence through self-attention mechanisms. The encoder output is finally passed through a linear layer to produce class predictions. By leveraging the global attention capabilities of transformers, AST can capture both local and contextual features within audio data. Figure 3b shows a view of the basic architectural model, as we implemented in our evaluation.

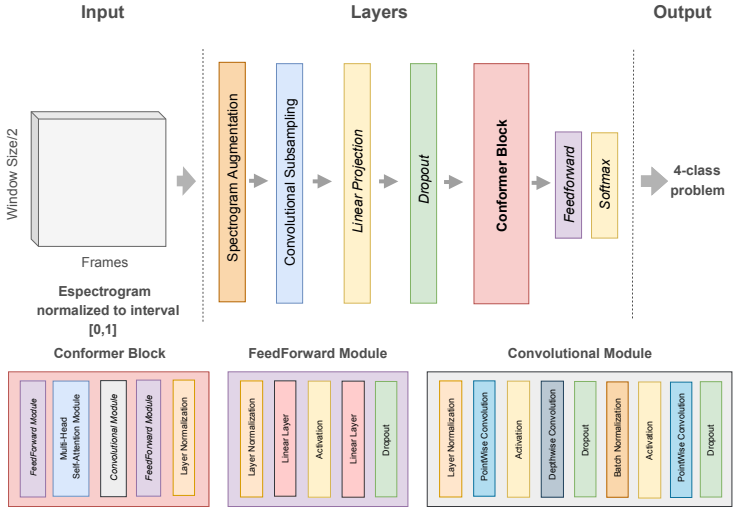


Figure 4. Model architecture considered for the Conformer model.

Finally the **Conformer** model, whose reference implementation considered in this research is shown in Figure 4, is specifically designed for audio signal processing tasks, such as automatic speech recognition [Gulati et al. 2020]. It begins with a data augmentation step using SpecAugment, followed by convolutional subsampling to reduce the temporal resolution while preserving salient features. The subsampled input is transformed via a linear layer and passed through a dropout layer to reduce the risk of overfitting. The core of the model consists of a series of Conformer blocks, which integrate convolutional modules with transformer-based self-attention. This hybrid design enables the model to effectively capture both local patterns and global dependencies, making it well-suited for complex audio modeling tasks.

3.3. Evaluation Pipeline

Figure 5 summarizes the evaluation pipeline, depicting (1) the transformation of raw audio signals into model inputs and (2) the subsequent evaluation workflow.

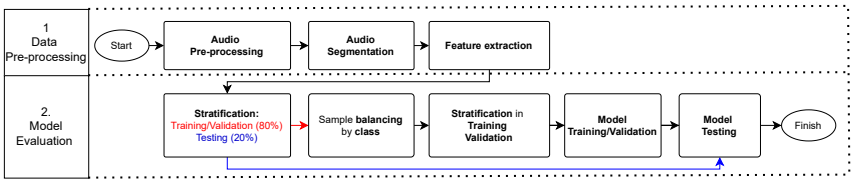


Figure 5. Data pre-processing and evaluation methodology used in our research.

1. Data Pre-processing: First, we standardize the audio files in terms of frequency content and sampling rate, thereby ensuring consistency across the dataset. Initially, a low-pass filter with a cutoff frequency of 4 kHz is applied to eliminate high-frequency noise. Subsequently, the audio signals are downsampled to 8 kHz, which corresponds to the lowest sampling rate among the datasets utilized. Segmentation is implemented using a sliding window approach, wherein each segment is of fixed duration and the stride is set to half the window length.

Feature extraction depends on the model: for the Residual Model, AST and Conformer, each segmented audio frame is transformed into a time-frequency representation using the Short-Time Fourier Transform (STFT). The resulting complex-valued spectrogram is converted into a magnitude spectrogram, which is then scaled in decibels using a dynamic range threshold of 80 dB below the maximum amplitude. For the MLP and LSTM models, we preprocess the input waveforms by extracting their complex number representations (magnitude and phase components). These components are then normalized to a [0,1] range with a mean value of 0.5.

2. Model Evaluation: For the experimental evaluation, we implemented five distinct models (Section 3.2). The dataset was first divided using the holdout method, generating the training/validation and testing sets in the 80/20 proportion (Step 1: Stratification). Model hyperparameters were tuned empirically, and all models were trained for 20 epochs. The input audio segments used for training and evaluation were standardized to a fixed duration of approximately 2.5 seconds to ensure consistency across samples. Next, a 5-fold cross-validation approach was employed over the training/validation set (Step 2: Stratification in Training and Validation).

To address class imbalance, we augmented the dataset by generating additional samples for the minority class (Step 3: Sample balancing by class). Performing balancing after stratification prevents information leakage by ensuring that class distributions are adjusted only within the training set. We partitioned the raw audio signals into overlapping frames (with a hop size of 1250 ms and window length of 2500 ms).

While the training and validation sets were organized using 5-fold stratified sampling, the test set was kept separate and reserved for final performance evaluation (Step 4: Model Training/Validation). To assess model performance, standard binary classification metrics were used, including accuracy, precision, recall, and F1-score. In addition, confusion matrices were analyzed to identify potential classification biases and better understand model behavior (Step 4: Model Testing).

All models were trained using a batch size of 32, the Adam optimizer, and a learning rate of 0.01%. The binary cross-entropy loss function was employed to guide the optimization process. The experiments were conducted on a computational environment equipped with an NVIDIA Tesla P100 GPU (2016) featuring 3584 CUDA cores and 16 GB of memory. The system also included 32 GB of RAM, an Intel(R) Xeon(R) processor running at 2.20 GHz, and 80 GB of available storage.

4. Experiments and Results

In this section, we present and analyze the key findings of our study. Adopting a top-down approach, we first discuss the overarching results before examining specific patterns and underlying mechanisms that explain these outcomes.

4.1. Comparative Evaluation

Figure 6 reports binary classification metrics—accuracy, precision, recall, and F1-score—across all evaluated models. The MLP achieves the highest overall performance, with precision and F1-score reaching 97, followed closely by the LSTM, which maintains balanced scores across all metrics. AST and Conformer yield intermediate results, peaking at 93 for both precision and recall, yet falling short of the dense architectures. The Residual Network, while stable, shows the lowest overall scores ($\approx$ 88), indicating reduced discriminative capacity in this task. These findings suggest that well-calibrated dense architectures, such as the MLP, can outperform more complex sequential or hybrid models, particularly in audio classification scenarios involving pre-processed spectral features and limited contextual variability.

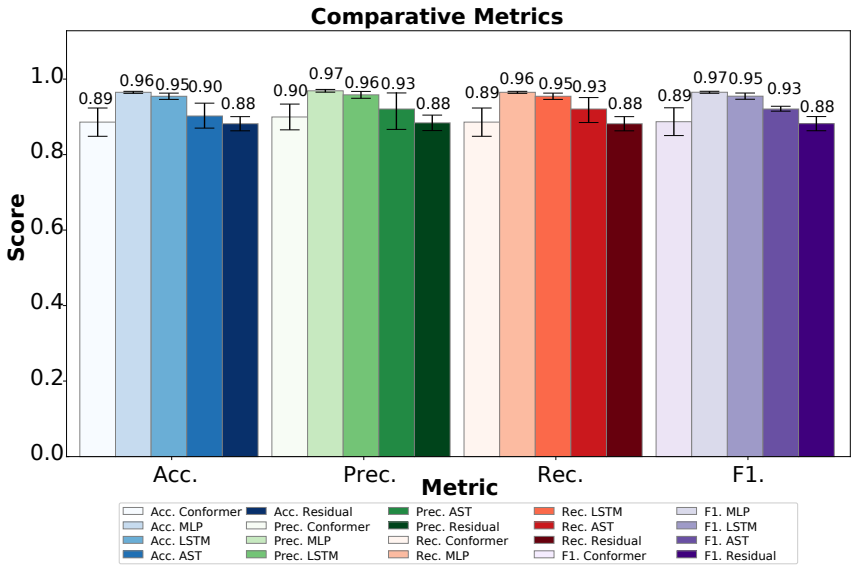


Figure 6. Comparison between different architectures.

4.2. Analysis of ROC Curves

Figure 7 presents the ROC curves for each of the evaluated models. The curves were generated for each class in the dataset using a one-vs-all classification approach. It is evident from the figure that the class with the highest error rate was generally Class 4, corresponding to non-*Aedes Aegypti* mosquitoes. Additionally, the models that demonstrated

the best overall performance were the MLP and LSTM, both showing superior area under the curve (AUC) values, indicating a strong ability to discriminate between classes.

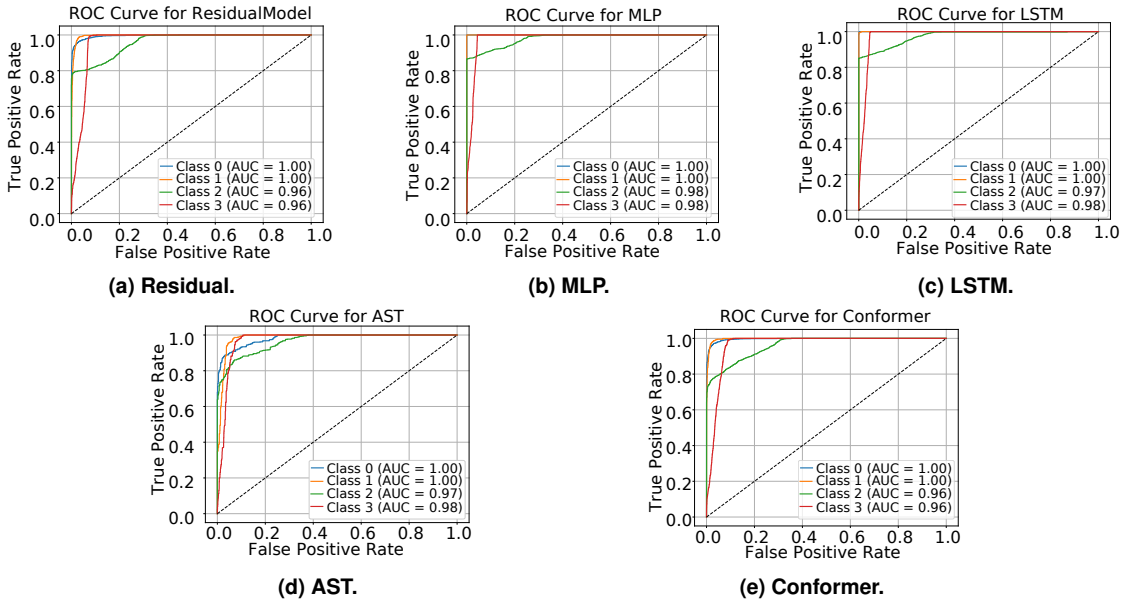


Figure 7. ROC curves for each model.

4.3. Analysis of Confusion Matrices

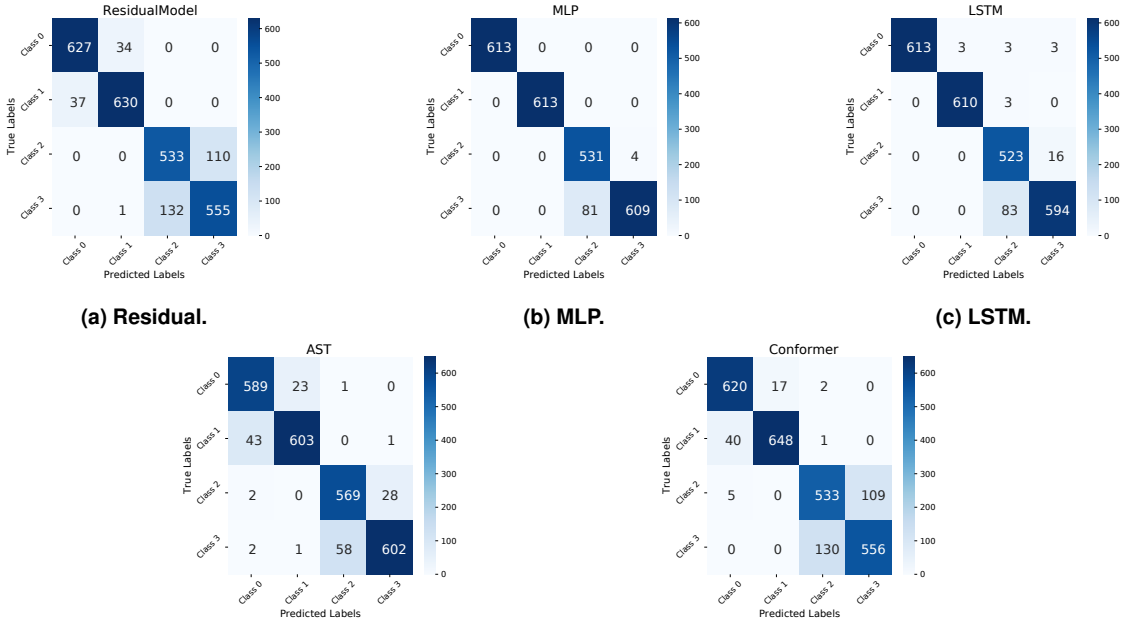
Figure 8 illustrates the multiclass confusion matrices for each of the evaluated models. Overall, all models demonstrated good classification ability, as evidenced by the predominance of diagonal elements in the matrices, which represent correct predictions. This highlights the effectiveness of the models in accurately classifying the audio segments. Among the models, the MLP and LSTM — the lightest architectures evaluated — showed the best predictive performance, with a higher concentration of correct predictions along the diagonal. On the other hand, the Conformer and Residual Network models exhibited lower predictive accuracy, with sparser confusion matrices and a more dispersed distribution of errors, indicating greater difficulty in correctly assigning class labels.

4.4. Training Evaluation

Figure 9 presents the training and validation loss curves over the epochs for each model. These curves provide insights into the comparative performance of the different models, revealing variations in their convergence behavior and learning dynamics. Overall, most models exhibited stable behavior, with a progressive reduction in both losses. It is observed that, starting from the seventh epoch, most models began to show stability in the training loss, indicating a maturation point in the learning process.

5. Final Considerations

In this paper, we present an evaluation of the effectiveness of various neural network architectures for the task of mosquito identification based on audio signals. Our results highlighted the strengths and limitations of each model, with MLP and LSTM—despite their simplicity—outperforming more complex architectures such as AST, Conformer,



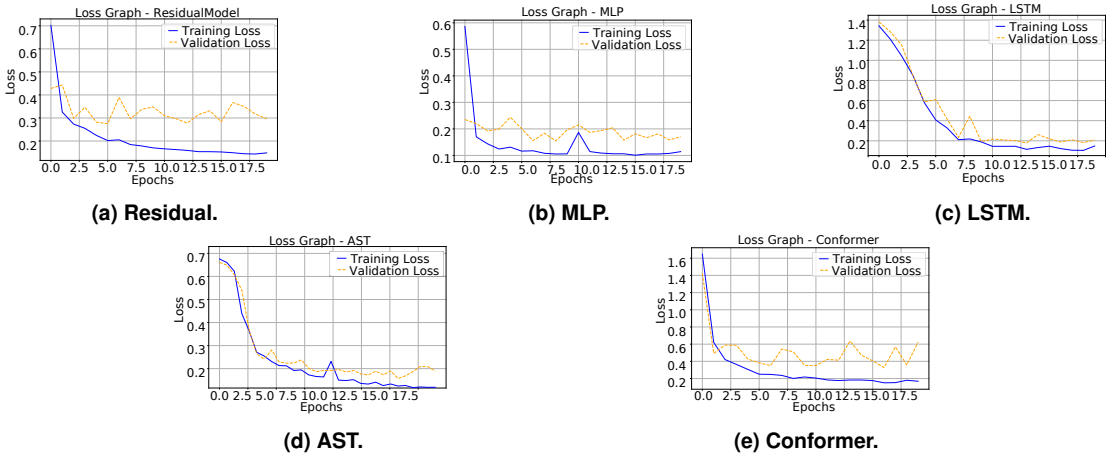
(a) Residual. **(b) MLP.** **(c) LSTM.**

(d) AST. **(e) Conformer.**

Figure 8. Confusion matrices for each model.

and Residual networks. In particular, simpler architectures like MLPs and LSTMs demonstrated advantages likely associated with faster convergence, allowing for better data fitting within reduced training time. This characteristic is of special interest for embedded applications, such as edge devices, where constraints on processing power, memory, and energy consumption are severe.

In spite of the results achieved, our work provides room for further research, including empirical hyperparameter tuning, a restricted experiment scope, and a small dataset, which may limit generalization. Future research should also explore hybrid architectures, larger datasets for better representativeness, and automated optimization strategies like Bayesian Optimization and NAS to improve model efficiency and adaptability. In summary, we expect that these results foster future research in the field, in the quest for



(a) Residual. **(b) MLP.** **(c) LSTM.**

(d) AST. **(e) Conformer.**

Figure 9. Training curves for each model.

tools that can detect *Aedes aegypti* mosquitoes in real-time using smartphone apps (just like Shazam) or intelligent mosquito traps.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 and Grant #88887.954253/2024-00 (INCT ICoNIoT). It was also financed in part by Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), grant nr. 22/2551-0000390-7 (RITE CIARS). It was also financed in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil (CNPq) - Grant #314506/2023-3 and #405940/2022-0 (INCT ICoNIoT).

References

Adhane, G. et al. (2022). On the use of uncertainty in classifying aedes albopictus mosquitoes. *IEEE Journal of Selected Topics in Signal Processing*, 16(2):224–233.

Ahmed, D. A. et al. (2022). Managing biological invasions: the cost of inaction. *Biological Invasions*, 24(7):1927–1946.

Balestrino, F., Iyaloo, D. P., Elahee, K. B., Bheecarry, A., Campedelli, F., Carrieri, M., and Bellini, R. (2016). A sound trap for aedes albopictus (skuse) male surveillance: response analysis to acoustic and visual stimuli. *Acta tropica*, 164:448–454.

Chen, Y., Why, A., Batista, G., Mafra-Neto, A., and Keogh, E. (2014). Flying insect classification with inexpensive sensors. *Journal of Insect Behavior*, 27(5):657–677.

Fernandes, M. S., Cordeiro, W., and Recamonde-Mendoza, M. (2021). Detecting aedes aegypti mosquitoes through audio classification with convolutional neural networks. *Computers in Biology and Medicine*, 129:104152.

Forsyth, J. et al. (2020). Source reduction with a purpose: Mosquito ecology and community perspectives offer insights for improving household mosquito management in coastal kenya. *PLoS neglected tropical diseases*, 14(5):e0008239.

Gong, Y., Chung, Y.-A., and Glass, J. R. (2021). Ast: Audio spectrogram transformer. *ArXiv*, abs/2104.01778.

Gulati, A., Qin, J., Chiu, C.-C., Parmar, N., Zhang, Y., Yu, J., Han, W., Wang, S., Zhang, Z., and Pang, R. (2020). Conformer: Convolution-augmented transformer for speech recognition. pages 5036–5040.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.*, 9(8):1735–1780.

Johnson, B. J. and Ritchie, S. A. (2016). The siren's song: Exploitation of female flight tones to passively capture male aedes aegypti (diptera: Culicidae). *Journal of medical entomology*, 53(1):245–248.

Kahn, M. C., Celestin, W., Offenhauser, W., et al. (1945). Recording of sounds produced by certain disease-carrying mosquitoes. *Science (Washington)*, pages 335–6.

Kim, D., DeBriere, T., Cherukumalli, S., White, G., and Burkett-Cadena, N. (2021). Infrared light sensors permit rapid recording of wingbeat frequency and bioacoustic species identification of mosquitoes. *Scientific Reports*, 11.

Kiskin, I., Zilli, D., Li, Y., Sinka, M., Willis, K., and Roberts, S. (2020). Bioacoustic detection with wavelet-conditioned convolutional neural networks. *Neural Computing and Applications*, 32(4):915–927.

Leandro, A. S. et al. (2022). Citywide integrated aedes aegypti mosquito surveillance as early warning system for arbovirus transmission, brazil. *Emerging Infectious Diseases*, 28(4):707.

Li, Z., Zhou, Z., Shen, Z., and Yao, Q. (2005). Automated identification of mosquito (diptera: Culicidae) wingbeat waveform by artificial neural network. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 483–489. Springer.

Motta, D. et al. (2019). Application of convolutional neural networks for classification of adult mosquitoes in the field. *PLOS ONE*, 14(1):1–18.

Mukundarajan, H., Hol, F. J. H., Castillo, E. A., Newby, C., and Prakash, M. (2017). Using mobile phones as acoustic sensors for high-throughput mosquito surveillance. *Elife*, 6:e27854.

Ouyang, T.-H., Yang, E.-C., Jiang, J.-A., and Lin, T.-T. (2015). Mosquito vector monitoring system based on optical wingbeat classification. *Computers and Electronics in Agriculture*, 118:47–55.

Paim, K. O. et al. (2024). Acoustic identification of ae. aegypti mosquitoes using smartphone apps and residual convolutional neural networks. *Biomedical Signal Processing and Control*, 95:106342.

Piczak, K. J. (2015). ESC: Dataset for Environmental Sound Classification. In *23rd Annual ACM Conference on Multimedia*, pages 1015–1018. ACM Press.

Qureshi, A. I. (2018). Chapter 2 - mosquito-borne diseases. In Qureshi, A. I., editor, *Zika Virus Disease*, pages 27–45. Academic Press.

Su Yin, M. et al. (2022). A deep learning-based pipeline for mosquito detection and classification from wingbeat sounds. *Multimedia Tools and Applications*, 2022.

Townson, H. et al. (2005). Exploiting the potential of vector control for disease prevention. *Bulletin of the World Health Organization*, 83:942–947.

Vasconcelos, D., Nunes, N., Ribeiro, M., Prandi, C., and Rogers, A. (2019). Locomobis: a low-cost acoustic-based sensing system to monitor and classify mosquitoes. In *2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, pages 1–6. IEEE.

Waltz, E. et al. (2021). First genetically modified mosquitoes released in the united states. *Nature*, 593(7858):175–176.

Wei, X., Hossain, M., and Ahmed, K. A. (2022). A resnet attention model for classifying mosquitoes from wing-beating sounds. *Scientific Reports*, 12:10334.

Yin, M. S. et al. (2021). A lightweight deep learning approach to mosquito classification from wingbeat sounds. In *Conference on Information Technology for Social Good*, GoodIT '21, page 37–42, New York, NY, USA. ACM.