Ecological Informatics 80 (2024) 102495

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Convolutional neural network-based real-time mosquito genus identification using wingbeat frequency: A binary and multiclass classification approach

Endra Joelianto[a,b,*], Miranti Indar Mandasari[c], Daniel Beltsazar Marpaung[d], Naufal Dzaki Hafizhan[d], Teddy Heryono[e], Maria Ekawati Prasetyo[f], Dani[g], Susy Tjahjani[g], Tjandra Anggraeni[h], Intan Ahmad[h]

[a] Instrumentation, Control and Automation Research Group, Institut Teknologi Bandung, Bandung, Indonesia
[b] University Center of Excellence Artificial Intelligence on Vision, NLP and Big Data Analytics, Institut Teknologi Bandung, Bandung, Indonesia
[c] Build Environment Performance Engineering Research Group, Institut Teknologi Bandung, Bandung, Indonesia
[d] Engineering Physics Study Program, Institut Teknologi Bandung, Bandung, Indonesia
[e] Instrumentation and Control Graduate Program, Institut Teknologi Bandung, Bandung, Indonesia
[f] Biomedical Research Centre of Indonesia, Bandung, Indonesia
[g] Faculty of Medicine, Maranatha Christian University, Bandung, Indonesia
[h] School of Life Sciences and Technology, Institut Teknologi Bandung, Bandung, Indonesia

ARTICLE INFO

Keywords:
Dengue infection
Aedes aegypti
Mosquito vector
Wingbeat frequency
Deep learning
Sustainability monitoring

ABSTRACT

Global rises in dengue hemorrhagic fever, especially in Asia and Latin America, underscore the necessity for enhanced public health interventions. *Aedes* spp. mosquitoes are the primary vectors; however, species such as *Culex quinquefasciatus* pose significant health risks by transmitting diseases such as filariasis, impacting millions of people worldwide. This study introduces a real-time convolutional neural network-based mosquito classification system using wingbeat frequency for identifying various mosquito species, with emphasis on *Aedes* sp. We proposed and assessed two models: a binary classification and a multiclass system. The binary system exhibited an outstanding accuracy of 91.76% in distinguishing between *Aedes aegypti* and *Culex quinquefasciatus*. The multiclass system accurately identified female and male *Aedes aegypti* and *Culex quinquefasciatus* with a precision of 87.16%. This innovative approach serves as a potential tool for dengue infection control and a versatile instrument for combating various mosquito-borne illnesses, enhancing vector surveillance for comprehensive disease management.

1. Introduction

The dengue virus, transmitted through *Aedes* spp. mosquitoes, primarily *Aedes aegypti*, causes dengue hemorrhagic fever (DHF). Environmental conditions, such as rainfall, relative humidity, and temperature markedly influence the spread of the dengue fever, and it is prevalent in tropical regions worldwide (World Health Organization, 2014). Over the past two decades, the reported dengue cases have surged eight-fold, particularly affecting Asian and Latin American countries (World Health Organization, 2014). Indonesia reported 137,761 dengue cases with 917 deaths in 2019 alone (Indonesia Ministry of Health, 2019). *Culex* mosquitoes, which inhabit tropical to temperate climates, pose a substantial health risk by transmitting diseases like bancroftian filariasis and various encephalitis viruses (Soh and Aik, 2021; Zoladek and Nisole, 2023). Approximately 1 billion people live in filariasis-endemic areas, with 36 million suffering from chronic disabilities (World Health Organization, 2017). By the end of 2021, Indonesia reported 9354 cases of filariasis with long-term disabilities across 34 provinces, according to the country's Health Ministry (Kemenkes, 2022).

Abbreviations: CNN, convolutional neural network; MFCC, Mel Frequency Cepstral Coefficient; DCT, Discrete cosine transform; FFT, Fast Fourier Transform.

* Corresponding author at: Instrumentation, Control and Automation Research Group, Institut Teknologi Bandung, Bandung, Indonesia.
E-mail addresses: ejoel@itb.ac.id (E. Joelianto), mandasari@itb.ac.id (M.I. Mandasari), beltsazar59@gmail.com (D.B. Marpaung), naufaldh27@gmail.com (N.D. Hafizhan), 23820305@mahasiswa.itb.ac.id (T. Heryono), mariaprasetyo1@gmail.com (M.E. Prasetyo), marcello.dani.dd@gmail.com (Dani), susy.tjahjani@med.maranatha.edu (S. Tjahjani), tjandra@sith.itb.ac.id (T. Anggraeni), intan@itb.ac.id (I. Ahmad).

https://doi.org/10.1016/j.ecoinf.2024.102495
Received 14 August 2023; Received in revised form 21 January 2024; Accepted 21 January 2024
Available online 23 January 2024

Control strategies, including the use of ovitraps, have been developed to monitor dengue vectors (Gonzalez et al., 2023; Hamesse et al., 2023; Sasmita et al., 2021). In Indonesia, these devices have been instrumental in calculating the ovitrap index, assessing the minimum infection rate, and studying their relationship with the dengue incidence rate (Agustiningtyas and Lusiyana, 2017; Juhdi et al., 2019; Sasmita et al., 2021). Ovitraps have proven useful in tracking the vertical transmission of the virus and analyzing the spread of four dengue strains in Makassar City, South Sulawesi (Juhdi et al., 2019). Additionally, they have been adapted to estimate *Culex* mosquito populations and their connection to disease incidence (Agustiningtyas and Lusiyana, 2017; Juhdi et al., 2019; Sasmita et al., 2021). However, they require laboratory analysis and expert handling for species identification, presenting a significant limitation in rapid and on-site vector surveillance (Agustiningtyas and Lusiyana, 2017; Hamesse et al., 2023; Juhdi et al., 2019; Martin et al., 2019; Ortega-Morales et al., 2018; Reed et al., 2019; Sasmita et al., 2021; Zuhriyah et al., 2016).

Advancements in Internet of Things (IoT) technology offer a promising solution to the limitations of conventional ovitraps. IoT-enabled ovitraps have the potential to identify mosquito species in real-time, crucial for early intervention in dengue and filariasis outbreaks. These innovative systems can also provide insights into the oviposition patterns of species such as *Aedes aegypti*, *Aedes albopictus*, and *Culex* sp. Studies have highlighted the relationship between the wingbeat frequency of *Aedes aegypti* and environmental factors such as temperature and humidity (Fanioudakis et al., 2018; Jhaveri et al., 2022; Santos et al., 2019). Further research has expanded this to include various mosquito species, using the wingbeat frequency and Welch power spectral density analysis as identification markers, which are crucial in regions with DHF and filariasis, such as Indonesia (Agustiningtyas and Lusiyana, 2017; Juhdi et al., 2019; Sasmita et al., 2021; Zuhriyah et al., 2016).

Machine learning techniques have successfully categorized the acoustic signatures of wildlife such as those of bees and birds (Tang et al., 2023; Truong et al., 2023). For mosquitoes, convolutional neural network (CNN) models enhanced with log-Mel spectral features have proven effective in species classification (Kiskin et al., 2020). However, a recent literature review highlighted a limited focus on the auditory dimensions of mosquito control in urban settings, with only 9% (11 out of 122 articles) addressing this aspect (Joshi and Miller, 2021). Thus, this study aimed to address this gap by examining the sound-based characteristics of mosquitoes, thus substantially contributing to mosquito control through audio analysis.

Herein, we developed an automated system utilizing mosquito wingbeat frequencies to identify species within a genus, which is an innovative set to enhance the functionality of ovitraps. A crucial component of this system is designed to extract features from audio data, in which the Mel Frequency Cepstral Coefficient (MFCC) plays a critical role. The MFCC provides a crafted acoustic signature, especially beneficial with limited datasets (Ng, 2018). This system is poised to significantly advance the dengue vector surveillance capabilities of traditional ovitraps by harnessing the power of the Internet of Things (IoT) and cloud computing technologies (Agustiningtyas and Lusiyana, 2017; Hamesse et al., 2023; Juhdi et al., 2019; Martin et al., 2019; Ortega-Morales et al., 2018; Reed et al., 2019; Sasmita et al., 2021; Zuhriyah et al., 2016). Our study presents two novel mosquito classification systems, binary and multiclass. Using audio recordings of wingbeats from female and male *Aedes aegypti* and *Culex quinquefasciatus* mosquitoes, both systems have been rigorously tested.

Integrating MFCC with a CNN offers several advantages. MFCC effectively distills audio content by filtering out irrelevant characteristics such as speaker pitch and provides a robust feature set unaffected by noncontent elements (Aucouturier et al., 2007; Ghoraani and Krishnan, 2011; Ng, 2018). The system when coupled with a CNN, a deep learning algorithm capable of hierarchical data processing, can identify complex patterns within audio data, catering to frequency and time shifts (LeCun et al., 2015). This synergy between the MFCC and CNN enhances the robustness of the model to variations in sound and eliminates the requirement for manual feature extraction, enabling the algorithm to autonomously discern the relevant features (Abdel-Hamid et al., 2014; Deng and Yu, 2014). This potent combination has shown superior performance in speech recognition tasks, demonstrating particular efficacy in complex scenarios where traditional machine learning algorithms may fail (Abdel-Hamid et al., 2014; Hinton et al., 2012). Thus, by leveraging MFCC with a CNN, our model provides a deeper understanding of the acoustic features crucial for identifying mosquito species (LeCun et al., 2015).

This paper is organized to guide readers through the research process and findings. In Section 2, we outline the methodology, explore the theoretical underpinnings of processing wingbeat audio recordings, focus on the MFCC and introduce deep learning technologies particularly CNNs, and provide a detailed breakdown of the wingbeat frequency processing steps. In Section 3, we showcase the efficacy of our approach, examining and interpreting the outcomes of the two classification methods. Additionally, this section engages with broader scholarly discourse, situating our work within the context of existing literature. Section 4 reflects on the broader implications of our algorithms, assesses their potential impacts, and suggests avenues for future research on mosquito-borne disease prevention.

2. Material and methods

2.1. Designing the mosquito classification system

Deep learning, which utilizes audio recordings of mosquito wingbeats as input, was employed to classify the mosquito species. The model was developed through a training process using data gathered from recordings made using a smartphone placed inside plastic bottles containing mosquitoes (Fig. 1a). To capture clear sound recordings with minimal environmental noise, this method was implemented using an ovitrap with a smartphone to obtain test data in a laboratory setting (Fig. 1b). The specific recording location was 6°53′27.25″S, 107°36′37.36″E, Bandung 40132, Indonesia. Subsequently, audio processing involved two stages: manual preprocessing to prepare the data and automatic processing for feature extraction.

2.2. Fundamental theory

2.2.1. Aedes wingbeat frequency

The wingbeat frequency of *Aedes* mosquitoes has been previously demonstrated to be influenced by ambient temperature, with variations observed between 8 and 13 Hz for each degree of Celsius change (Santos et al., 2019; Villarreal et al., 2017). Additionally, humidity is crucial in modulating the wing flapping frequency (Fanioudakis et al., 2018). Notably, the microphone type used for data collection affects the measurement accuracy (Potamitis and Rigakis, 2016). Table 1 illustrates the recorded wingbeat frequencies of *Aedes aegypti* across various countries, highlighting their environmental effects (Santos et al., 2019).

Previous studies have used acoustic methodologies to measure the wingbeat frequency of *Aedes aegypti* mosquitoes, revealing variances based on the microphone type. The average wingbeat frequency of female *Aedes aegypti* was recorded at 511 Hz with a pressure-gradient microphone, compared to 480.6 Hz using an electric condenser microphone. Optical sensors have also been employed for their ability to measure wingbeats in larger mosquito populations (Santos et al., 2019). Furthermore, environmental factors such as temperature and humidity have been established to affect wingbeat frequency (Fanioudakis et al., 2018). Advancements in artificial intelligence have refined the classification of mosquitoes based on their wingbeat frequency using machine learning algorithms, providing more precise real-time results (Fanioudakis et al., 2018; Fernandes et al., 2021). Two primary approaches are used in the classification realm: similarity-based (computes a distance

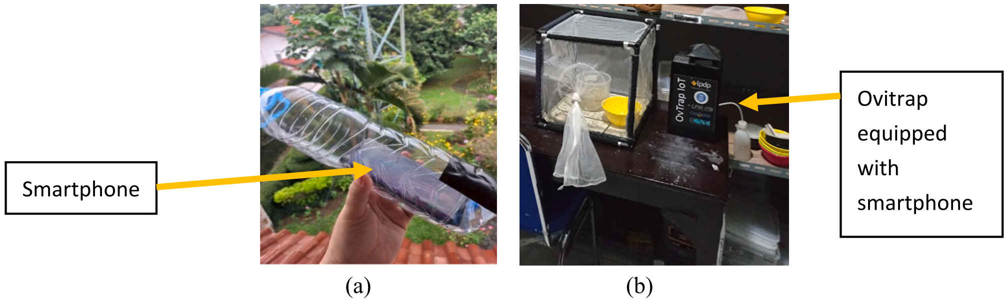

Fig. 1. Mosquito wingbeat recording process (a) training data (b) test data.

Table 1
The wingbeat frequency of the *Aedes aegypti* in several countries.

Weather	Wingbeat frequency	Country	Author
–	480.0 Hz	Germany	O. Sotavalta, 1952 (Sotavalta, 1952)
22.0 °C	508.0 Hz	USA	A. Moore, J. R. Miller, B. E. Tabashnik, S. H. Gage, 1986 (Moore et al., 1986)
23.0 °C	511.0 Hz	USA	B. J. Arthur, K. S. Emr, R. A. Wyttenbach, R. R. Hoy, 2014 (Arthur et al., 2014)
26.0 °C	480.6 Hz	UK	A. Aldersley, A. Champneys, M. Homer, D. Robert, 2014 (Aldersley et al., 2014)
32.6 °C	664.3 Hz	Thailand	L. J. Cator, B. J. Arthur, A. Ponlawat, L. C. Harrington, 2011 (Cator et al., 2011)

function among the data points) and feature-extraction-based (can be segmented into temporal, spectral, and cepstral categories). Temporal extraction is crucial for audio signal analysis, spectral extraction pinpoints insect wing frequencies, and cepstral features are derived based on the response of the human auditory system to sound (Silva et al., 2015).

2.2.2. Mel spectrogram

This study employs the Mel spectrogram for feature extraction, generated by applying a Fast Fourier Transform (FFT) to various windowed segments of an audio signal (Jiang et al., 2022). The acoustic profile of mosquitoes, similar to several animals, encompasses a broad spectrum of sound frequencies, including components beyond fundamental frequencies, such as those produced by wing flapping. The Mel spectrogram visually represents the amplitude of the sound signal over time across a range of frequencies. It is based on the Mel scale, constructed to match the human perception of pitch, with an equidistant spacing of pitches according to the listener's experience. The comparative reference between the Mel scale and standard frequency measurement is set by matching a 1000 Hz tone, perceived as 40 dB above the listener's threshold, to a pitch of 1000 Mels. At frequencies below approximately 500 Hz, the Mel and Hertz measurements closely align; however, as the frequency increases beyond this point, the Mel scale expands, indicating that human listeners perceive equal pitch increments at progressively larger frequency intervals (Rao and Manjunath, 2017; Stevens et al., 1937).

The Mel spectrogram, a variant of the traditional spectrogram, maps frequencies onto the Mel scale to further reflect human auditory perception. This approach is gaining traction as an essential feature for sound classification tasks because it encapsulates the temporal and spectral characteristics of the signal, which are crucial for identifying distinct sounds, such as insect wingbeats (Santos et al., 2019). The Mel spectrogram offers the distinct advantage of effectively capturing salient sound features necessary for tasks such as speech recognition while also producing a more compact dataset. It aligns with the human auditory system by applying a linear filter to frequencies <1000 Hz and a logarithmic filter above this threshold. Mel spectrogram generation involves a series of methodical steps that enable the transformation from raw sound waves to a format more suited for pattern recognition and classification tasks (O'Shaughnessy, 1987).

2.2.2.1. Pre-emphasis. Pre-emphasis refers to the filtering that emphasizes higher frequencies. The purpose is to balance the spectrum of voiced sounds with a steep roll-off in the high-frequency region. The most commonly used pre-emphasis filter is given by the following transfer function:

$$H(z) = 1/\left(1 - az^{-1}\right) \tag{1}$$

where the a value controls the slope of the filter and is usually between 0.4 and 1.0.

2.2.2.2. Frame blocking. The analysis of speech signals necessitates examination within brief time intervals to ensure quasi-stationarity of the signal. Speech analysis should be conducted on these short segments to maintain the assumption of stationarity. Typically, speech's spectral measurements are performed using windows of 20 ms duration, with a stride of 10 ms between successive windows. The rationale for this overlapping approach is to position each speech element within the center of a window frame, ensuring that the signal is symmetrically weighted and tapered at the edges to minimize discontinuities. The application of a window function, typically Hanning or Hamming, emphasizes harmonic components and mitigates spectral leakage during the discrete Fourier transform (DFT) process. These windows are favored for their ability to smooth transitions and diminish the "edge effect," leading to a more accurate spectral representation of the speech signal.

2.2.2.3. DFT and Mel spectrum. By applying the DFT and considering N number of points used to compute the DFT, each windowed frame was converted into a magnitude spectrum. The Mel spectrum is calculated by passing a signal that has been Fourier transformed through a set of bandpass filters known as a Mel filter bank. The Mel is a measurement unit based on the frequency perceived by the human ear. The approximation of Mel for the physical frequency can be written as

$$f_{Mel} = 2595 \times log_{10}\left(1 + \frac{f}{700}\right) \tag{2}$$

where f is the physical frequency in Hz and f_{Mel} is the perceived frequency. The Mel spectrum of the $X(k)$ magnitude spectrum was calculated by multiplying the magnitude spectrum with each triangular Mel-weighted filter.

$$s(m) = \sum_{k=0}^{N-1} \left[|X(k)|^2 H_m(k) \right], 0 \leq m \leq M - 1 \tag{3}$$

where M is the total number of triangular Mel-weighted filters. $H_m(k)$ is the weight assigned to the k-th energy spectrum container contributing to the m-th output band and is expressed as follows:

$$H_m(k) = \begin{cases} 0, & k < f(m-1) \\ \dfrac{2(k - f(m-1))}{f(m) - f(m-1)}, & f(m-1) \le k \le f(m) \\ \dfrac{2(f(m+1) - k)}{f(m+1) - f(m)}, & f(m) < k \le f(m+1) \\ 0, & k > f(m+1) \end{cases} \tag{4}$$

with m ranging from 0 to M–1.

2.2.2.4. Discrete cosine transform (DCT). A Discrete Cosine Transform (DCT) is used to process the Mel frequency coefficients, yielding a series of cepstral coefficients. Traditionally, before applying the DCT, Mel spectra are logarithmically represented. This transformation results in a cepstral signal, where the prominent peak correlates with the pitch of the original audio signal, and additional peaks, known as formants, represent resonant frequencies. Notably, the most significant information in the signal is typically encapsulated within the first several MFCCs. The robustness of the system was enhanced by focusing on extracting these initial coefficients and disregarding the higher-order components. Conventionally, only the first 8–13 cepstral coefficients are used. The zeroth coefficient is often omitted because it denotes the average log energy of the signal, which generally contains minimal information specific to the speaker.

2.2.2.5. Dynamic MFCC features. Given that the cepstral coefficient contains only information from a particular frame, it is usually referred to as a static feature. The first and second derivatives of the cepstral coefficients were calculated to obtain more information on the temporal dynamics of the signal (Sahidullah and Saha, 2012). The first-order and second-order derivatives are called the delta and delta-delta coefficients, respectively. The delta coefficient contains information on speech rate, whereas the delta-delta coefficient provides information similar to speech speed. The equations used to calculate the dynamic parameters are as follows.

$$\Delta C_{m(n)} = \frac{\sum_{i=-T}^{T} k_i c_m(n+i)}{\sum_{i=-T}^{T} |i|} \tag{5}$$

where $C_{m(n)}$ represents the m-th feature for the n-th time frame, k_i is the i-th weight, and T is the number of consecutive frames used for computation. Generally, T is set to two. The delta-delta coefficient was calculated using its first-order derivative.

MFCCs are prominent audio features that have been demonstrated to have good performance (Aucouturier et al., 2007; Ng, 2018). MFCC-based features have been used for sound analysis in several cases, ranging from the sounds of humans and the environment to those of animals. Mosquito sounds are similar to those of other animals in terms of their wide spectral range. In addition to the fundamental frequency, the sound spectrum comprises several other frequencies. The MFCC is a method for extracting frequency-based features from the sound spectrum (Ghoraani and Krishnan, 2011).

2.2.3. Deep learning

Once transformed into a matrix or vector, the results of voice feature extraction serve as the input for a deep learning algorithm. This algorithm discerns patterns within the data, enabling it to make predictions based on these learned patterns, all without the need for explicit programming (França et al., 2021; Navamani, 2019). A critical distinction between deep learning and traditional machine learning is the feature-extraction process during classification. Feature extraction from data typically requires manual or human intervention in traditional machine learning. Contrastingly, deep learning models are designed to automatically perform feature extraction.

The structure of a deep learning algorithm or simple neural network comprises an input layer, several hidden layers, and an output layer (Fig. 2). Each neuron within these layers has its own weight and bias coupled with an activation function to introduce nonlinearity. During the forward and backward propagation phases, the values of these weights and biases are iteratively adjusted. The initial data or input must be converted into matrix format and processed by the network. Each matrix element is subsequently manipulated according to the weights and biases of the neural network. In this context, we used a robust deep learning framework available in Python, which is known for its efficacy in handling complex calculations.

2.2.4. Convolutional network

CNN represents an algorithm type or neural network model commonly used to process two or more dimensions of data with high complexity, such as images or sound data (Fernandes et al., 2021). A CNN consists of several parts, one of which is the feature extraction layer and the other is the fully connected layer (Albawi et al., 2017; Kim et al., 2018; Navamani, 2019). In the feature extraction section, several combinations of convolution layers, activation functions, and pooling are used to extract features from the given input data. In the fully connected layer, several matrix-merging processes and output mapping occur in the SoftMax activation function layer. Furthermore, the classification process occurs from the data processed in the feature extraction section.

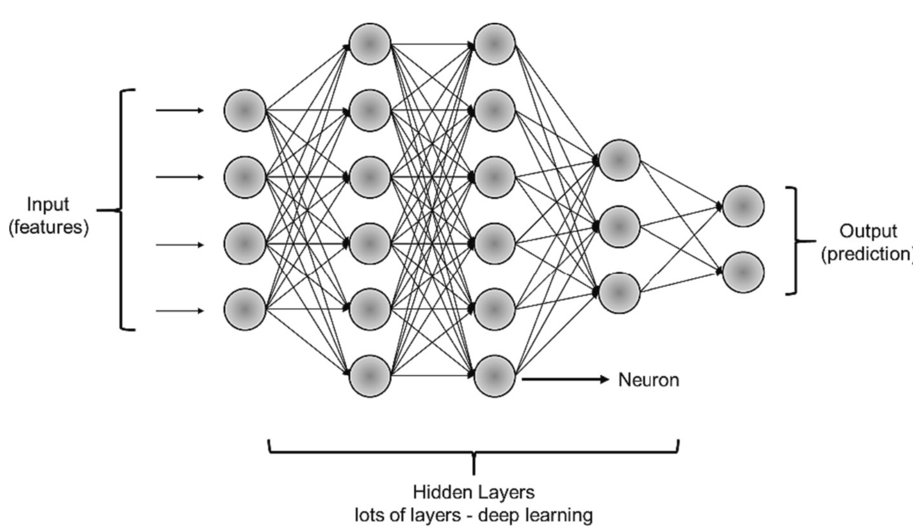

Fig. 2. Model of deep learning layer.

In a CNN, a convolution layer performs convolution operations between the pixels on the input and several kernels or filters. The mathematical equation of the convolution process is as follows, where f is the input data and h is the kernel size.

$$G[m,n] = (f^{*}h)[m,n] = \sum_{j}\sum_{k} h[j,k]f[m-j,n-k] \quad (6)$$

The filter moves to sweep all input pixels with a shift operation, with parameters such as stride and padding defining the movement. A parameter that sets the number of filter shifts is the stride. If the stride value is one, the convolution filter shifts by one pixel in the vertical and horizontal directions. Padding is a parameter that determines the number of pixels containing a value of zero added to each side of the input. Fig. 3 illustrates the identification process of the system.

In the CNN, feature extraction is automatically performed through a convolution process that applies a convolution kernel to the input data and pooling, which progressively reduces the spatial size of data representation (Fernandes et al., 2021). Using convolution layers, pooling, and the parameter-sharing feature also makes CNN effective and efficient in classifying data, such as images and sound recordings (Kim et al., 2018). When processing voice-recorded data, a visual representation of the sound in the form of a spectrogram is generated (Thakur and Kumar, 2021). Therefore, as with image data, a CNN can be used to map patterns from a given sound recording data (a mosquito voice recording in this case). Additionally, it can produce considerably higher accuracy than other deep learning algorithms.

2.3. Implementation of the Mel spectrogram feature extraction

Audacity was used for manual audio processing, which included cutting audio duration, noise reduction, and converting files into the WAV format, to increase the amount of data and eliminate environmental noise. Feature extraction was performed to obtain a Mel scale spectrogram feature representation of the sound. The feature representation acts as an input to the CNN architecture, which studies the data patterns for each mosquito species (Fernandes et al., 2021). Fig. 4 a, b, and c illustrate the wingbeat spectrogram for female *Aedes aegypti*, male *Aedes aegypti*, and *Culex quinquefasciatus*, respectively.

Fig. 5 illustrates a flowchart for the mosquito classification to obtain the mosquito type. Feature extraction was performed on the audio data (Fig. 6). The audio samples were prepared with a predetermined window size, and data augmentation was performed with an overlapping frame of 50% to ensure no features were wasted due to the windowing process. Each audio sample was calculated via a spectrogram using an FFT with several parameters, including the number of bands, number of frames, hop length, and length of the FFT window. The parameter values used were optimized based on model performance with a combination of parameters, similar to that used in a previous study (Fernandes et al., 2021), as follows:

Number of bands: 20, 40, 60, and 80
Number of frames: 20, 40, and 60
Hop length: 64, 128, 256, and 512
Length of the FFT window: 512, 1024, and 2048.

The spectrogram log form, normalized to the range of −80 to 0 dB, was obtained by changing the value on the Mel spectrogram. The results of feature extraction were obtained using a spectrogram log in the form of an array with dimensions of 1 × bands × frames as the input for CNN model training (Fernandes et al., 2021).

The recorded audio samples of the mosquitoes were cut before entering the FFT process using a windowing process with overlapping frames. The number of chunks was determined from the hop length and number of frame parameters. After going through the cutting process, FFT process, and conversion to the Mel scale, the amplitude or matrix values in this result were subsequently converted to a logarithmic scale because the value was extremely small. Subsequently, a log feature representation of the spectrogram was obtained. Fig. 6 illustrates the sizes of the input and output data for each process.

2.4. Architectural design of the CNN

The CNN architecture consists of feedforward networks in the form of a convolution layer and a pooling layer, which are sequentially arranged. In this case, the ReLU activation function of the convolution layer was used for binary and multiclass classifications. In binary classification, three convolution layers are used with 3 × 3 kernels and two max-pooling layers with 2 × 2 kernels, and the output was continued for flattening through a fully connected layer of 256 neurons with a ReLU activation function and two neurons with a sigmoid activation function. Meanwhile, in multiclass classification, two convolutional layers with 20 × 5 and 8 × 4 kernels were used, and two max-pooling layers with 2 × 2 kernels subsequently went through a fully connected layer of 3 neurons with the SoftMax activation function. Fig. 7 illustrates the CNN architecture used for mosquito classification.

The CNN model was trained using epochs and batch sizes of 10 and 32, respectively. This model was compiled using a categorical cross-entropy loss function, and the hyperparameters used in the CNN algorithm were tuned using a root mean square propagation (RMSProp) optimizer. The model generated in the training process was evaluated using stratified 10-fold cross-validation with calculated values of accuracy, precision, recall, F1-Score, area under the curve (AUC) score, confusion matrix, and model training time for binary and multiclass classifications (Forman and Scholz, 2010; Grandini et al., 2020; Ouyang et al., 2015). The model performance needs to be evaluated to determine the model's ability to classify mosquitoes, and a good model is characterized by high values of predetermined performance parameters.

The model parameters were tuned to increase model accuracy. Some model parameters or hyperparameters were varied to evaluate the performance of the given model. The model with the best performance was selected from several other models with different parameter variations. However, parameter tuning was limited to a few hyperparameters because of the large number of parameters in the CNN model. The model parameters that varied were the optimization algorithm or optimizer, learning rate, and dropout layer. The following are

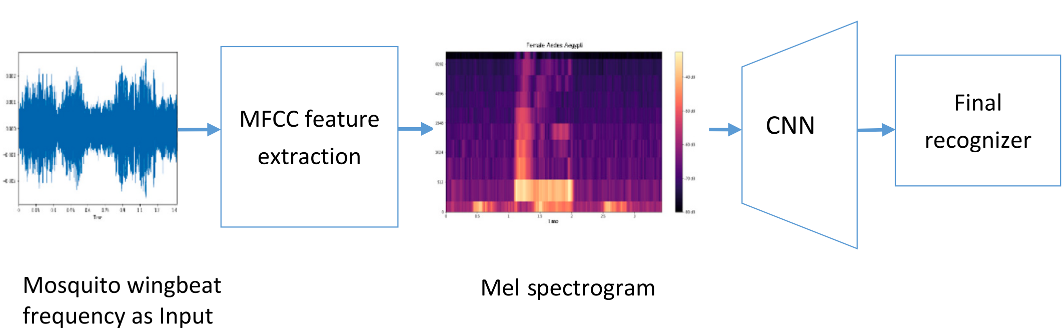

Fig. 3. Schematic of the process that occurs in the convolutional neural network algorithm.

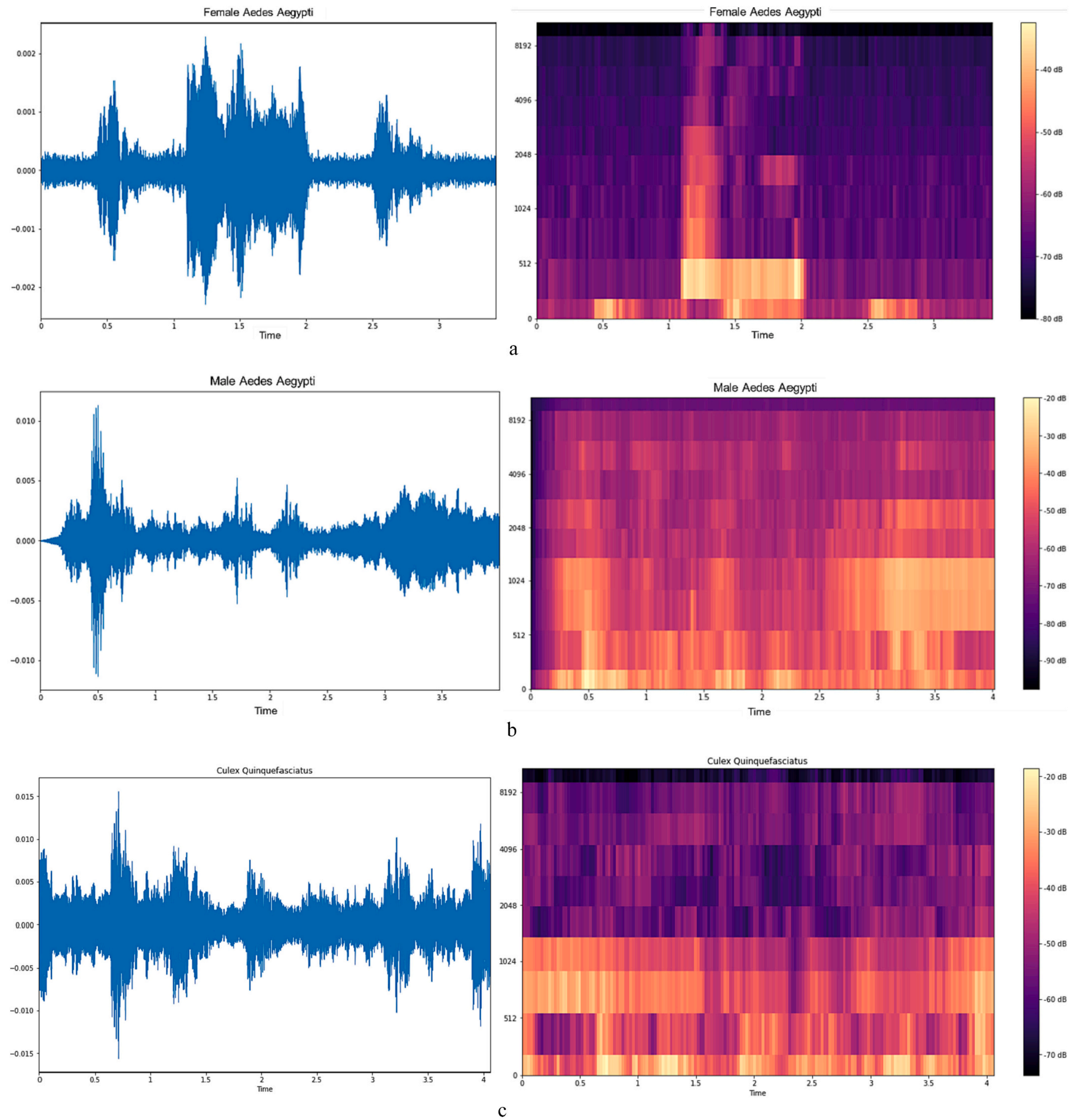

Fig. 4. a. Sound and Mel spectrogram of female *Aedes aegypti*.
b. Sound and Mel spectrogram of male *Aedes aegypti*.
c. Sound and Mel spectrogram of *Culex quinquefasciatus*.

the selected parameters that were combined and used in the model:

1. Optimizer: Stochastic gradient descent (SGD), root mean square propagation (RMSProp) and Adaptive Moment Estimation (Adam).
2. Learning rate: 0.001 and 0.01.
3. Dropout: Used and not used.

The recorded audio file required to be processed first considering that it needed to be used to predict the mosquito type. The audio file for mosquito recording in this test was obtained from the recording using a cell phone placed in the ovitrap, which was positioned inside a mosquito cage.

Subsequently, the audio was automatically processed, including converting the audio into WAV format and extracting the Mel-scaled spectrogram features. Predictions were made using the best deep learning model from the 10 stored training processes. The prediction results were the mosquito types predicted by the model along with the percentage of each mosquito type. For each predicted mosquito species,

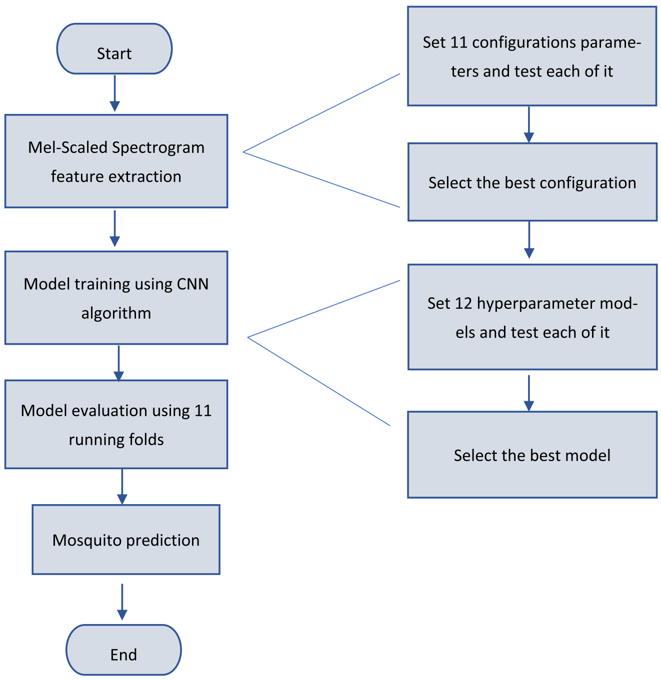

Fig. 5. Deep learning flowchart for mosquito classification.

the results were added to the number of detected mosquitoes within the predicted species.

3. Results and discussion

Audio was recorded using a mobile device placed in a 1.5-L plastic bottle to minimize the presence of other mosquitoes in the environment at the time of recording. The training and test data were collected in different temporal settings, with the training data obtained at an average temperature of 23.1 °C and an average humidity of 66.3%. In contrast, the test data was obtained at an average temperature of 22.7 °C and an average humidity of 82%. A dataset of 436 audio recordings was manually processed, 172, 105, and 159 recordings of female *Aedes aegypti*, male *Aedes aegypti*, and *Culex quinquefasciatus*, respectively. For the test data, 85 data points were selected from each recording class.

Optimization of the feature extraction parameters for binary classification (two classes) and multiclass classification (three classes) was conducted by considering a set of 11 configuration parameters (Table 2). The parameter values were contingent on the outcomes of the initial optimization of the first parameter. Furthermore, model optimization was conducted by employing a composite of 12 models (Table 3).

The purpose of these two optimizations was to improve the performance of the CNN deep learning model in classifying mosquito species. This is because selecting the correct parameters produces a CNN model with improved performance. All previous audio recordings had a 22,050 Hz sample rate and were used as inputs for deep learning optimization for mosquito classification, including binary and multiclass classification.

3.1. Binary classification

A binary classification was used to divide the mosquitoes into two classes *Aedes aegypti* and *Culex quinquefasciatus*. A total of 277 *Aedes aegypti* recordings and 159 *Culex quinquefasciatus* recordings were used for audio data analysis. Fig. 8 depicts the performance test results for each arrangement.

The standard deviation values for each performance indicator were compared to determine the optimal configuration (Table 4). Configuration 8 was determined as the ideal parameter (number of bands: 60, number of frames: 40, hop length: 256, and length of the FFT window: 1024) owing to its reduced standard deviation value. The feature extraction parameters that yielded the best performance were employed during the feature extraction phase in the optimization process of the CNN deep learning model. The model optimization outcomes were derived using performance criteria identical to those employed to optimize feature extraction parameters. Fig. 9 illustrates the performance evaluation of each model. Additionally, Model 1 exhibited the most favorable average performance outcome, with an 89.43% accuracy rate, 91.38% precision, 95.52% recall, and 0.816 F1-Score. The obtained results demonstrated an improved optimality level compared to the alternative models. Consequently, Model 1 that lacked dropout, used the RMSprop optimizer, and had a 0.001 learning rate is the most optimal deep learning model.

Figs. 10 and 11 illustrate the performance of Model 1 in each running fold of model training. The first and tenth running fold models performed well. Compared with the first running fold, the tenth running fold model had greater Recall and F1-Scores (96.46% and 0.9497, respectively). We chose Model 1 in the tenth running fold as the best model for evaluating ovitraps in binary classification. Subsequently, mosquito audio data from the recordings were used to test the optimal binary classification model. Table 5 presents the confusion matrix of the prediction results from the best model on 85 audio recordings for each mosquito class or type. From the results in the table, the performance measure can be remade with the formula of accuracy, precision, Recall, F1-Score, weighted accuracy, and micro F1 (Grandini et al., 2020) to determine how well the model works on audio data recorded by mosquitoes in the ovitrap. Table 6 indicates the model's *Aedes aegypti* mosquito detection accuracy of 91.76% and an F1-Score of 0.9230.

3.2. Multiclass classification

In the realm of multiclass categorization, mosquitoes are categorized into three distinct groups: male *Aedes aegypti*, female *Aedes aegypti*, and both male and female *Culex quinquefasciatus*. This study used 172, 105, and 159 audio data samples from female *Aedes aegypti* mosquitoes, male *Aedes aegypti* mosquitoes, and *Culex quinquefasciatus* mosquitoes, respectively. Following the acquisition of the outcomes of the optimization process for the feature extraction parameters, preliminary test findings revealed that Configurations 9 and 11 exhibit the highest average performance values (Fig. 12).

The standard deviation values for each performance indicator were compared to determine the ideal configuration (Table 7). Based on the observation that configuration 9 had a decreased standard deviation value, the optimal parameter combination was identified as follows: number of bands, 60; number of frames, 60; hop length, 512; and length

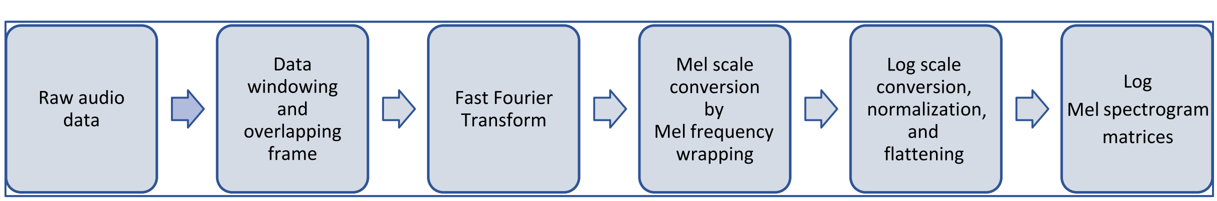

Fig. 6. Feature extraction flowchart.

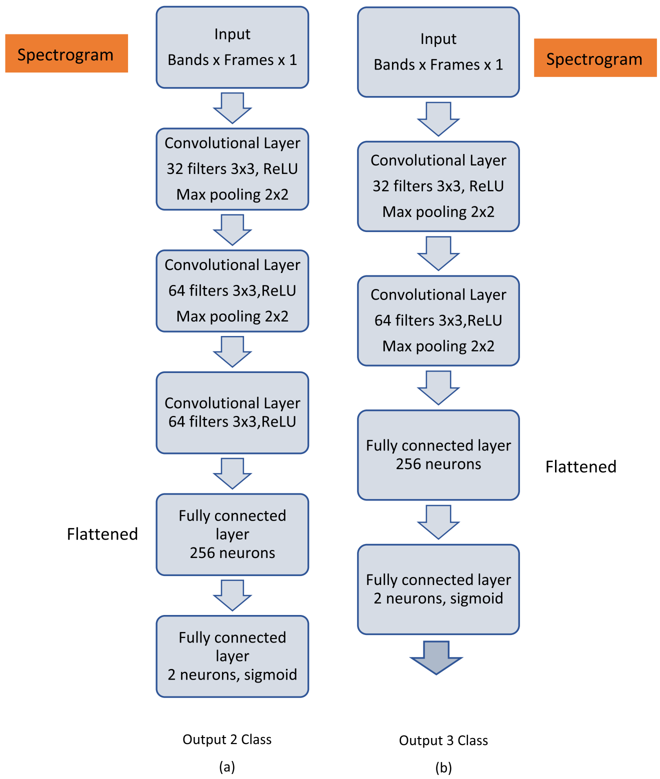

Fig. 7. Convolutional neural network architecture for (a) binary classification and (b) multiclass classification.

Table 2
Configuration of the feature extraction parameters for deep learning.

No	Number of Bands	Number of Frames	Hop Length	FFT Window Length
1	20	40	128	1024
2	40	40	128	1024
3	60	40	128	1024
4	80	40	128	1024
5	20/40/60/80	20	128	1024
6	20/40/60/80	60	128	1024
7	20/40/60/80	20/40/60	64	1024
8	20/40/60/80	20/40/60	256	1024
9	20/40/60/80	20/40/60	512	1024
10	20/40/60/80	20/40/60	64/256/512	512
11	20/40/60/80	20/40/60	64/256/512	2048

Table 3
Combination of the convolutional neural network deep learning model parameters.

No	Model	Optimizer	Learning Rate
1	Without Dropout	RMSProp	0.001
2			0.01
3		Adam	0.001
4			0.01
5		SGD	0.001
6			0.01
7	With Dropout	RMSProp	0.001
8			0.01
9		Adam	0.001
10			0.01
11		SGD	0.001
12			0.01

of the FFT window, 1024. The feature extraction parameters that yielded the best performance were employed during the feature extraction phase in optimizing the CNN deep learning model. The results of model optimization were acquired by employing identical performance metrics used in optimizing the feature extraction parameters (Fig. 13).

The findings indicate that Model 9 exhibits the highest average performance, with an 89.30% accuracy rate, 87.22% precision rate, 84.62% recall rate, and 0.8587 F1-Score. The obtained result surpasses that of the other models, indicating that Model 9, which incorporates dropout, Adam's optimizer, and a 0.001 learning rate, is the most optimal deep-learning model. The ovitrap test employed Model 9, which had the highest running fold and demonstrated the most optimal performance value. Figs. 14 and 15 indicate Model 9's performance at each iteration of the training process.

The model demonstrating the best performance was observed in the first and tenth running folds. A comparison between the two reveals that the recall performance metrics and the F1-Score in the tenth running fold model exhibited higher values, specifically 90.85% and 0.9153, respectively, compared to the first running fold. Based on these findings, model 9, specifically in the tenth iteration of the cross-validation process, was determined to be the ideal choice for ovitrap testing in multiclass classification.

Additionally, the efficacy of the ideal multiclass classification model was assessed using audio data recorded from mosquitoes in an ovitrap. The experiment employed the most efficient model on a dataset with 85 audio recordings for each mosquito class and type. The outcome of the experiment was a confusion matrix that displayed the prediction results (Table 8). Table 9 displays the outcomes of the model's performance in identifying different mosquito types (Fig. 16). The average accuracy rate and average F1-Score were 87.16% and 0.8167, respectively.

3.3. Further discussion

Ecologists often grapple with accurately categorizing temporal data into predefined classes, such as biological entities or ecological states. A common approach involves converting time-series data into user-defined features for use as predictors in conventional statistics (da Silva et al., 2023; Fisher et al., 2018; Peterson et al., 2008) or machine learning models (Capinha et al., 2021; Ghani et al., 2023; Lee et al., 2016; Nolasco et al., 2023). Deep learning models, a subset of machine learning, employ artificial neural networks to emulate the learning process of the human brain and have gained popularity for classification. Deep learning techniques enable direct classification from time series, eliminating subjective and resource-consuming data transformation steps (Ceia-Hasse et al., 2023). This has the potential to improve classification results, such as in insect species identification from wingbeat spectrograms (Capinha et al., 2021). This method utilizes wingbeat spectrograms and various deep learning modeling frameworks. The InceptionTime architecture achieved the highest accuracy with a throughput of 85% and the CNN model reached 70%.

A similar study employed two lightweight deep learning models for mosquito species and sex classification based on wingbeat audio signals (Yin et al., 2021). The 1D CNN and the combination of 1D CNN with LSTM models directly operate on raw audio signals at low sample rates without preprocessing. They achieved a remarkable classification accuracy rate of >93% for a dataset of male and female mosquitoes from five mosquito species. Furthermore, fine-tuning the model reduced its size by approximately 60% with only a slight 3% decrease in the classification accuracy.

Herein, a mosquito classification system was designed using wingbeat frequency and subsequently transformed with MFCC into a Mel spectrogram as the input to the CNN model. Additionally, two classification system models were proposed: binary and multiclass. The binary system distinguishes two mosquito types: *Aedes aegypti* and *Culex quinquefasciatus*. The best result was achieved with extraction feature configuration no. 8 with the number of bands: 60, number of frames: 40,

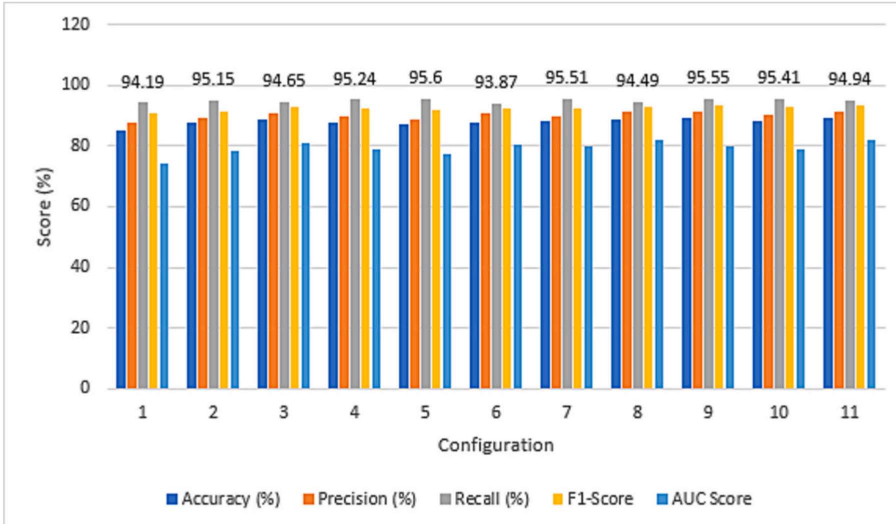

Fig. 8. Performance test of each configuration.

Table 4
The standard deviation values for each performance metric.

Configurations	Accuracy (%)	Precision (%)	Recall (%)	F1-Score	AUC Score
8	0.8356	1.4505	2.2526	0.006	0.0424
11	1.2006	1.973	3.2649	0.00889	0.0302
Length of the Optimal FFT Window = 1024					

hop length: 256, length of the FFT window: 1024, and CNN model no. 1, which used the following parameters: without dropout, RMSprop Optimizer, and a learning rate of 0.001. The accuracy score was 91.76% and the precision, recall, and F1-Scores were 86.59%, 98.82%, and 0.9230, respectively.

The multiclass mosquito classification system categorizes three mosquito types: female *Aedes aegypti*, male *Aedes aegypti*, and *Culex quinquefasciatus*. Optimal results were obtained from the extraction feature configuration no. 9 with number of bands: 60, number of frames: 60, hop length: 512, length of the FFT window: 1024, and CNN model no. 9, where the following parameters were used: dropout, Adam's optimizer, and a 0.001 learning rate. The average accuracy was 87.16%, and the precision, recall, and F1-Scores were 89.09%, 77.23%, and 0.8167, respectively. Thus, the MFCC and CNN combination led to higher accuracy compared with that described in two previous studies (Capinha et al., 2021; Yin et al., 2021).

Fig. 17 illustrates the proposed wingbeat recognition system. Implemented in a specially designed Ovitrap, it incorporates the internet of things (IoT) for communication, a cloud server for computing features, and utilizes a smartphone for recording mosquito wing flapping. Smartphone technologies simplify the IoT architecture, reducing the required components. The developed wingbeat data collection system is simpler than that proposed in a previous study (Yin et al., 2023). The mosquito classification algorithm was executed on a cloud server. Furthermore, applying visual identification between *Aedes aegypti* and *Culex quinquefasciatus* mosquito eggs in an ovitrap using IoT has previously been developed (Gunara et al., 2023). A powerful model of mosquito control in urban environments requires available geospatial, audio, and visual data, rather than one data source (Joshi and Miller, 2021).

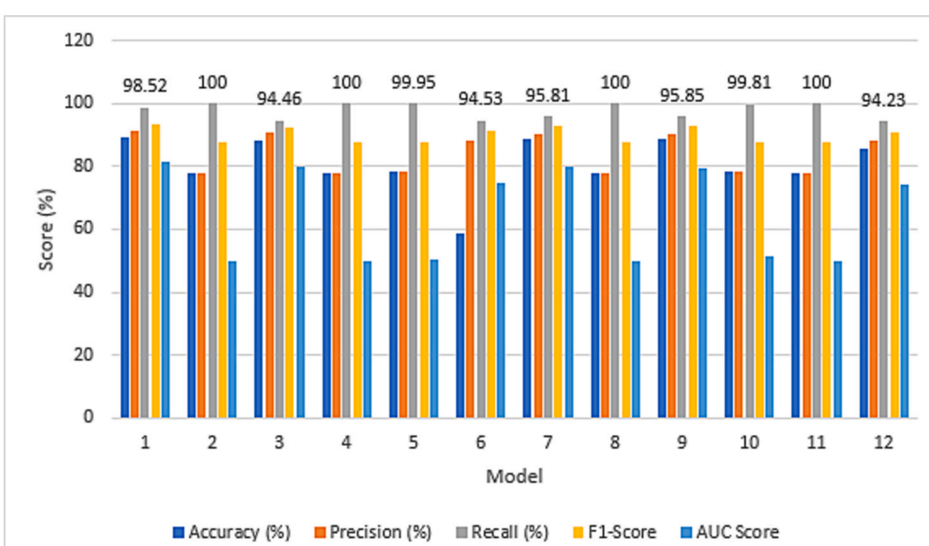

Fig. 9. Model optimization results.

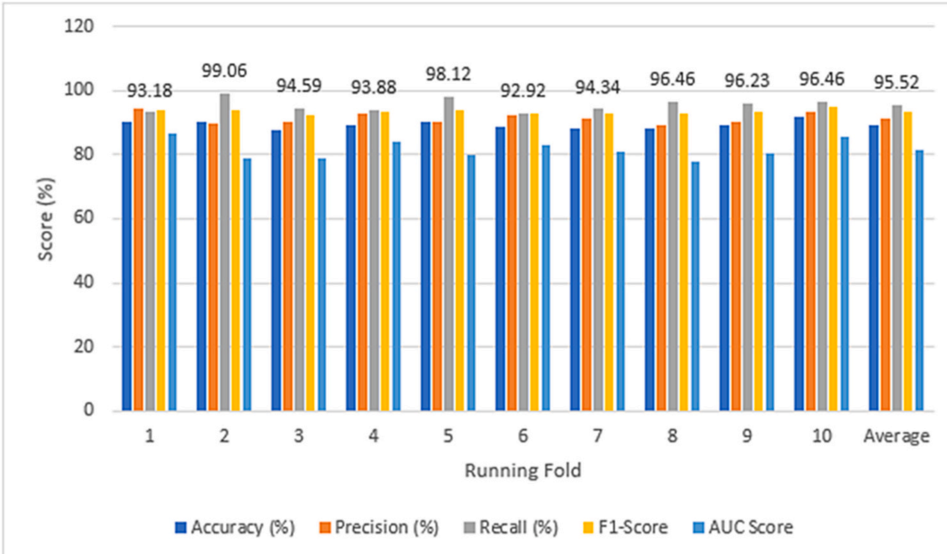

Fig. 10. The performance of Model 1 In each Running Fold of the model training for binary classification.

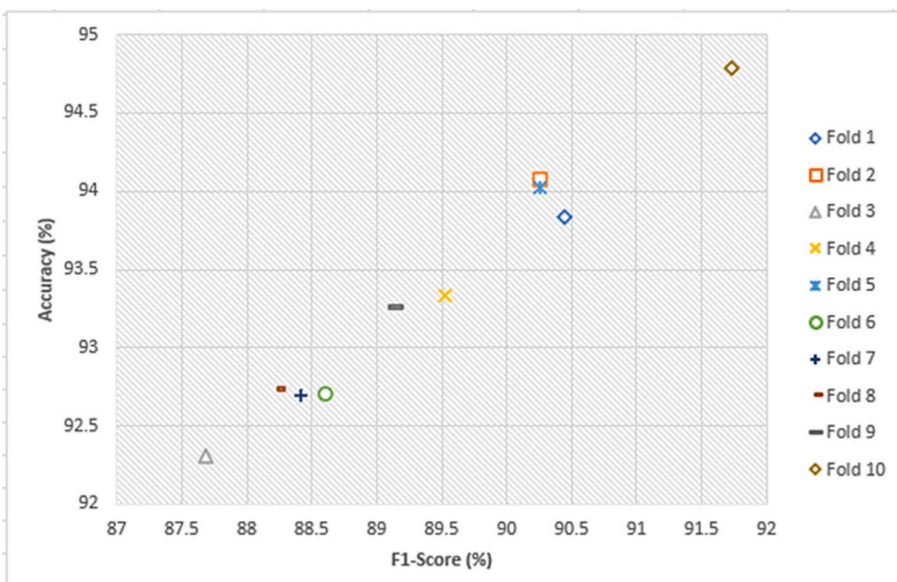

Fig. 11. Scatterplot of the Model 1's performance on each Running Fold for binary classification.

Table 5
Confusion matrix of the binary classification test.

	Aedes aegypti	*Culex quinquefasciatus*
Aedes aegypti	84	13
Culex quinquefasciatus	1	72

Table 6
Model performance on the binary classification test.

Accuracy (%)	Precision (%)	Recall (%)	F1-Score
91.76	86.59	98.82	0.9230

Successfully identifying mosquitoes from audio data is pivotal for maintaining vector surveillance efficacy and comprehensive disease prevention strategies. This breakthrough has the potential to dramatically transform public health measures by facilitating the enhanced early detection and monitoring of vector-borne diseases. These capabilities promise to reduce transmission rates, improve response times, and ultimately save lives, marking a substantial leap in global health and disease prevention efforts.

In addition to the immediate impact on public health, the broader implications of this study extend to various ecological and biological domains. The developed techniques could revolutionize bird-song identification, offering ornithologists more effective tools for avian population monitoring. Moreover, these methods have potential significance in amphibian population studies and marine biology, providing noninvasive tools for nuanced cetacean sound classification. This cross-disciplinary application contributes to the protection and understanding of vital species, opening new avenues for ecological research and conservation efforts. Taken together, these advancements promise a future where public health and environmental conservation benefit from the audio classification techniques.

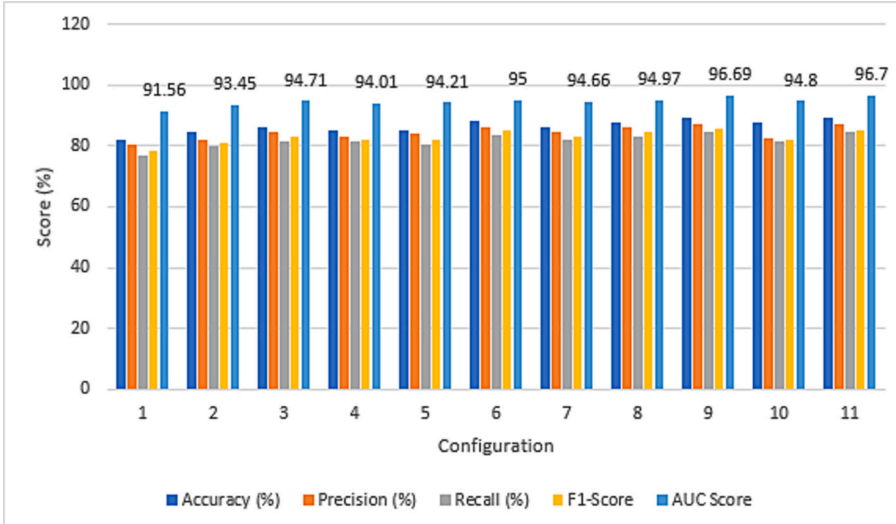

Fig. 12. Optimization of the feature extraction parameters.

Table 7
The standard deviation values for each performance metric.

Configurations	Accuracy (%)	Precision (%)	Recall (%)	F1-Score	AUC Score
9	1.5956	1.3839	2.4915	2.0305	0.0117
11	1.6132	1.59	2.9082	2.0401	0.007
Length of the optimal Fast Fourier Transform window = 1024					

4. Conclusions

Herein, a mosquito classification system was proposed using wingbeat frequency as the input and a CNN as the classification model. The proposed classification system was divided into two models: binary and multiclass. The binary mosquito classification system classified two mosquito types: *Aedes aegypti* and *Culex quinquefasciatus*. The multiclass mosquito classification system classified three mosquito types: female *Aedes aegypti*, male *Aedes aegypti*, and *Culex quinquefasciatus*. The performance of the binary classification system was superior to that of the multiclass classification system. This was because the algorithm worked to classify male and female *Aedes aegypti* mosquitoes of the same species with similar wingbeat frequencies in a multiclass system. Consequently, the performance of the *Culex* mosquito classification was not as optimal as that of the binary classification. The developed algorithm was able to provide a high confidence level in identifying *Aedes aegypti* mosquitoes against *Culex*. However, the algorithm's ability to identify *Culex* was low. Based on the high identification results for *Aedes aegypti* (female or male), a multiclass algorithm can be designed so that the application of identification is *Aedes aegypti* or other mosquito species (with the ability to identify *Culex quinquefasciatus* mosquitoes). The mosquito classification system was implemented in an ovitrap equipped with a smartphone, IoT, and cloud server. The use of new technologies has enabled the development of a simple system with a high computing performance capability for faster complex calculations.

Understanding the complexities of combining male and female *Aedes aegypti* necessitates more intensive research, particularly focusing on developmental models to discern these subtle differences. Herein, a careful approach is required to apply data augmentation to increase model variation and generalization, especially in the multiclass classification. Further exploration of the dropouts is necessary to enhance the robustness of the model against overfitting. Identifying the optimal dropout rate will have a positive impact on the modeling. Designing a

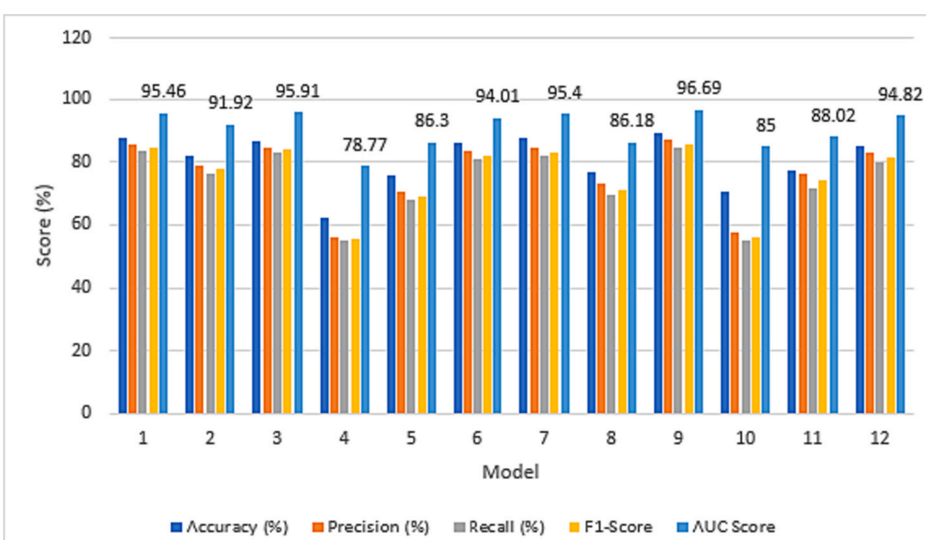

Fig. 13. The model optimization results.

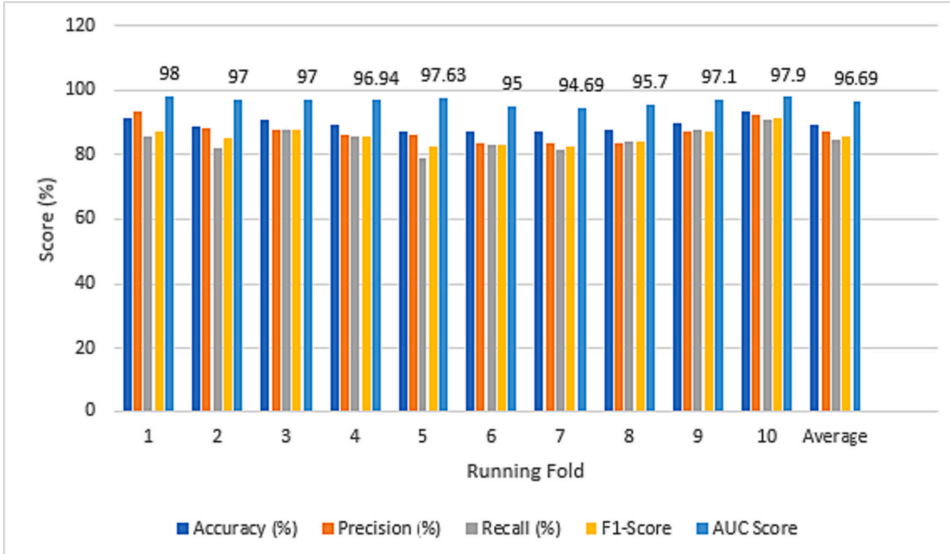

Fig. 14. The performance of Model 9 in each Running Fold for multiclass classification.

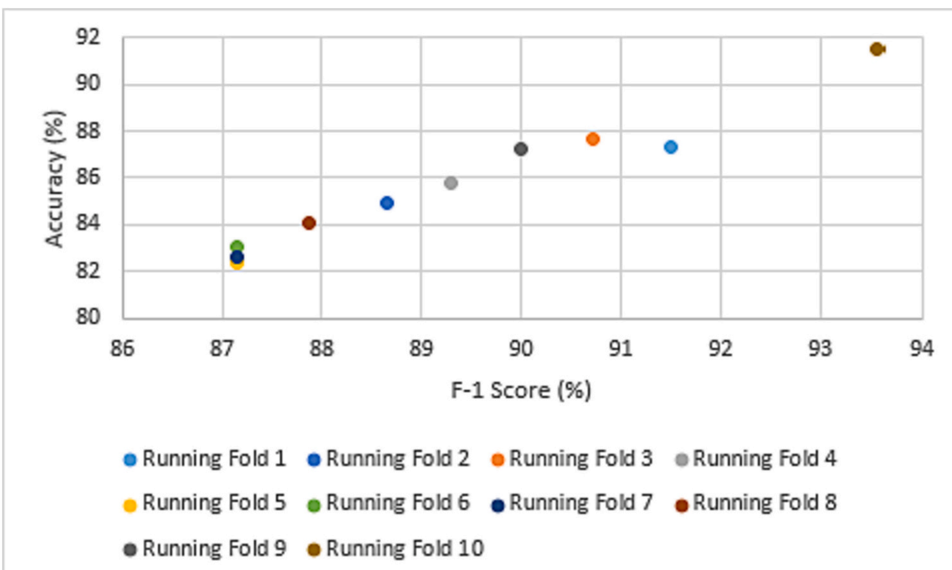

Fig. 15. Scatterplot of the performance of Model 9 in each Running Fold.

Table 8
Confusion matrix of the multiclass Classification Test.

	Female *Aedes aegypti*	Male *Aedes aegypti*	*Culex quinquefasciatus*
Female *Aedes aegypti*	84	8	15
Male *Aedes aegypti*	1	79	9
Culex quinquefasciatus	0	0	61

Table 9
Model performance on the multiclass classification test.

Metrics	Mosquito species			Average
	Female *Ae. aegypti*	Male *Ae. aegypti*	*Culex quinquefasciatus*	
Accuracy (%)		87.16		
Precision (%)	78.05	88.76	100.00	89.09
Recall (%)	89.36	77.45	64.89	77.23
F1-Score (%)	83.58	82.72	78.70	81.67
Micro F1		87.16		
Balanced Accuracy		87.84		

multiclass application capable of identifying *Aedes aegypti* or other mosquito species, including *Culex quinquefasciatus*, will open new opportunities for monitoring and controlling mosquito-borne diseases.

This achievement is crucial in the context of mosquito-borne diseases, as accurately classifying mosquito species is essential for implementing effective vector control strategies. In broader terms, the successful development of a robust classification system has implications for the wider field of disease prevention and control. Precise mosquito classification aids in comprehending and monitoring disease vectors, facilitating timely and targeted interventions to reduce the spread of diseases, such as dengue fever, filariasis, and Zika.

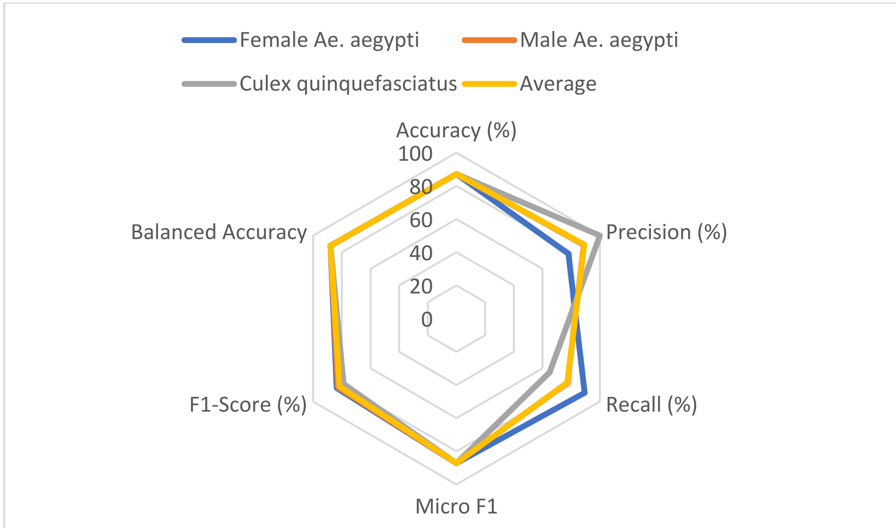

Fig. 16. Model performance in a multiclass classification test.

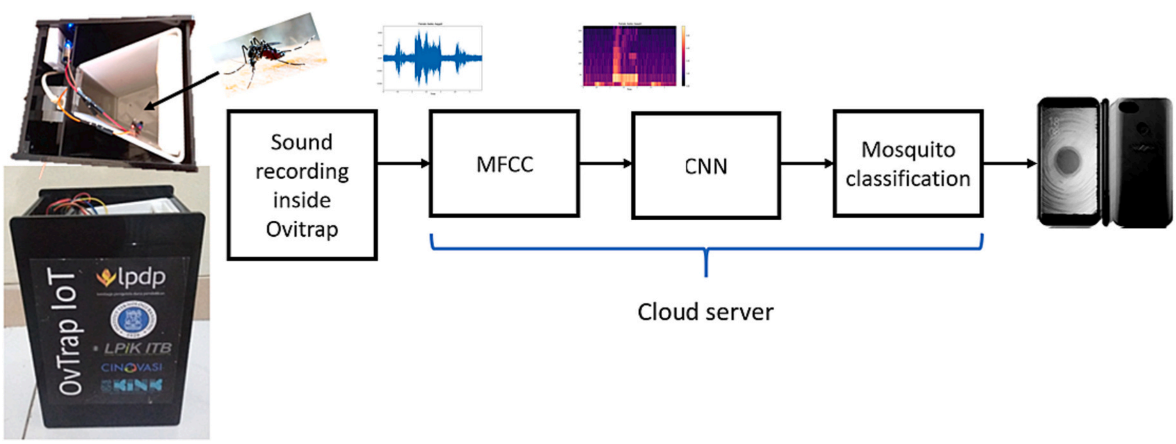

Fig. 17. Mosquito wingbeat classification system in the internet of things (IoT) enabled ovitrap with cloud server.

Although the results are promising, being aware of potential uncertainties and limitations is crucial. Factors such as dataset variability, environmental conditions, and model generalization to different regions warrant further investigation. Additionally, the model's robustness and its adaptability to changes in mosquito populations must be carefully considered. Thus, although the proposed classification system is promising, ongoing research and validation efforts are required to address uncertainties and improve the applicability of the model under various settings and conditions. Through these improvements, the mosquito classification system will hopefully continue to develop, significantly contributing to the understanding of mosquito-based diseases and opening the door for further innovative research.

Author contributions

Conceptualization, E.J., M.I.; methodology, E.J., M.I., and T.A.; software, D.B., and N.D.; validation, T.A.; formal analysis, S.T.; investigation, D.B., and N.D.; resources, I.A.; data curation, D.B., and N.D.; writing—original draft preparation, T.H.; writing—review and editing, E.J., M.I., I.A., S.T., M.E., D., and T.A.; visualization, E.J., M.I., and T.H.; supervision, I.A., S.T., M.E., and D.; project administration, E.J.; funding acquisition, E.J. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the Innovative-Productive Research Invitation-RISPRO KEP-52/LPDP/2019 and PRJ-23/LPDP/2019, the Indonesia Endowment Fund for Education, Ministry of Finance, 2019–2023, Indonesia.

CRediT authorship contribution statement

Endra Joelianto: Writing – review & editing, Visualization, Project administration, Methodology, Funding acquisition, Conceptualization. **Miranti Indar Mandasari:** Writing – review & editing, Visualization, Methodology, Conceptualization. **Daniel Beltsazar Marpaung:** Software, Investigation, Data curation. **Naufal Dzaki Hafizhan:** Software, Investigation, Data curation. **Teddy Heryono:** Writing – original draft, Visualization. **Maria Ekawati Prasetyo:** Writing – review & editing, Supervision. **Dani:** Writing – review & editing, Supervision. **Susy Tjahjani:** Writing – review & editing, Supervision. **Tjandra Anggraeni:** Writing – review & editing, Validation, Methodology. **Intan Ahmad:** Writing – review & editing, Supervision, Resources.

Declaration of competing interest

The authors declare no conflicts of interest. The funders had no role in the study design; data collection, analyses, or interpretation; manuscript writing; or decision to publish the results.

Data availability

Requests for materials should be addressed to E. Joelianto.

Acknowledgements

This study was supported by an Innovative-Productive Research Invitation-RISPRO KEP-52/LPDP/2019 and PRJ-23/LPDP/2019, the Indonesia Endowment Fund for Education, Ministry of Finance, 2019-2023, Indonesia. The authors also thank CINOVASI Ltd., Bandung, Indonesia for providing technical support for this research. The authors would like to acknowledge the Associate Editor and anonymous reviewers for their valuable comments on improving the manuscript.

References

Abdel-Hamid, O., Mohamed, A.R., Jiang, H., Deng, L., Penn, G., Yu, D., 2014. Convolutional neural networks for speech recognition. IEEE/ACM Trans. Audio Speech Lang. Process 22 (10), 1533–1545.

Agustiningtyas, I., Lusiyana, N., 2017. Ovitrap survey and serotype identification of dengue virus on *Aedes* sp. mosquito in Potorono, Banguntapan, Bantul, Indonesia. Int. J. Mosq. Res. 4 (5), 32–37.

Albawi, S., Mohammed, T.A., Al-Zawi, S., 2017. Understanding of a convolutional neural network. In: 2017 International Conference on Engineering and Technology (ICET). IEEE, pp. 1–6.

Aldersley, A., Champneys, A., Homer, M., Robert, D., 2014. Time-frequency composition of mosquito flight tones obtained using Hilbert spectral analysis. J. Acoust. Soc. Am. 136 (4), 1982–1989.

Arthur, B.J., Emr, K.S., Wyttenbach, R.A., Hoy, R.R., 2014. Mosquito (*Aedes aegypti*) flight tones: frequency, harmonicity, spherical spreading, and phase relationships. J. Acoust. Soc. Am. 135 (2), 933–941.

Aucouturier, J.J., Defreville, B., Pachet, F., 2007. The bag-of-frames approach to audio pattern recognition: a sufficient model for urban soundscapes but not for polyphonic music. J. Acoust. Soc. Am. 122 (2), 881–891.

Capinha, C., Ceia-Hasse, A., Kramer, A.M., Meijer, C., 2021. Deep learning for supervised classification of temporal data in ecology. Ecol. Inform. 61, 101252.

Cator, L.J., Arthur, B.J., Ponlawat, A., Harrington, L.C., 2011. Behavioral observations and sound recordings of free-flight mating swarms of *ae. Aegypti* (Diptera: Culicidae) in Thailand. J. Med. Entomol. 48 (4), 941–946.

Ceia-Hasse, A., Sousa, C.A., Gouveia, B.R., Capinha, C., 2023. Forecasting the abundance of disease vectors with deep learning. Ecol. Inform. 78, 102272.

da Silva, B.F., dos Santos Rodrigues, R.Z., Heiskanen, J., Abera, T.A., Gasparetto, S.C., Biase, A.G., Ballester, M.V.R., de Moura, Y.M., de Stefano Piedade, S.M., de Oliveira Silva, A.K., de Camargo, P.B., 2023. Evaluating the temporal patterns of land use and precipitation under desertification in the semi-arid region of Brazil. Ecol. Inform. 77, 102192.

Deng, L., Yu, D., 2014. Deep learning: methods and applications. FNT Sign. Proc. 7 (3–4), 197–387.

Fanioudakis, E., Geismar, M., Potamitis, I., 2018. Mosquito wingbeat analysis and classification using deep learning. In: 2018 26th European signal processing conference (EUSIPCO), 2410–2414. IEEE.

Fernandes, M.S., Cordeiro, W., Recamonde-Mendoza, M., 2021. Detecting *Aedes aegypti* mosquitoes through audio classification with convolutional neural networks. Comput. Biol. Med. 129, 104152.

Fisher, R., Wilson, S.K., Sin, T.M., Lee, A.C., Langlois, T.J., 2018. A simple function for full-subsets multiple regression in ecology with R. Ecol. Evol. 8 (12), 6104–6113.

Forman, G., Scholz, M., 2010. Apples-to-apples in cross-validation studies. SIGKDD Explor. Newsl. 12 (1), 49–57. https://doi.org/10.1145/1882471.1882479.

França, R.P., Monteiro, A.C.B., Arthur, R., Iano, Y., 2021. Chapter 3 - An overview of deep learning in big data, image, and signal processing in the modern digital age. Trends Deep Learn. Methodol. 63–87.

Ghani, B., Klinck, H., Hallerberg, S., 2023. Classification of group-specific variations in songs within house wren species using machine learning models. Ecol. Inform. 74, 101946.

Ghoraani, B., Krishnan, S., 2011. Time–frequency matrix feature extraction and classification of environmental audio signals. IEEE Trans. Audio Speech Lang. Process. 19 (7), 2197–2209.

Gonzalez, C.R., Guzman, C., Andreo, V., 2023. Using VHR satellite imagery, OBIA and landscape metrics to improve mosquito surveillance in urban areas. Ecol. Inform. 77, 10222.

Grandini, M., Bagli, E., Visani, G., 2020. Metrics for multiclass classification: an overview. arXiv, 1-17, Preprint ArXiv:2008.05756.

Gunara, N.P., Joelianto, E., Ahmad, I., 2023. Identification of *Aedes aegypti* and *Aedes albopictus* eggs based on image processing and elliptic Fourier analysis. Sci. Rep. 13 (1), 17395.

Hamesse, C., Andreo, V., Rodriguez Gonzalez, C.R., Beumier, C., Rubio, J., Porcasi, X., Lopez, L., Guzman, C., Haelterman, R., Shimoni, M., Scavuzzo, C.M., 2023. Ovitrap monitor-online application for counting mosquito eggs and visualisation toolbox in support of health services. Ecol. Inform. 75, 102105.

Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N., Kingsbury, B., 2012. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Process. Mag. 29 (6), 82–97.

Indonesia Ministry of Health, 2019. Profil Kesehatan Indonesia 2019. Available online: https://www.kemkes.go.id/id/profil-kesehatan-indonesia-2019 (Accessed 1.15.2024).

Jhaveri, A., Sangwan, K.S., Maan, V., Dhiraj, 2022. Deep learning-based mosquito species detection using wingbeat frequencies. In: Smart Innov. Syst. Technol. Springer Science+Business Media Deutschland GmbH, pp. 71–80. https://doi.org/10.1007/978-981-16-6624-7_8.

Jiang, W., Majumder, S., Kumar, S., Subramaniam, S., Li, X., Khedri, R., Mondal, T., Abolghasemian, M., Satia, I., Deen, M.J., 2022. A wearable tele-health system towards monitoring COVID-19 and chronic diseases. IEEE Rev. Biomed. Eng. 15, 61–84.

Joshi, A., Miller, C., 2021. Review of machine learning techniques for mosquito control in urban environments. Ecol. Inform. 61, 101241.

Juhdi, I., Fitri, L., Zuhriyah, L., Arasy, A., 2019. Ovitrap index and transovarial transmission rate of dengue virus of male and female *Aedes aegypti* mosquitoes in Makassar, South Sulawesi, Indonesia. J. Trop. Life Sci. 9 (1), 95–103. https://doi.org/10.11594/JTLS.09.01.13.

Kemenkes, RI, 2022. Penguatan Kapasitas Petugas Filariasis dalam Rangka Peringatan Hari NTD Sedunia 2022. Newsletter DITJEN P2P 1, 4. https://p2p.kemkes.go.id/newsletter-ditjen-p2p-edisi-i-tahun-2022/.

Kim, T., Lee, J., Nam, J., 2018. Sample-level CNN architectures for music auto-tagging using raw waveforms. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 366–370.

Kiskin, I.A., Cobb, D., Wang, L., Roberts, S., 2020. Humbug zooniverse: a crowd-sourced acoustic mosquito dataset. In: 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 916–920.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Lee, K.Y., Chung, N., Hwang, S., 2016. Application of an artificial neural network (ANN) model for predicting mosquito abundances in urban areas. Ecol. Inform. 36, 172–180.

Martin, E., Medeiros, M.C.I., Carbajal, E., Valdez, E., Juarez, J.G., Garcia-Luna, S., Salazar, A., Qualls, W.A., Hinojosa, S., Borucki, M.K., Manley, H.A., Badillo-Vargas, I.E., Frank, M., Hamer, G.L., 2019. Surveillance of *Aedes aegypti* indoors and outdoors using Autocidal gravid Ovitraps in South Texas during local transmission of Zika virus, 2016 to 2018. Acta Trop. 192, 129–137. https://doi.org/10.1016/J.ACTATROPICA.2019.02.006.

Moore, A., Miller, J.R., Tabashnik, B.E., Gage, S.H., 1986. Automated identification of flying insects by analysis of wingbeat frequencies. J. Econ. Entomol. 79 (6), 1703–1706.

Navamani, T.M., 2019. Chapter 7 - Efficient deep learning approaches for health informatics. In: Deep Learning and Parallel Computing Environment for Bioengineering Systems. Academic Press, pp. 123–137.

Ng, A., 2018. Machine Learning Yearning: Technical Strategy for AI Engineers, In the Era of Deep Learning, deeplearning. AI Project. Available online. https://info.deeplearning.ai/machine-learning-yearning-book (Accessed 1.15.2024).

Nolasco, I., Singh, S., Morfi, V., Lostanlen, V., Strandburg-Peshkin, A., Vidaña-Vila, E., Gill, L., Pamuła, H., Whitehead, H., Kiskin, I., Jensen, F.H., Morford, J., Emmerson, M.G., Versace, E., Grout, E., Liu, H., Ghani, B., Stowell, D., 2023. Learning to detect an animal sound from five examples. Ecol. Inform. 77, 102258.

Ortega-Morales, A.I., Moreno-García, M., González-Acosta, C., Correa-Morales, F., Dec. 2018. Mosquito surveillance in Mexico: the use of ovitraps for *Aedes aegypti*, *ae. Albopictus*, and non-target species. Fla. Entomol. 101 (4), 623–626. https://doi.org/10.1653/024.101.0425.

O'Shaughnessy, D., 1987. Speech Communications: Human and Machine. Addison-Wesley Publishing Company.

Ouyang, T.H., Yang, E.C., Jiang, J.A., Te Lin, T., Oct. 2015. Mosquito vector monitoring system based on optical wingbeat classification. Comput. Electron. Agric. 118, 47–55. https://doi.org/10.1016/j.compag.2015.08.021.

Peterson, A.T., Papeş, M., Soberón, J., 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecol. Model. 213 (1), 63–72.

Potamitis, I., Rigakis, I., 2016. Measuring the fundamental frequency and the harmonic properties of the wingbeat of a large number of mosquitoes in flight using 2D optoacoustic sensors. Appl. Acoust. 109, 54–60.

Rao, K.S., Manjunath, K.E., 2017. Speech Recognition Using Articulatory and Excitation Source Features. Springer. https://doi.org/10.1007/978-3-319-49220-9.

Reed, E.M.X., Byrd, B.D., Richards, S.L., Eckardt, M., Williams, C., Reiskind, M.H., Feb. 2019. A statewide survey of container *aedes* mosquitoes (Diptera: Culicidae) in North Carolina, 2016: a multiagency surveillance response to Zika using ovitraps. J. Med. Entomol. 56 (2), 483–490. https://doi.org/10.1093/JME/TJY190.

Sahidullah, M., Saha, G., 2012. Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. Speech Comm. 54 (4), 543–565.

Santos, D.A.A., Rodrigues, J.J.P.C., Furtado, V., Saleem, K., Korotaev, V., Apr. 2019. Automated electronic approaches for detecting disease vectors mosquitoes through the wing-beat frequency. J. Clean. Prod. 217, 767–775.

Sasmita, H.I., Neoh, K.B., Yusmalinar, S., Anggraeni, T., Chang, N.T., Bong, L.J., Putra, R. E., Sebayang, A., Silalahi, C.N., Ahmad, I., Tu, W.C., 2021. Ovitrap surveillance of dengue vector mosquitoes in Bandung city, West Java province, Indonesia. PLoS Negl. Trop. Dis. 15 (10), e0009896 https://doi.org/10.1371/journal.pntd.0009896.
Silva, D.F., Souza, V.M.A., Ellis, D.P.W., Keogh, E.J., Batista, G.E.A.P.A., 2015. Exploring low cost laser sensors to identify flying insect species: Evaluation of machine learning and signal processing methods. J. Intell. Robot. Syst. 80 (S1), 313–330.
Soh, S., Aik, J., 2021. The abundance of *Culex* mosquito vectors for West Nile virus and other flaviviruses: a time-series analysis of rainfall and temperature dependence in Singapore. Sci. Total Environ. 754, 142420.
Sotavalta, O., 1952. The essential factor regulating the wing-stroke frequency of insects in wing mutilation and loading experiments and in experiments at subatmospheric pressure. Annales Zoologici Societatis Zoologicae Botanicae Fennicae "Vanamo"; Tom. 15 (2).
Stevens, S.S., Volkmann, J., Newman, E.B., 1937. A scale for the measurement of the psychological magnitude pitch. J. Acoust. Soc. Am. 8 (3), 185–190.
Tang, Q., Xu, L., Zheng, B., He, C., 2023. Transound: hyper-head attention transformer for birds sound recognition. Ecol. Inform. 75, 102001.
Thakur, S., Kumar, A., Aug. 2021. X-ray and CT-scan-based automated detection and classification of Covid-19 using convolutional neural networks (CNN). Biomed. Signal Process. Control 69, 102920. https://doi.org/10.1016/J.BSPC.2021.102920.
Truong, T.H., Du Nguyen, H.D., Mai, T.Q.A., Nguyen, H.L., Dang, T.N.M., Phan, T., 2023. A deep learning-based approach for bee sound identification. Ecol. Inform. 78, 102274.
Villarreal, S.M., Winokur, O., Harrington, L., 2017. The impact of temperature and body size on fundamental flight tone variation in the mosquito vector *Aedes aegypti* (Diptera: Culicidae): implications for acoustic lures. J. Med. Entomol. 54 (5), 1116–1121.
World Health Organization, 2014. Dengue and severe dengue. WHO Fact Sheet.
World Health Organization, 2017. Guideline: Alternative Mass Drug Administration Regimens to Eliminate Lymphatic Filariasis (No. WHO/HTM/NTD/PCT/2017.07). World Health Organization.
Yin, M.S., Haddawy, P., Nirandmongkol, B., Kongthaworn, T., Chaisumritchoke, C., Supratak, A., Sa-Ngamuang, C., Sriwichai, P., 2021. A lightweight deep learning approach to mosquito classification from wingbeat sounds. In: Proceedings of the Conference on Information Technology for Social Good. ACM, pp. 37–42.
Yin, M.S., Gunatilaka, D., Assawavinijkulchai, K., Thampakorn, T., Heesawat, W., Haddawy, P., Eiamsamang, S., Sriwichai, P., Weber, M., 2023. MosquitoSongSense: IoT-based mosquito wingbeat data collection system. In: Proceedings of the 2023 ACM Conference on Information Technology for Social Good. ACM, pp. 285–290.
Zoladek, J., Nisole, S., 2023. Mosquito-borne flaviviruses and type I interferon: catch me if you can! Front. Microbiol. 14, 1257024.
Zuhriyah, L., Satoto, T.B.T., Kusnanto, H., 2016. Efektifitas Modifikasi Ovitrap Model Kepanjen Untuk Menurunkan Angka Kepadatan Larva *Aedes aegypti* di Malang. Jurnal Kedokteran Brawijaya 29 (2), 157–164. https://doi.org/10.21776/UB.JKB.2016.029.02.10.