Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 6, No. 4, Agustus 2025, Page. 2067-2078
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4922

Mosquito Species Classification Using Wingbeat Acoustic Signals Based on Bidirectional Long Short-Term Memory

Bella Melati Wiranur Dwifani[1], Fatan Kasyidi*[2], Ridwan Ilyas[3]

[1, 2, 3]Informatics, Faculty of Science and Informatics, Jenderal Achmad Yani University, Indonesia
Email: [2]fatan.kasyidi@lecture.unjani.ac.id

Received : Jun 20, 2025; Revised : Jul 1, 2025; Accepted : Jul 6, 2025; Published : Aug 18, 2025

Abstract

The increasing prevalence of mosquito-borne diseases such as Dengue, chikungunya, and malaria underscores the urgent need for effective mosquito vector monitoring. This study proposes a non-invasive classification system of mosquito species based on wingbeat acoustic signals using a deep learning approach with Bidirectional Long Short-Term Memory (BiLSTM). The audio dataset was collected from the Wingbeats repository, consisting of six major mosquito species. Preprocessing was performed using Discrete Wavelet Transform (DWT) to reduce noise. Feature extraction combined Linear Predictive Coding (LPC) and Mel-Spectrogram to represent spectral and temporal signal characteristics. Each binary model was trained in a one-vs-rest scheme to recognize a target species against others, and a BaggingClassifier was used to fuse predictions from six BiLSTM models. Evaluation showed that the proposed system achieved a final accuracy of 96.85% and F1-score of 95.03%, with confusion matrices showing near-diagonal performance. The results indicate that the hybrid LPC-Mel features and ensemble BiLSTM architecture are effective for mosquito species classification using acoustic signals.

Keywords: *Acoustic Classification, BiLSTM, Ensemble Learning, LPC, Wingbeat.*

1. INTRODUCTION

Mosquitoes, especially those from the Aedes aegypti and Aedes albopictus species, are known vectors of serious diseases such as dengue fever, chikungunya, and malaria, which remain a major health concern in Indonesia, particularly in West Java[1], [2]. The surge in dengue cases during the dry season has been exacerbated by climate phenomena like El Niño, which increases temperature and facilitates mosquito breeding in stagnant water[3]. According to Circular Letter No. HK.02.02/C/466/2025 issued by the Indonesian Ministry of Health, dengue and chikungunya cases have risen significantly in early 2025, with 38.740 reported infections and 182 deaths by April 13, 2025[4].

Mosquito wingbeat frequencies vary by species: Aedes albopictus (463–541 Hz), Culex quinquefasciatus (366–437 Hz), and Anopheles crawfordi (295–338 Hz). These frequencies also increase with age, stabilizing in adulthood[5].

Previous research using the LSTM method and audio features like Mel-Spectrogram, Log-Mel, and Mel-Frequency Cepstral Coefficients (MFCC) successfully classified Aedes, Anopheles, and Culex mosquitoes. The best accuracy, 96.71%, was achieved with the Log-Mel feature, highlighting the effectiveness of frequency-based audio classification[6]. The researchers proposed a mosquito classification method based on images using Bootstrap Your Own Latent (BYOL), which achieved 96.77% accuracy with only 10% labeled data[7].

Recent studies show mosquito wingbeat sounds can be used for species classification, with MFCC and Convolutional Neural Network (CNN) achieving up to 99.46% accuracy. However, these methods struggle in noisy environments, as seen in Abuzz and HumBug datasets[8]. Another study using Gammatone Frequency Cepstral Coefficients (GFCC) with a Multi-Layer Perceptron (MLP) model

Vol. 6, No. 4, Agustus 2025, Page. 2067-2078
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4922

achieved 93.82% accuracy for classifying Aedes species using the Wingbeats dataset[9]. The WbNet model, based on ResNet and self-attention, successfully classified six mosquito species from wingbeat sounds, 89.9% on the Wingbeats dataset[10]. The ResNet-9 model based on 1D convolution was used for classifying insect wingbeat sounds, utilizing raw audio data without spectrograms. As a result, the small model achieved an average accuracy of 95.37% on the Wingbeats dataset, outperforming DenseNet121 and WbNet[11]. The use of a CNN based on a Residual Attention Network for classifying six mosquito species from wingbeat signals, using Mel-Spectrogram as input. The best model achieved an accuracy of 89.9% on the Wingbeats dataset[12].

This study introduces a novel approach using Linear Predictive Coding (LPC) combined with a BiLSTM neural network to analyze wingbeat audio signals. The method focuses on capturing both spectral and temporal patterns that characterize different mosquito species. The use of BiLSTM allows for bidirectional sequence modeling, enhancing the system's capability to learn context-aware representations. To further boost classification performance, an ensemble strategy was employed by training six separate BiLSTM models in a one-vs-rest scheme and combining their outputs using a BaggingClassifier. This system is evaluated on a publicly available dataset (Wingbeats), showing significant performance improvements in accuracy and reliability.

While previous studies have explored mosquito species classification using deep learning approaches such as CNN, ResNet, and LSTM with audio features like MFCC, GFCC, and Mel-Spectrogram, many of these efforts focused on binary classification or faced challenges in noisy recording conditions. Furthermore, the integration of complementary spectral features like LPC has not been widely examined. In this study, a multi-class classification framework is presented by combining LPC and Mel-Spectrogram features within a BiLSTM architecture trained in a one-vs-rest strategy, followed by ensemble fusion using a BaggingClassifier. The aim is to enhance species-level recognition across six mosquito classes, even in cases of overlapping acoustic characteristics. This contribution is expected to complement prior approaches and support the development of adaptive, non-invasive bioacoustic classification systems for mosquito vector surveillance.

2. RESEARCH METHOD

This research uses experimental methods to classify mosquito species from noisy wingbeat audio, employing BiLSTM with hybrid features and ensemble learning for robust performance.

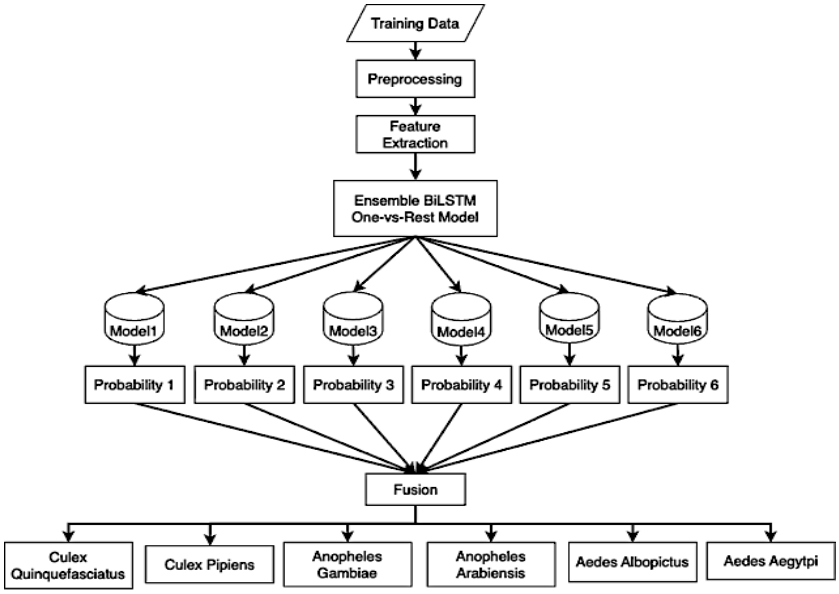

Figure 1. Classification of Mosquito Sounds Using BiLSTM

Vol. 6, No. 4, Agustus 2025, Page. 2067-2078
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4922

Figure 1 shows the ensemble architecture, six one-vs-rest BiLSTM models output class probabilities, which are fused using a BaggingClassifier to produce the final prediction[13]. This ensemble strategy enhances classification robustness by reducing the influence of individual model errors and improving generalization[14].

2.1. Dataset and Data Distribution

The dataset used in this study is the Wingbeats dataset, a publicly available audio collection of mosquito wingbeat sounds frequently used in bioacoustic research[10]. It contains 279.566 audio samples from six mosquito species. Each audio file lasts approximately one second and is labeled by species. Table 1 presents the distribution of mosquito species in the original dataset.

Table 1. Initial Class Distribution in the Wingbeats Dataset

Mosquito Species	Number of Samples
Ae. Aegypti	85.553
Ae. Albopictus	20.231
An. Arabiensis	19.297
An. Gambiae	49.471
Cu. Pipiens	30.415
Cu. Quinquefasciatus	74.599
Total	**279.556**

As shown in Table 1, Ae. aegypti and Cu. quinquefasciatus dominate the dataset, while species like An. arabiensis and Ae. albopictus have considerably fewer samples. Such class imbalance can lead to a biased model that performs poorly on underrepresented classes.

2.2. Signal Preprocessing

The raw mosquito wingbeat recordings often contain background noise and frequency fluctuations due to varying environmental conditions. To ensure that relevant spectral features are preserved while reducing unwanted components, this study uses Discrete Wavelet Transform (DWT) as a denoising method. Unlike traditional techniques such as STFT or Fourier Transform, DWT provides a multi-resolution analysis, capturing both time and frequency domain representations with high precision[15].

$$c^{w}_{m-1,k} = \sum_{n} h_{n-2k} \cdot c^{w}_{m,n} \quad (1)$$

$$d^{w}_{m-1,k} = \sum_{n} \tau_{n-2k} \cdot c^{w}_{m,n} \quad (2)$$

This decomposition helps isolate relevant signal components while discarding high-frequency noise. The processed signal is then reconstructed using inverse DWT, resulting in a cleaner waveform that retains the mosquito wingbeat patterns[15].

2.3. Feature Extraction

After denoising, the next step involves extracting features that can represent both temporal and spectral characteristics of mosquito wingbeats. Three main types of features are used: Linear Predictive Coding (LPC), Mel-Spectrogram, and frame-level statistical features.

Vol. 6, No. 4, Agustus 2025, Page. 2067-2078
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4922

2.3.1. Linear Predictive Coding (LPC)

The LPC method estimates the current audio sample as a linear combination of its previous values. The prediction formula is given by[16]:

$$s(n) = \sum_{i=1}^{p} a_i s(n-i) \quad (3)$$

Where a_i represents the predictive coefficients and p is the prediction order. This equation assumes that the audio signal behaves linearly over short time intervals. The residual error of this prediction, calculated as[16]:

$$e(n) = s(n) - \sum_{i=1}^{p} a_i s(n-i) \quad (4)$$

Provides insight into the accuracy of the linear model. LPC is particularly effective in representing the envelope of the speech-like signals present in mosquito wingbeats. It compresses the spectral content into a small set of coefficients, which can then be used as compact descriptors of the sound[17].

2.3.2. Mel Spectogram

To complement the spectral modeling offered by LPC, this study also utilizes the Mel-Spectrogram, which transforms the signal into a perceptually scaled frequency representation, approximating how the human ear perceives sound[18]. Frequencies are first converted to the Mel scale using the following formula[19]:

$$m = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right) \quad (5)$$

This nonlinear transformation aligns with the sensitivity of the auditory system, emphasizing lower frequencies where mosquito wingbeats typically reside. The result is a time–frequency matrix that is further enhanced with its delta and delta-delta components to capture temporal variation in spectral content[19].\

$$H_m(k) = \begin{cases} 0 \\ \frac{k - f(m-1)}{f(m) - f(m-1)} \\ \frac{f(m+1) - k}{f(m+1) - f(m)} \\ 0 \end{cases} \quad (6)$$

The equation defines the response of the m-th Mel filter as a triangular function that increases linearly up to its center frequency and then decreases linearly. Frequencies outside the filter's range have zero contribution. This structure simulates the human auditory perception, which is more sensitive to lower frequency changes, and ensures smooth spectral transitions across Mel bands.

Spectrograms are enhanced with delta and delta-delta features. Additionally, frame-wise statistics such as energy, mean, standard deviation, and skewness are concatenated, resulting in a final feature tensor of shape [T,F][20].

2.4. Model Architecture

To process the sequential feature vectors generated in the previous stage, a deep neural network based on Bidirectional Long Short-Term Memory (BiLSTM) is employed. BiLSTM is capable of capturing dependencies in both forward and backward directions, making it suitable for modeling the temporal dynamics of wingbeat patterns. The forward hidden state at time t, denote by $\overrightarrow{h_t}$, is computed as[21]:

Vol. 6, No. 4, Agustus 2025, Page. 2067-2078
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4922

$$\overrightarrow{h_t} = \text{LSTM}\left(\overrightarrow{h_{t-1}}, x_t\right) \quad (7)$$

While the backward hidden state $\overleftarrow{h_t}$, which processes the sequence in reverse, is given by[21]:

$$h_t = \left[\overrightarrow{h_t}; \overleftarrow{h_t}\right] \quad (8)$$

The final hidden representation h_t is the concatenation of both directions:

$$h_t = \left[\overrightarrow{h_t}; \overleftarrow{h_t}\right] \quad (9)$$

This architecture allows the model to make predictions based not only on past information but also future context within the audio sequence[21]. The final output is passed through a sigmoid-activated dense layer for binary classification in a one-vs-rest scheme[14].

2.5. Training and Evaluation Strategy

Each of the six BiLSTM models is trained individually using a one-vs-rest strategy. All models receive balanced input 10.000 positive samples for the target class and 10.000 negative samples drawn evenly from the remaining five classes. The dataset is divided using a stratified split with proportions of 60% for training, 20% for validation, and 20% for testing. Preprocessing and fiture extraction includes DWT denoising (db8, level 5), LPC (order 30), Mel-Spectrogram (64 filters, delta, delta-delta), and statistical features.

To optimize model performance and ensure convergence, we used the AdamW optimizer, which integrates weight decay with Adam to reduce overfitting and improve generalization[22]. The binary crossentropy loss function was chosen since each classifier performs binary prediction[23].

To evaluate classification performance, we use standard metrics derived from the confusion matrix: accuracy, precision, recall, and F1-score[24]. These metrics are defined mathematically as follows[25]:

$$\text{Accuracy} = \frac{\text{TN+TP}}{\text{TN+FP+TP+FN}} \quad (10)$$

Measures the proportion of correct predictions made by the model out of all predictions[24]. It reflects the overall correctness but may be misleading in imbalanced datasets where one class dominates[25].

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (11)$$

Indicates how many of the samples predicted as positive are actually positive[24]. It is useful in scenarios where false positives are costly and should be minimized[25].

$$\text{Recall} = \frac{\text{TP}}{\text{TF+FN}} \quad (12)$$

Measures the model's ability to correctly identify all actual positive cases[24]. High recall means few false negatives, which is important when missing positive instances is critical[25].

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (13)$$

Vol. 6, No. 4, Agustus 2025, Page. 2067-2078
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4922

Is the harmonic mean of precision and recall. It provides a balanced metric that is particularly valuable when the dataset is imbalanced or when both false positives and false negatives need to be considered equally[25].

After training, the output probabilities from all six binary models are fused using a BaggingClassifier, which aggregates predictions by selecting the class with the highest probability score[14]. This strategy significantly enhances the overall classification robustness and resolves conflicts among individual classifiers.

3. RESULT

This section presents results from six one-vs-rest BiLSTM models and their fusion for multi-class classification.

3.1. Performance of Individual Models

The classification results of each BiLSTM model were evaluated using a one-vs-rest scheme for six mosquito species. Figure 2 summarizes the values of precision, recall, F1-score, and accuracy. Among all models, the classifier for C. quinquefasciatus achieved the highest F1-score of 0.8403 and accuracy of 83.48%, indicating that this species may have more consistent or distinguishable acoustic patterns compared to others. Its high recall value (0.8695) also suggests that the model was able to correctly capture most of the true positive instances.

The model for An. arabiensis also showed relatively stable performance with an F1-score of 0.7880 and accuracy of 78.17%. Meanwhile, An. gambiae achieved slightly lower results with an F1-score of 0.7660, indicating minor differences in the temporal or spectral patterns that may contribute to overlapping feature representations. Models for C. pipiens and Ae. albopictus had F1-scores of 0.7398 and 0.7236 respectively, with moderate recall values between 0.7375 and 0.7570. This may reflect more variability in the signal patterns of these species, possibly due to wider recording conditions or more variation within class.

Classification Summary (One-vs-Rest):

	Precision	Recall	F1-Score	Accuracy
Ae. aegypti vs all	0.7009	0.7160	0.7084	0.7052
C. quinquefasciatus vs all	0.8130	0.8695	0.8403	0.8348
An. gambiae vs all	0.7488	0.7840	0.7660	0.7605
C. pipiens vs all	0.7234	0.7570	0.7398	0.7338
Ae. albopictus vs all	0.7102	0.7375	0.7236	0.7183
An. arabiensis vs all	0.7662	0.8110	0.7880	0.7817

Figure 1. One-vs-Rest Classification Results

The model with the lowest performance was found in the Ae. aegypti classifier, which recorded an F1-score of 0.7084 and accuracy of 70.52%. One possible explanation is the frequency proximity and signal shape similarity between Ae. aegypti and other species such as Ae. albopictus, which may have contributed to more frequent misclassifications. These findings also highlight that classification difficulty may not solely depend on the number of training samples but on the distinctiveness of the acoustic patterns of each species.

To further observe the model's behavior over training time, validation accuracy and loss curves were monitored throughout 100 epochs for each binary classifier. Figure 3 shows the trend of validation accuracy per class.

Vol. 6, No. 4, Agustus 2025, Page. 2067-2078
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4922

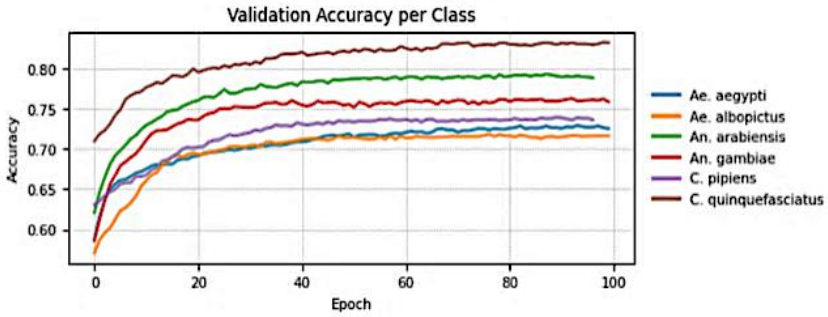

Figure 2. Accuracy of Six Mosquito Using BiLSTM

From the early epochs, Culex quinquefasciatus displayed a steep and consistent rise in accuracy, surpassing 0.80 around epoch 40 and maintaining a stable trend above 0.83. This could indicate that the species has signal characteristics that are relatively more distinguishable by the BiLSTM model, potentially due to its spectral distribution or stability across samples. In contrast, Aedes aegypti and Aedes albopictus consistently produced lower validation accuracies, both converging below 0.73. The curves for these classes flattened early and did not exhibit significant upward trends in the later epochs. This may suggest that both species have overlapping frequency components or temporal signal patterns, making it difficult for the model to build a strongly discriminative representation. Anopheles arabiensis and Anopheles gambiae presented moderately increasing accuracy curves that plateaued near 0.78 and 0.76 respectively. Meanwhile, Culex pipiens achieved a relatively stable accuracy in the range of 0.73–0.74 after epoch 40. These trends reinforce the idea that certain species may inherently carry more acoustic regularity, while others tend to fluctuate or resemble one another more closely.

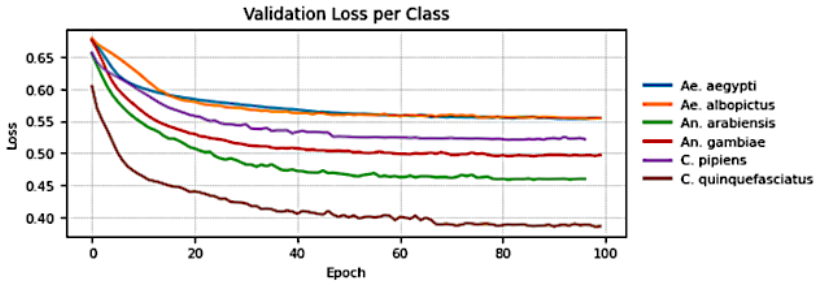

Figure 3. Loss of Six Mosquito Using BiLSTM

Figure 4 displays the validation loss curves across the same training epochs. Once again, the C. quinquefasciatus model exhibited the most favorable trend, showing a rapid and smooth decrease in validation loss to a final value below 0.40. In comparison, Ae. aegypti and Ae. albopictus settled at the highest loss values—both remaining above 0.55—indicating that the models for these classes had greater difficulty minimizing classification error.

Another interesting observation can be seen in the early phase (epoch 0–20), where all models experienced rapid loss reduction, implying effective early-stage learning. However, loss convergence patterns diverged across classes, where models such as An. arabiensis continued to refine well into later epochs, while others like Ae. albopictus stabilized prematurely. From both accuracy and loss perspectives, C. quinquefasciatus not only reached higher classification consistency but also showed signs of more stable optimization behavior. This aligns with earlier findings in the classification report and confusion matrix, where this species achieved higher precision and recall.

Overall, the visual trends offer additional insights beyond static evaluation metrics. Specifically, they help uncover learning dynamics and possible saturation points for each class. The consistently lower accuracy and higher loss in Ae. aegypti and Ae. albopictus further support the hypothesis that

Vol. 6, No. 4, Agustus 2025, Page. 2067-2078
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4922

class overlap or feature redundancy may have hindered clearer class separation. These findings highlight the need for deeper exploration of signal characteristics and potential refinement of feature extraction strategies for species with marginal classification performance. It may also be beneficial to investigate whether more fine-tuned frame segmentation or signal normalization techniques could enhance performance on classes with slower learning curves.

An important observation is that none of the individual models showed extreme overfitting or instability, as all results fall within a relatively narrow band between 0.70 and 0.84 for F1-scores, and 70% to 83% for accuracy. This suggests that the learning process was consistent across species, and that the feature representation (LPC + statistical) provided a reasonable basis for temporal modeling.

3.2. Fusion Results

As shown in Figure 5, the multi-class classification performance after applying the BaggingClassifier fusion approach illustrates how well the system was able to recognize each mosquito species. The highest F1-score of 0.9943 was achieved for Ae. aegypti, with an exceptionally high recall of 0.9990. This suggests that nearly all actual instances of Ae. aegypti in the dataset were correctly recognized by the model, indicating a very high sensitivity toward this class.

Classification Report:

	precision	recall	f1-score
Ae. aegypti	0.9896	0.9990	0.9943
C. quinquefasciatus	0.9231	0.9706	0.9462
An. gambiae	0.9592	0.9691	0.9641
C. pipiens	0.9403	0.9306	0.9354
Ae. albopictus	0.9487	0.9151	0.9316
An. arabiensis	0.9754	0.9130	0.9432
accuracy			0.9692
macro avg	0.9560	0.9496	0.9525
weighted avg	0.9693	0.9692	0.9691

Figure 4. BiLSTM Fusion Results

The model also demonstrated strong recognition of An. gambiae and C. pipiens, with F1-scores of 0.9641 and 0.9354, respectively. These values reflect a good balance between precision and recall, suggesting that the model had a stable classification pattern for these species. In the case of C. quinquefasciatus, the recall was notably high (0.9706), indicating that the model was generally able to detect this species. However, the precision of 0.9231 implies that some samples from other species may have been incorrectly predicted as C. quinquefasciatus, possibly due to shared spectral features with certain classes. For Ae. albopictus and An. arabiensis, F1-scores were slightly lower at 0.9316 and 0.9432, respectively. Although both are still high, the recall values of 0.9151 and 0.9130 suggest that the model was slightly less sensitive in capturing all instances of these two species. This could indicate acoustic similarity with other classes, which can lead to misclassifications.

Overall, as visualized in Figure 5, the macro-average recall of 0.9496 and weighted recall of 0.9692 indicate that no species was significantly neglected. The fusion-based classification maintained consistent recognition across all classes, even though inter-class differences still played a role in performance variations.

3.3. Confusion Matrix Analysis

The confusion matrix presented in Figure 5 provides a more detailed view of the classification outcomes for each class after the fusion stage. Diagonal values close to 1.00 indicate correct predictions,

Vol. 6, No. 4, Agustus 2025, Page. 2067-2078
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4922

while off-diagonal entries represent misclassifications. The matrix confirms that the model was able to classify Ae. aegypti samples with perfect accuracy (1.00), showing zero misclassifications for this class.

For C. quinquefasciatus and An. gambiae, correct predictions reached 97%, with minimal confusion primarily coming from neighboring classes. For instance, 2% of An. gambiae samples were incorrectly predicted as C. quinquefasciatus, and 1% of C. quinquefasciatus as Ae. albopictus. Although minor, these misclassifications suggest that certain acoustic similarities exist between these species. C. pipiens achieved 93% correct predictions, with some misclassifications spread thinly across other classes such as Ae. aegypti and Ae. albopictus. Ae. albopictus and An. arabiensis recorded slightly lower values at 92% and 91%, respectively. Notably, Ae. albopictus was occasionally misclassified as C. quinquefasciatus (3%) and C. pipiens (3%), suggesting that those species may share overlapping signal patterns or ambiguous features within the extracted representations.

Overall, Figure 5 supports the classification report findings and highlights that while most classes were well distinguished by the fusion model, minor confusion persisted between acoustically similar species. The near-diagonal dominance of the matrix indicates that the model preserved a strong capacity to isolate and assign each instance to the correct class.

Figure 5. Confusion Matrix

4. DISCUSSIONS

This study adopts an acoustic-based classification approach using wingbeat signals of mosquitoes from the public Wingbeats dataset, which was consistently recorded using optoelectronic devices at the Biogents facility in Germany[9]. This consistent recording setup serves as a key advantage by minimizing channel variations commonly caused by differences in recording equipment—an issue frequently encountered in other datasets such as HumbugDB and Dryad[9].

In comparison, the model developed in this study demonstrates relatively stable and competitive performance. Evaluation results show an accuracy of 96.85% and an F1-score of 95.03% after applying the fusion method. This is slightly higher than the GFCC + MLP model, which achieved an accuracy of 93.82% on the Wingbeats dataset[9]. This difference may be attributed to the use of BiLSTM, which is more sensitive to temporal dynamics than MLP, and the combination of two complementary features:

Vol. 6, No. 4, Agustus 2025, Page. 2067-2078
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4922

LPC, which captures predictive spectral information, and Mel-Spectrogram, which represents energy distribution across frequencies.

From a model architecture perspective, the BiLSTM approach used in this study differs from two ResNet-based approaches [10], [11]. While ResNet models with attention mechanisms have shown high accuracy (up to 99%) in binary classification scenarios, they typically require complex tuning and large data volumes. Moreover, most of these studies focus only on binary classification (e.g., Ae. aegypti vs. non-Ae. aegypti), whereas this research addresses multi-class classification involving six mosquito species, including acoustically overlapping species like Ae. aegypti and Ae. albopictus. This makes the classification task more complex and closer to real-world conditions.

Although the ensemble model effectively improves aggregate performance, the confusion matrix reveals that misclassifications still occur between certain species pairs such as An. gambiae and C. pipiens. This suggests that while the acoustic features are informative, spectral overlap between species remains a challenge.

For future development, adopting Mamba, a recent state space model architecture designed for long sequential data, could be beneficial. Without relying on self-attention mechanisms, Mamba effectively captures complex temporal dynamics, making it highly suitable for classifying audio signals such as mosquito wingbeats.

5. CONCLUSION

This study presents a multi-class classification system for mosquito species based on their wingbeat acoustic signals using BiLSTM models with hybrid LPC and Mel-Spectrogram features. Through one-vs-rest training and ensemble fusion, the proposed approach demonstrates consistent performance across six mosquito classes, despite spectral overlaps and class imbalance. The Wingbeats dataset, recorded uniformly using optoelectronic sensors, provides a reliable benchmark for validating acoustic-based mosquito classification models. Experimental results indicate that the combination of spectral compression via LPC and perceptual mapping through Mel-Spectrogram effectively captures species-specific signal patterns. The BiLSTM architecture, when coupled with ensemble voting, enhances temporal learning and mitigates misclassification from acoustically similar species. While the confusion matrix reveals some inter-class confusion, especially between Culex and Anopheles species, the overall performance remains stable with macro-average recall above 94%. Future improvements may involve incorporating advanced temporal models like Mamba to better capture long-range dependencies in mosquito wingbeat signals. This research contributes to the development of non-invasive, data-driven tools for mosquito species identification, with potential applications in public health surveillance and vector control strategies.

REFERENCES

[1] VOA Indonesia, “Mewaspadai Peningkatan Kasus Demam Berdarah di Musim Kemarau,” Jun. 14, 2023. [Online]. Available: https://www.voaindonesia.com/a/mewaspadai-peningkatan-kasus-demam-berdarah-di-musim-kemarau/7136639.html (accessed Sep. 25, 2024).

[2] Alodokter, “Chikungunya,” 2025. [Online]. Available: https://www.alodokter.com/chikungunya (accessed Sep. 25, 2024).

[3] M. Pirani, C. Lorenz, T. S. de Azevedo, G. L. Barbosa, M. Blangiardo, and F. Chiaravalloti-Neto, “Effects of the El Niño-Southern Oscillation and seasonal weather conditions on Aedes aegypti infestation in the State of São Paulo (Brazil): A Bayesian spatio-temporal study,” *PLoS Negl Trop Dis*, vol. 18, no. 9, p. e0012397, Sep. 2024, doi: 10.1371/journal.pntd.0012397.

[4] CNN Indonesia, “April 2025, Kemenkes Justru Catat Kematian Akibat DBD Hampir 200 Jiwa,” Apr. 24, 2025. [Online]. Available: https://www.cnnindonesia.com/gaya-hidup/20250424083315-255-1222001/april-2025-kemenkes-justru-catat-kematian-akibat-dbd-hampir-200-jiwa (accessed May. 1, 2025).

[5] P. Rajan, D. Goswami, Vanlalhmuaka, S. Datta, B. Rabha, and D. V. Kamboj, "Acoustic behaviour and flight tone frequency changes in adult Aedes albopictus and Culex quinquefasciatus mosquitoes," *Sci Rep*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-89608-7.

[6] E. Toledo *et al.*, "LSTM-based mosquito genus classification using their wingbeat sound," in *Frontiers in Artificial Intelligence and Applications*, IOS Press BV, Sep. 2021, pp. 293–302. doi: 10.3233/FAIA210028.

[7] R. Charoenpanyakul *et al.*, "Enhancing mosquito classification through self-supervised learning," *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-78260-2.

[8] J. H. Rony, N. Karim, M. A. Rouf, S. B. Noor, and F. H. Siddiqui, "Mosquito Species Classification through Wingbeat Analysis: A Hybrid Machine Learning Approach," in *2023 International Conference on Next-Generation Computing, IoT and Machine Learning, NCIM 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/NCIM59001.2023.10212515.

[9] Mariyanto and H. F. Pardede, "Identification of Aedes Mosquitoes with Audio Data Using Gammatone Filter," in *Proceeding - 2024 International Conference on Information Technology Research and Innovation, ICITRI 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 164–169. doi: 10.1109/ICITRI62858.2024.10699195.

[10] X. Wei, M. Z. Hossain, and K. A. Ahmed, "A ResNet attention model for classifying mosquitoes from wing-beating sounds," *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-14372-x.

[11] B. J. Szekeres, M. Natabara Gyongyossy, and J. Botzheim, "A ResNet-9 Model for Insect Wingbeat Sound Classification," in *2023 IEEE Symposium Series on Computational Intelligence, SSCI 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 587–592. doi: 10.1109/SSCI52147.2023.10371871.

[12] D. Vasconcelos, N. J. Nunes, A. Förster, and J. P. Gomes, "Optimal 2D audio features estimation for a lightweight application in mosquitoes species: Ecoacoustics detection and classification purposes," *Comput Biol Med*, vol. 168, Jan. 2024, doi: 10.1016/j.compbiomed.2023.107787.

[13] A. Al Maruf, M. M. Haque, R. A. Rumy, J. J. Puspo, and Z. Aung, "TransembleNet: Enhancing vector mosquito species classification through transfer learning-based ensemble model," *PLoS One*, vol. 20, no. 5 May, May 2025, doi: 10.1371/journal.pone.0322171.

[14] G. Ngo, R. Beard, and R. Chandra, "Evolutionary bagging for ensemble learning," *Neurocomputing*, vol. 510, pp. 1–14, Oct. 2022, doi: 10.1016/j.neucom.2022.08.055.

[15] I. Kiskin, D. Zilli, Y. Li, M. Sinka, K. Willis, and S. Roberts, "Bioacoustic detection with wavelet-conditioned convolutional neural networks," *Neural Comput Appl*, vol. 32, no. 4, pp. 915–927, Feb. 2020, doi: 10.1007/s00521-018-3626-7.

[16] J. Xu, M. Davis, and R. de Frein, "New robust LPC-based method for time-resolved morphology of high-noise multiple frequency signals," in *Proc. IEEE*, 2020, doi: 10.1109/ISSC49989.2020.9180212.

[17] S. M. Williams, N. Aldabashi, P. Cross, and C. Palego, "Challenges in Developing a Real-Time Bee-Counting Radar," *Sensors*, vol. 23, no. 11, Jun. 2023, doi: 10.3390/s23115250.

[18] H. Li, J. Li, H. Liu, T. Liu, Q. Chen, and X. You, "MelTrans: Mel-Spectrogram Relationship-Learning for Speech Emotion Recognition via Transformers," *Sensors*, vol. 24, no. 17, Sep. 2024, doi: 10.3390/s24175506.

[19] K. Kushwaha, "Exploring Mel Spectrograms: A Powerful Feature Extraction Tool for Audio Signals," *Medium*, Aug. 3, 2023. [Online]. Available: https://medium.com/@kunalkushwahatg/exploring-mel-spectrograms-a-powerful-feature-extraction-tool-for-audio-signals-3b68ff6fcf96 (accessed: Feb. 3, 2025).

[20] P. Raghuwanshi and R. Kaushik, "Insect Classification Using Mel-CSTFT: A Fusion of Mel Spectrogram and Chroma STFT Features," in *Proceedings - 1st International Conference on Electronics, Communication and Signal Processing, ICECSP 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICECSP61809.2024.10698728.

[21] Z. Neili and K. Sundaraj, "Addressing Varying Lengths in PCG Signal Classification with BiLSTM Model and MFCC Features," in *Proceedings - 8th IEEE International Conference on*

Vol. 6, No. 4, Agustus 2025, Page. 2067-2078
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4922

Image and Signal Processing and their Applications, ISPA 2024, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ISPA59904.2024.10536851.

[22] P. Zhou, X. Xie, Z. Lin, and S. Yan, "Towards Understanding Convergence and Generalization of AdamW," *IEEE Trans Pattern Anal Mach Intell*, vol. 46, no. 9, pp. 6486–6493, 2024, doi: 10.1109/TPAMI.2024.3382294.

[23] Q. Wang, Y. Ma, K. Zhao, and Y. Tian, "A Comprehensive Survey of Loss Functions in Machine Learning," *Annals of Data Science*, vol. 9, no. 2, pp. 187–212, Apr. 2022, doi: 10.1007/s40745-020-00253-5.

[24] J. Kozak, B. Probierz, K. Kania, and P. Juszczuk, "Preference-Driven Classification Measure," *Entropy*, vol. 24, no. 4, Apr. 2022, doi: 10.3390/e24040531.

[25] R. Rachmawan, "Mengurai metrics: Apa itu presisi, recall, dan bagaimana menginterpretasikannya?," *Medium*, Apr. 18, 2022. [Online]. Available: https://medium.com/himit-pens/mengurai-metrics-apa-itu-presisi-recall-dan-bagaimana-menginterpretasikannya-a15e7f90411e (accessed: April. 10, 2025).