African J. Trop. Entomol. Res. 4(2):49-61
https://doi.org/10.58697/AJTER040201 Yamoa et al., 2025

Research article

African Journal of Tropical Entomology Research
ISSN: 2791-0113 (Online)
Journal homepage: www.ajter.com, Lenaf homepage: www.lenaf-ifa.org

African Journal Of Tropical Entomology Research

Open Access

Classification of Mosquito Species (Culicidae): Simplified Approach with Scheimpflug Lidar for Optimized Monitoring

Assoumou Saint-Doria YAMOA[1*], Yao Taky Alvarez KOSSONOU[1], Marie-Florence YEBOUET[1], Benoit KOUASSI KOUAKOU[1,2], Adolphe YATANA GBOGBO[1] & Jeremie T. ZOUEU[1,2]

[1]Instrumentation, Imaging and Spectroscopy Laboratory, Félix Houphouët-Boigny Polytechnic Institute, Yamoussoukro, Ivory Coast.
[2]Dept. Physics, University of San Pedro, San Pedro, Ivory Coast

**Corresponding author*, E-mail: assoumou.yamoa21@inphb.ci

Received: 12 December 2024 | **Accepted: 28 May 2025** | **Published: 30 June 2025**

ABSTRACT

Vector monitoring remains a challenging task for large-scale continuous surveillance. Light Detection and Ranging (Lidar) systems, which use laser pulses to detect and characterize objects at a distance (Lidar systems), offer a solution by capturing the wingbeat modulation frequencies of flying insects, especially mosquitoes. However, the wingbeat frequencies of different mosquito species are often very similar. The use of additional parameters is necessary to improve the classification. In this study, we used a Scheimpflug lidar system to record the kHz-modulated backscattered light of mosquitoes and applied the Random Forest method to select the most discriminating parameters. The simplified model, based only on two parameters, the fundamental frequency (Freq) and the first harmonic (Harm1), achieved an accuracy of 44.14%, as shown in the confusion matrix. Although this accuracy may seem modest, it is remarkable considering the inherent challenges of the problem, such as the reduced parameter set and the complexity of the data. These parameters enabled better separation of mosquito specimens collected by the lidar, demonstrating that reducing the number of parameters can not only maintain but also improve the model's classification accuracy while reducing its complexity.

Keywords: Scheimpflug Lidar, Wingbeat Frequency, Harmonic, Mosquito Classification

RÉSUMÉ

Classification des espèces de moustiques (Culicidae) : approche simplifiée avec Scheimpflug Lidar pour une surveillance optimisée

La surveillance des vecteurs demeure une tâche complexe dans le cadre d'une surveillance continue à grande échelle. Les systèmes de télédétection par laser (Lidar), qui utilisent des impulsions laser pour détecter et caractériser des objets à distance, offrent une solution en capturant les fréquences de modulation des battements d'ailes des insectes volants, en particulier des moustiques. Cependant, les fréquences des battements d'ailes des différentes espèces de moustiques sont souvent très similaires. L'utilisation de paramètres supplémentaires est donc nécessaire pour améliorer la classification. Dans cette étude, nous avons utilisé un système Lidar de type Scheimpflug pour enregistrer la lumière rétrodiffusée modulée en kHz par les moustiques et avons appliqué la méthode Random Forest afin de sélectionner les paramètres les plus discriminants. Le modèle simplifié, basé uniquement sur deux paramètres (la fréquence fondamentale «Freq» et la première harmonique «Harm1») a atteint une précision de 44,14 %, comme le montre la matrice de confusion. Bien que cette précision puisse paraître modeste, elle est remarquable compte tenu des défis inhérents au problème, tels que le nombre réduit de paramètres et la complexité des données. Ces paramètres ont permis une meilleure distinction des spécimens de moustiques détectés par le lidar, démontrant que la réduction du nombre de paramètres peut non seulement préserver mais aussi améliorer la précision de classification du modèle tout en réduisant sa complexité.

Mots-clés : Lidar Scheimpflug, Fréquence de battement d'ailes, Harmonique, Classification des moustiques

INTRODUCTION

Vector-borne diseases continue to pose a serious threat to global public health, particularly in tropical regions (Gubler, 1998; Organization, 2017). Caused by the *Plasmodium falciparum* parasite and spread by *Anopheles* mosquitoes, mosquitoes are the deadliest animals in the world, killing around one million people each year (Murray et al., 2012). According to the report (OMS, 2023), the African Region bears a disproportionate and significant share of the global malaria burden. In 2022, it

recorded 94% of malaria cases, or 233 million cases, as well as 95% of malaria-related deaths, totaling 580,000 deaths. Among these deaths, 80% were children under five, highlighting the devastating impact of malaria on this age group in the region. To overcome this challenge, medical treatment with affordable drugs, rapid and accurate diagnosis, and prevention are all necessary. The primary method employed to eradicate the disease is to implement preventive measures, such as landscape drainage, spraying, urban planning, and marsh management according to (Organization, 2015). However, the rapid adaptation of the parasite and vector remains a major problem. Recently, drug-resistant parasites have been discovered throughout sub-Saharan Africa (Corbel et al., 2012), and national mosquito distribution campaigns have shown a shift in mosquito temporal niche to earlier hours in the evening (Gatton et al., 2013; Yohannes & Boelee, 2012). Other vector-borne infectious diseases on the African continent are considered curable with precise knowledge of population dynamics and landscape flows (Benelli, 2016; Dao et al., 2014; Ferguson et al., 2010). This would allow for improved existing intervention methods. However, quantifying the sustainable ecosystem and life stages of mosquitoes is not an easy task. Within a few minutes, the activity and airborne proliferation of a mosquito specimen can vary by orders of magnitude. These changes are due to meteorological circumstances (Kirkeby et al., 2016; Malmqvist et al., 2018). Moreover, preferred breeding environments and reproductive experiences can be limited to plants and small landmarks within a few meters (Chen et al., 2024; Liu et al., 2017; Silver, 2007).

Researchers have used insect samples collected from moving vehicles (Silver, 2007), light traps (Liu et al., 2017), electronic traps (Chen et al., 2024; Potamitis et al., 2015), kairomone and CO_2 traps (Costantini et al., 1996; Mboera et al., 2000), and human appendages (Govella et al., 2009; Kenea et al., 2017; Maliti et al., 2015) to assess mosquito population dynamics. It is well known that each method may be biased towards specific species, sexes, or life stages, and that extensive temporal and spatial coverage can make surveys very time-consuming. There are few methods available to track mosquito flows and estimate their spread, often unfeasible for ethical reasons, even in cases where powder marking (Brydegaard et al., 2016; Hagler & Jackson, 2001) is possible. The best scenario for sampling is several weeks when the sky is clear, despite the possibility of directly correlating mosquitoes' preferred habitats with geographic data, such as satellite images or aerial topographic lidars (Hartfield et al., 2011; Millette et al., 2010). Consequently, such methods are unable to provide a complete description of daily activities and flows.

In recent years, entomological lidar —a specialized form of Light Detection and Ranging (Lidar) adapted to detect and analyze flying insects using their optical signatures and wingbeat frequencies— and remote modulation spectroscopy have been developed by (Brydegaard et al., 2017; Repasky et al., 2006) to ensure direct insect monitoring (Brydegaard et al., 2014, 2017; Malmqvist et al., 2016). Lidar devices used to distinguish captured insects rely on wingbeat modulation. However, it can be difficult to differentiate mosquito species based on wingbeat frequency, as these frequencies frequently overlap and are very different between and within species (Jansson et al., 2019). For this reason, (Gebru et al., 2018) used additional factors to distinguish flying mosquito species, including wingbeat frequency, the optical cross-section of the body and wings, and the harmonic content of signals recorded in various spectral and the degree of polarization. In this report, we present the parameters for disease vector species discrimination obtained through classification models (SVM, KNN, Neural Network, and Deep Learning) based on three distinct approaches (use of original parameters, use of parameters after selection, and use of parameters after applying Principal Component Analysis (PCA)). The primary goal was to identify differentiation parameters for specimens collected by the Scheimpflug lidar using classification models. The approaches were ranked based on the classification models' accuracy, determining their ability to distinguish insect specimens with similar frequencies.

Data were collected at the Félix Houphouët-Boigny Polytechnic Institute (INP-HB South) site in Yamoussoukro, Côte d'Ivoire, and cover 54,133 records of four different mosquito species: *Anopheles coluzzii* (*A. coluzzii*), *Anopheles arabiensis* (*A. arabiensis*), *Aedes aegypti* (*Ae. aegypti*), and *Culex quinquefasciatus* (*Cu. quinquefasciatus*) of both sexes (male and female). The wingbeat frequencies used for data extraction are from the study (Jansson et al., 2019), where these frequencies are well-documented.

MATERIALS AND METHODS

Experimental Setup

The instrument used in this study was installed near the geology lab buildings at INP-HB South in Yamoussoukro, at coordinates (Lat: 6.872231°, Lon: -5.237526°), see Fig. 1. A 3.2 W multimode laser diode with a central wavelength of 808 nm was used to expand the light through a refractive telescope (f = 500 mm, ø102 mm). The beam was projected across the landscape and ended on a board covered by black neoprene foam with a diffuse reflectance of 1.8% at 808 nm. The laser spot measured 20 cm wide and 1 cm high at the end of the beam. The board was fixed at a height of 9 m on a tree located 970 m from the lidar system (Lat: 6.877774°, Lon: -5.231588°). A Newtonian telescope (f = 800 mm, ø 200 mm) was used to capture the backscattered light, which was then focused onto a 2048-pixel CMOS linear sensor with a 16-bit dynamic range, measuring 14 × 200 µm². To exclude ambient light, an absorption filter (RG780) and a bandpass filter (3 nm full width at half-maximum, centered at 808 nm) were used. The receiver and expander

telescopes were mounted on a base fixed to a tripod at 814 mm from each other. A 45° angle was created between the CMOS sensor and the optical axis of the receiver telescope. A combined laser driver and time-multiplexed switch were connected to the CMOS sensor and laser diode. With this module, the laser can operate in continuous or modulated mode. In modulated mode, a signal from the detector to a multiplexer that activated and deactivated the laser was used to expose the sensor's odd and even lines. As the line sensors had a frequency of 3.5 kHz, the background signal and backscattered signal could be subtracted in real-time since each was sampled at 1.75 kHz. The laser beam and linear camera were aligned on the black neoprene foam. Using this black terminating surface reduced the amount of light interference in the monitored aerial transect and prevented detector saturation. The collected data were stored in 10-second files, each containing 35,000-line exposures.

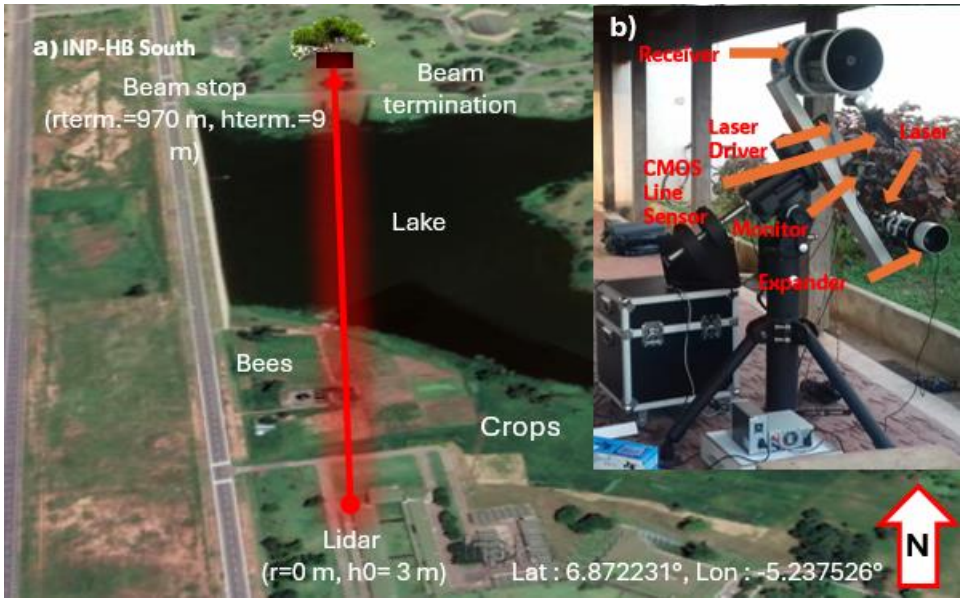

Fig. 1. (a) Aerial photograph of the measurement location. The light is transmitted from the lidar near the geology lab buildings, 3 m above the ground, and ends on a neoprene-covered board mounted on a tree, 970 m from the lidar system. (b) Photograph of the Scheimpflug lidar and its components. The linear sensor in the Newtonian receiver is tilted at 45° relative to the optical axis to satisfy the Scheimpflug criterion and achieve infinite depth of field.

Analysis of wing beat frequency of flying insects

For entomological lidar data, the fast Fourier transform is used to calculate the wing beat frequency of flying insects. As a reminder, this fundamental wing beat frequency is a discriminating parameter of flying insects.

$$\mathbf{S(f)} = \int_{-\infty}^{+\infty} \mathbf{s(t)}.\mathbf{e^{-2\pi ft}}\,\mathbf{dt} \qquad \text{Eq.1}$$

Fig.2 shows the optical cross section (OCS and in French SEO) of a flying insect detected by the entomological Scheimpflug lidar after calibration of the intensity backscattered by the latter.At the end of this transform, an initial estimate of the fundamental frequency $f_{0init}$is obtained and it will be used to find the combination of harmonics that best reconstructs the original OCS. This original cross section contains the contribution of the body and that of the wings of the insect Eq.2.

$$\mathrm{OCS} = \mathrm{OCS}_{body} + \mathrm{OCS}_{wing} \qquad \text{Eq.2}$$

The OCS for wing beats $\mathrm{OCS}_{wing}$is considered as a linear combination of harmonics with:

$\mathrm{OCS}_{wing} = \sum_n a_n \varphi_n$, with $\varphi_n = sin(2\pi n f_{0init} t) + cos(2\pi n f_{0init} t)$ Eq.3

The coefficients, an, are obtained by the least squares method (Eq.4)

$$\mathrm{a_n} = (\varphi_\mathrm{n}^\mathrm{T}\varphi_\mathrm{n})^{-1}\varphi_\mathrm{n}^\mathrm{T}\mathrm{OCS_{wing}} \qquad \text{Eq.4}$$

The OCS of the wings $\mathrm{OCS}_{wing}$is reconstructed for frequencies lower and higher than 30% of the initial estimate f0init. Finally, each frequency is estimated using least squares fit (Eq. 5):

$$\left|\frac{1}{N}\sum_0^N\left(\mathrm{OCS} - \mathrm{OCS}_{body} - \mathrm{OCS}_{wing}\right)^2\right|_{min} \qquad \text{Eq.5}$$

The frequency that best reconstructs the OCS is defined as the fundamental wingbeat frequency of the insect. Fig. 3 shows us an OCS and its fundamental wingbeat frequency.

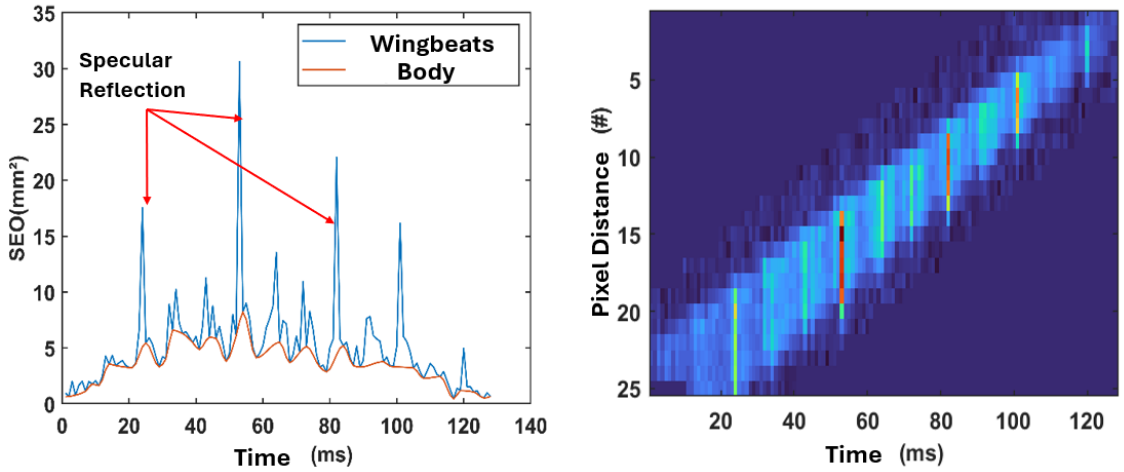

Fig. 2: The temporal OCS of an insect. The red curve represents the contribution of the insect body. The large peaks represent a spectacular reflection. Time-distance image of the passage of a flying insect is shown on the right.

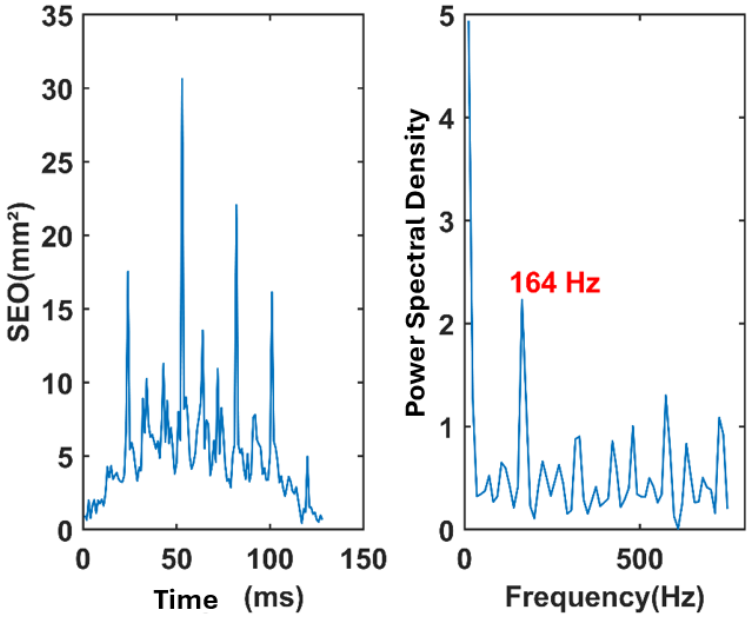

Fig. 3: Fig. 3: a) Optical cross section of an insect as a function of time. b) Power spectral density of the insect wing beats after Fourier transformation. This signal corresponds to a random insect used as an example to illustrate the method. The measured fundamental frequency is 164 Hz, just for the demonstration of the spectral analysis process applied in this study.

Mosquito Specimens

In this article, the wingbeat frequencies of mosquitoes, used to retrieve various events in the lidar data collected at the INP-HB site, were proposed in a previous study by (Gebru et al., 2018), including both sexes (male and female) of 4 different species, namely: A. coluzzii, A. arabiensis, Ae. aegypti, and Cu. quinquefasciatus. For more details on obtaining these frequencies, see (Gebru et al., 2018; Jansson et al., 2019).

METHODOLOGY

Calculation of lidar parameters

To improve the accuracy of flying insect identification, parameters were selected for this study, namely: the fundamental frequency (f0), first and second-order harmonics (Harm1 and Harm2), apparent size (appSize), optical cross-section (OCS), Mel-Frequency Cepstral Coefficients (MFCC) (MFCC1 to MFCC12), flight speed (Speed), the ratio of optical cross-sections of the wings and body (ratioOCS), the insect's actual size (reelSize), and signal length (Length).

To identify the wingbeat frequency (WBF), f0, a combination of several frequency analysis methods was implemented to improve WBF estimation accuracy (Gebru et al., 2018). Harmonics were obtained through autocorrelation, cepstrum, and spectrum as previously described (Brydegaard et al., 2021; de Cheveigné & Kawahara, 2002; Oppenheim & Schafer, 2004) describes the apparent size (appSize) of insects as an additional measure of size derived from the opening angle within the beam and the distribution of echo pixels. "Apparent size" is calculated from the distribution of echo pixels

according to equation 6.

$$appSize = \frac{r * \cos(\theta_{sens}) * l_{pix}}{F_{rec}} \sqrt{\frac{\sum_{mask} I_{insect(p,t)}(p - p_{cent.})^2}{\sum_{mask} I_{insect(p,t)}}} \quad \text{Eq. 6}$$

$$\text{where } p_{cent} = \frac{\sum_{mask} I_{insect(p,t)} p}{\sum_{mask} I_{insect(p,t)}}$$

Where $p_{cent}$is the central pixel calculated by the first statistical moment, mask refers to pixels and exposures exceeding the noise level, and appSize is then calculated as the square root of the second statistical moment of the echo (pixel distribution) and recalibrated by r and system parameters described in Table I (Brydegaard et al., 2021).

The apparent size, appSize, thus represents the spatial distribution of light echoes captured by the lidar system and is affected by distance r, the inclination angle of the linear camera θ_{sens}, and the point spread function (PSF) of the receiving instrument. This size is measured by considering the distribution of light echoes across the pixels of the detector, enabling an estimation of insect size.

The actual size of the insect is determined from the apparent size according to equation 7.

$$realSize = \frac{appSize * D}{\sqrt{(p_y)^2 + (p_z)^2}} \quad \text{Eq. 7}$$

Where D is the detector-insect distance, and p_y, p_z are the pixel coordinates. The authors (Brydegaard et al., 2021; Kouakou et al., 2020) estimate that the OCS of flying insects can be expressed as a function of insect wing modulation through equation 8.

$$\text{OCS} = \sigma_{term} \frac{r_{insect}^2 (I_{insect} - I_{static})}{r_{term}^2 I_{static}} \quad \text{Eq. 8}$$

Where σ_{term} is the optical cross-section of the black neoprene termination (mm²), I_{insect} is the insect signal intensity (16 bits), I_{static} is the atmospheric signal (16 bits), r_{insect} is the detection distance of the observed insect (m), and r_{term} is the distance to the beam termination (m). For more information, see (Kouakou et al., 2020). The calculation of Speed was done according to the approximations of (Li et al., 2020), and the formula is given by equation 9.

$$Speed = \frac{u_x}{u'_x} * W * cos(\theta) * k \quad \text{Eq. 9}$$

Where u_x is the object-to-lens distance (image focal length), u'_x is the distance from pixel p_x to the lens plane (object focal length), W is the pixel pitch (14 µm), θ is the inclination angle of the linear camera (45° in this case), and k is the number of fields of view covered by the insect per unit of time. The ratioOCS is defined as the ratio between the median optical cross-section of the insect's body and the median optical cross-section of the insect's wings. To calculate it, we first define the object interval. Then, we calculate the median optical cross-section of the body and the wings. We also define the maximum body cross-section of the object to check for saturation, aiming to obtain the optical cross-section that best matches the captured object. Its formula is given in equation 10.

$$\text{ratioOCS} = \frac{\text{OCS}_b}{\text{OCS}_w} \quad \text{Eq. 10}$$

Where ratioOCS is the ratio of the optical cross-section of the body to the optical cross-section of the wings, OCS_b is the median optical cross-section of the insect's body, and OCS_w is the median optical cross-section of the insect's wings. The signal length (Length) in the context of our study refers to the number of pixels that detected the insect within the detection field of the optical sensor. The actual length is obtained by multiplying the pixel pitch by the total number of pixels that detected the insect. Its formula is given in equation 11.

$$\text{Length} = \prod_{n=1}^{N} \omega_n * p_{pix} \quad \text{Eq. 11}$$

Where ω_n is the total number of pixels that detected the insect, and p_{pix} is the pixel pitch (14 µm). The Mel Frequency Cepstral Coefficients (MFCC) were calculated based on the work of (Zheng et al., 2001) using equation 12.

$$MFCC_n = \sum_{m=1}^{M} E(m) * \cos\left(\frac{\pi n(2m+1)}{2M}\right) \quad \text{Eq. 12}$$

Where $E(m)$ is the logarithm of the energy filtered by the Mel filter, n is the index of the desired cepstral coefficient, and M is the total number of Mel filters.

Evaluation of mosquito species identification parameters by classification models.

Four classification models were used in this article, namely: Support Vector Machine (SVM), K-Nearest Neighbors (k-NN), Multi-Layer Perceptron (MLP) - Neural Network, and Deep Learning - Deep Neural Network. The idea was to see which of these four models would give the best result in terms of classification. To carry out this study, three different objectives were adopted.

The first objective involved using all 21 original parameters that we were able to extract from our mosquito specimens, namely f0, Harm1, Harm2, appSize, OCS, MFCC1 to MFCC12, Speed, ratioOCS, realSize, and Length, which were then used as classifier inputs.

For the second objective, parameters were pre-selected using importance-based selection methods, namely Random Forest, which ranks parameters based on their ability to improve model accuracy when used in decision trees; Lasso, which applies L1 regularization to set the coefficients of less important parameters to zero; and Recursive Feature Elimination (RFE), which recursively eliminates the least important features to improve model accuracy. The parameters with high weights in each

method were retained and used as classifier inputs.

For the third objective, a PCA technique was used to reduce the dimensionality of the parameters while retaining at least 81% cumulative variance. The parameters whose linear combination yielded these principal components were retained and used as classifier inputs.

Of these three objectives, objective two provided the best accuracy, with the Random Forest selection method identifying the following 10 parameters as the most relevant with their associated weights: Freq (54.25%), Harm1 (14.56%), Harm2 (11.60%), MFCC5 (2.07%), MFCC6 (1.51%), MFCC4 (1.46%), MFCC7 (1.45%), MFCC8 (1.06%), MFCC9 (1.04%), MFCC2 (1.01%). These parameters allowed the classifiers to achieve the following accuracies: 88.62% for SVM, 85.53% for K-NN, 89.24% for MLP, and Deep Learning, which were significantly higher than the classifier results from objectives 1 and 3. To assess the accuracy of these selected parameters in discriminating our mosquito specimens, we visualized them through 4 different combinations (combination 1: Freq, Harm1, Harm2, MFCC5, MFCC6, MFCC4, MFCC7, MFCC8, MFCC9, MFCC2; combination 2: Freq, Harm1, Harm2, MFCC5; combination 3: f0, Harm1, Harm2; combination 4: f0, Harm1) using the confusion matrix to identify which combination had the highest accuracy. Finally, a 2D projection in the PCA space was performed to confirm the selected parameter combination for better differentiation of each mosquito specimen. We have outlined the methodology in the diagram in Fig. 4.

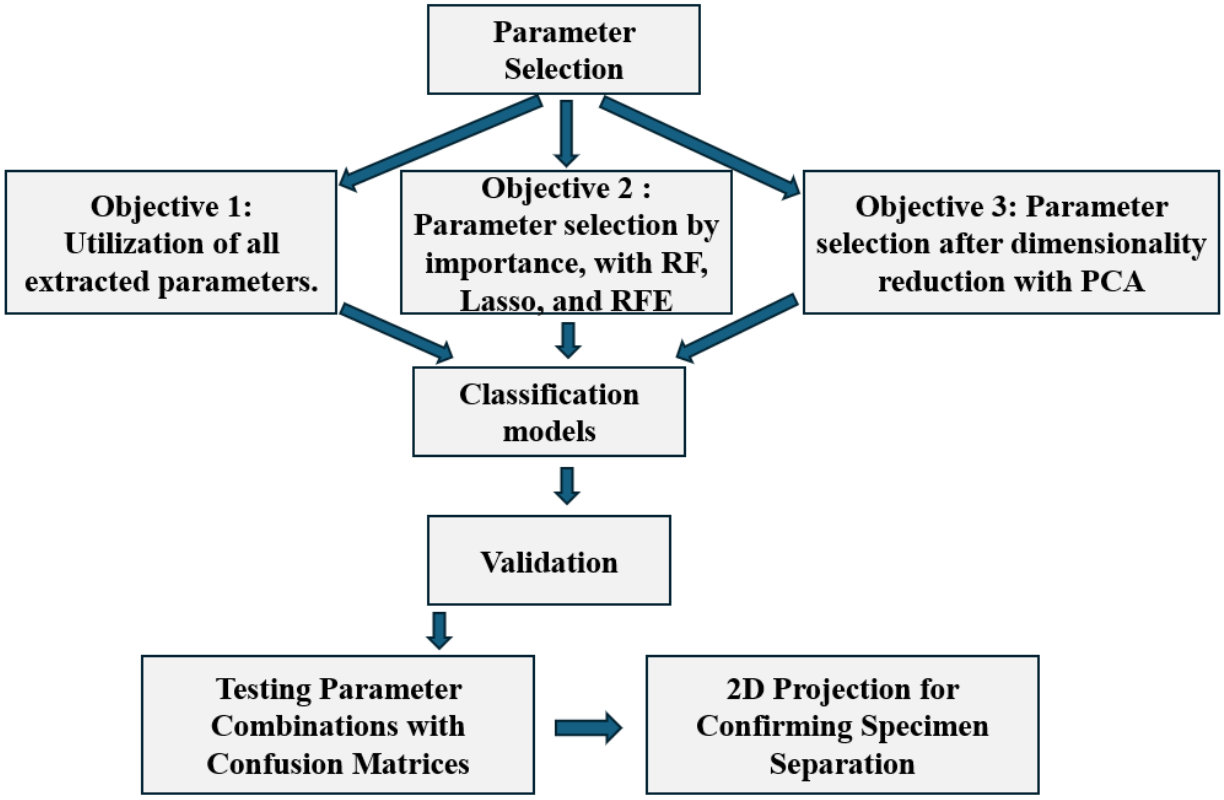

Fig.4: Our flowchart for best parameters selection for mosquitoes' specimen classification. To enhance mosquito identification, key characteristic parameters were meticulously selected. Subsequently, three objectives were outlined to deepen the study, each designed to evaluate the effectiveness of these parameters in different contexts. To assess performance, four classification models were employed, providing a robust framework for comparing results. The effectiveness of the selected parameters was validated through testing various combinations using confusion matrices. Additionally, a 2D projection was conducted to visually confirm the clear separation of specimens, as presented in the flowchart below.

RESULTS

The probability density for male and female specimens detected by the Scheimpflug lidar is presented in Fig. 5 as a function of mosquito wingbeat frequencies. These data were obtained by measuring the optical signals generated by the lidar, which capture periodic variations caused by wingbeats. Although we do not have direct ground truth for everyone, we used published data on wingbeat frequencies specific to some species, such as *Ae. aegypti* and *Cu. quinquefasciatus*, to calibrate our models.

However, these species have similar frequency ranges, leading to overlaps in the frequency distributions. This overlap can lead to mis classifications, as mentioned. To minimize this effect, we calculated probability densities to estimate the most likely ranges for each category (males/females and species).

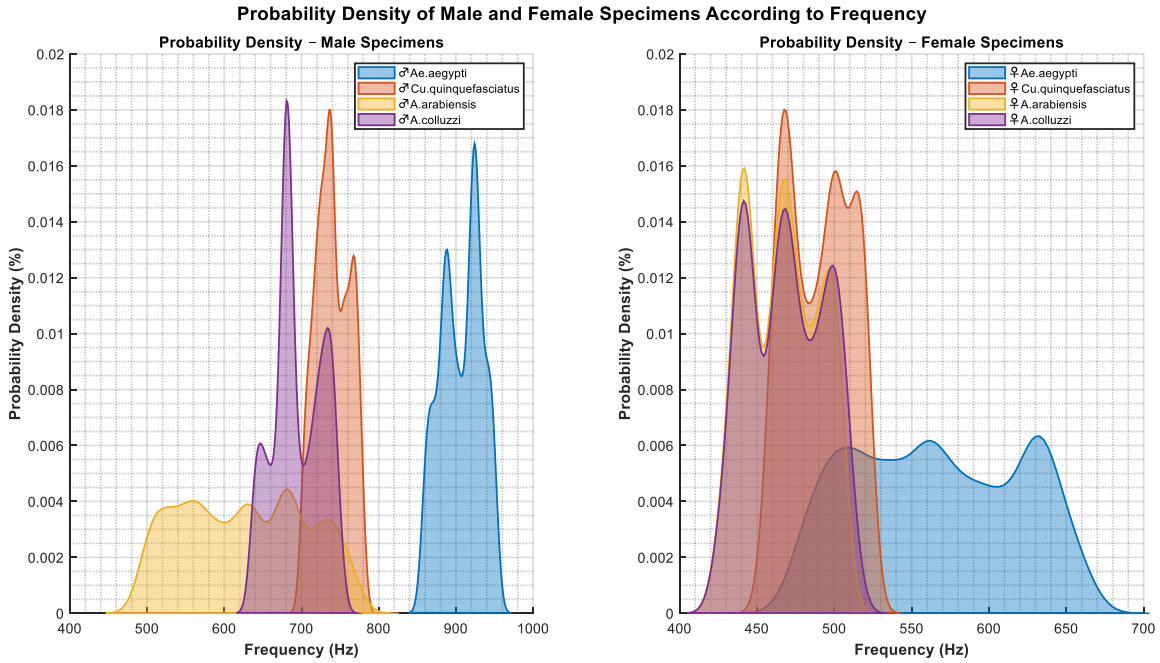

Fig.5. Probability density of selected specimens. a) male specimens. b) female specimens.

Methods such as Random Forest, Lasso, and Recursive Feature Elimination (RFE) were tested to select the most discriminative available parameters to resolve this ambiguity. Frequency (Freq: 54.25%), the first harmonic (Harm1: 14.56%), and the second harmonic (Harm2: 11.60%) were identified as the most important parameters by the Random Forest method (Fig. 6). The cepstral coefficients (MFCC) had significantly lower weights (between 1 and 2%). However, the selection of these parameters enabled the Deep Learning model to achieve an impressive accuracy of 89.24%, demonstrating the relevance of the chosen parameters.

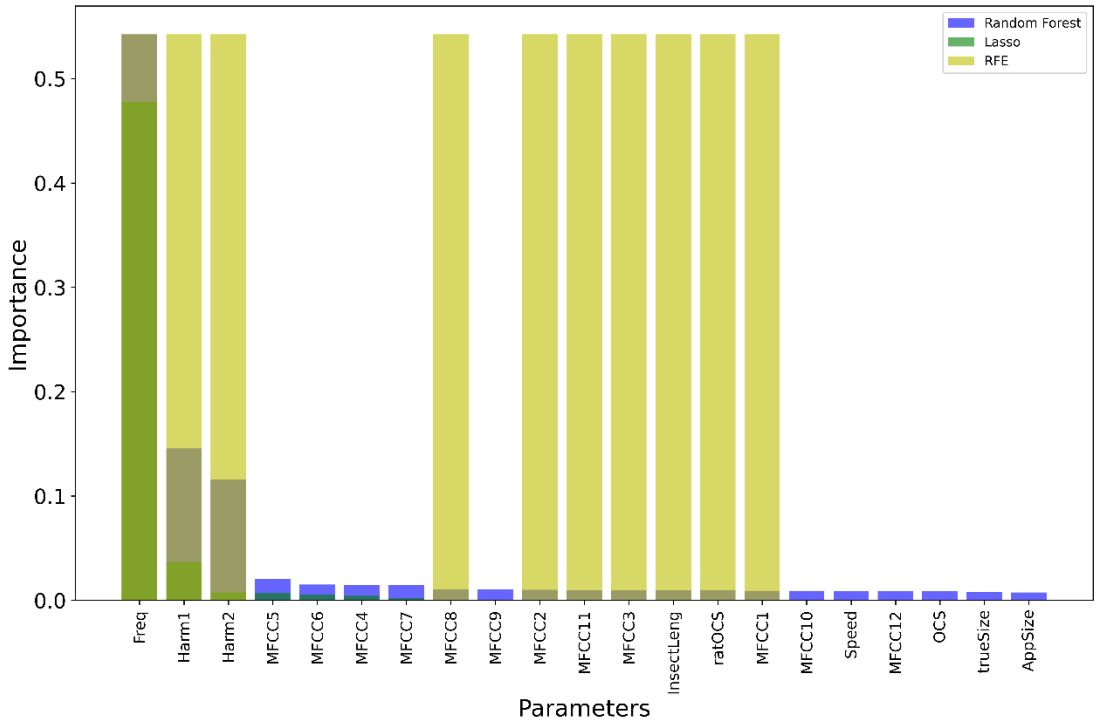

Fig.6. Comparison of parameter importances (Random Forest, Lasso, RFE)

Table 1 compares the performance of classification models based on the three main objectives. Objective 2, which uses parameter selection with Random Forest, Lasso, and RFE, achieved a better balance between performance (up to 89.24% with the MLP and Deep Learning models) and simplicity, outperforming the other methods tested.

Table 1: Comparison of objectives with classification models

Objective	Description	Classification Models (%)	Performance	Complexity	Interpretability
Objective 1 (Using all parameters)	Good performance but increased complexity and reduced interpretability.	SVM: Accuracy = 88.62%, K-NN: Accuracy = 85.53%, MLP: Accuracy = 89.24%, Deep Learning: Accuracy = 89.09%	High	High	Lower
Objective 2 (Selection with Random Forest, Lasso, and RFE)	Most suitable method. Maintains high performance while simplifying the model. Random Forest is effective for selecting discriminative parameters.	SVM: Accuracy = 88.62%, K-NN: Accuracy = 85.53%, MLP: Accuracy = 89.24%, Deep Learning: Accuracy = 89.24%	High	Reduced	Good
Objective 3 (Dimensionality reduction with PCA)	Useful for dimensionality reduction, but model performance decreases and is less suitable for biological interpretation.	SVM: Accuracy = 78.51%, K-NN: Accuracy = 76.34%, MLP: Accuracy = 86.25%, Deep Learning: Accuracy = 83.15%	Reduced	Reduced	Less suitable

The effectiveness of parameter combinations in distinguishing between two insect specimens—specifically mosquitoes—is evaluated using a confusion matrix, as shown in Fig. 7; here, they are mosquitoes, through the confusion matrix. Fig. 7-a, uses the combination of 10 parameters, thus reaching an accuracy of 43.33%. However, this accuracy is not likely, because the MFCCs have a lower weight, therefore do not bring significant information.

By limiting ourselves to the significant parameters, that is, those with relatively large weights [Freq, Harm1, Harm2, MFCC5] (Fig. 7-b), the accuracy increases slightly to 43.74%. On the other hand, using only frequency and harmonics (Fig. 7-c), we obtain an accuracy of about 43.97%, indicating that these parameters are the most relevant than the set of 10. The best accuracy is obtained with a very simple set of two first parameters [Freq, Harm1] because they have the highest weights (Fig. 7-d), reaching an accuracy of 44.14%. This accuracy with two parameters is indeed the highest in this case. This demonstrates that the fundamental frequency and the first harmonic are sufficient in this study to discriminate two overlapping species. Figs 8 and 9 present the distribution of male and female specimens projected onto the first two principal components (PCA) based on the selected parameters, with 10 parameters in Fig. 4 and 2 parameters in Fig. 5. These Fig. s allow visualization in 2D space, highlighting the value of using 2 parameters instead of the selected 10. In Fig. 4, mosquito specimens are projected using the full set of 10 parameters ([Freq, Harm1, Harm2, MFCC5, MFCC6, MFCC4, MFCC7, MFCC8, MFCC9, MFCC2]), with males on the left and females on the right. Male and female specimens are relatively dispersed along the principal axes (PC1 and PC2), with significant overlap between certain species. For example, *A. arabiensis* and *A. coluzzii* overlap substantially among males, as do *A. arabiensis* and *Cu. quinquefasciatus* among females, indicating difficulty in distinguishing these species using only the 10 selected parameters. The overall accuracy associated with this parameter set (43.33%) reflects this difficulty in accurately differentiating certain species.

Harm1, Harm2, and MFCC combinations on model accuracy. a) [Freq, Harm1, Harm2, MFCC5, MFCC6, MFCC4, MFCC7, MFCC8, MFCC9, MFCC2]. b) [Freq, Harm1, Harm2, MFCC5]. c) [Freq, Harm1, Harm2]. d) [Freq, Harm1].

In contrast, the specimens are projected using only two parameters, Freq and Harm1, as in Fig. 3-d.

Males (left): A better separation between the different species is observed. Notably, *Ae. aegypti* and *Cu. quinquefasciatus* are well separated from the other species along the PC1 axis. However, slight overlap remains between *A. arabiensis* and *A. coluzzii*.

Females (right): Their projections also show better species distinction compared to Fig. 4. Here, *Ae. aegypti* and *Cu. quinquefasciatus* are clearly separated from other species, with more compact groupings of each specimen.

With only two parameters, the overall accuracy slightly improves (44.14%), confirming that Freq and Harm1 are sufficient to better separate certain mosquito species, simplifying the model while achieving a performance slightly above the previous set.

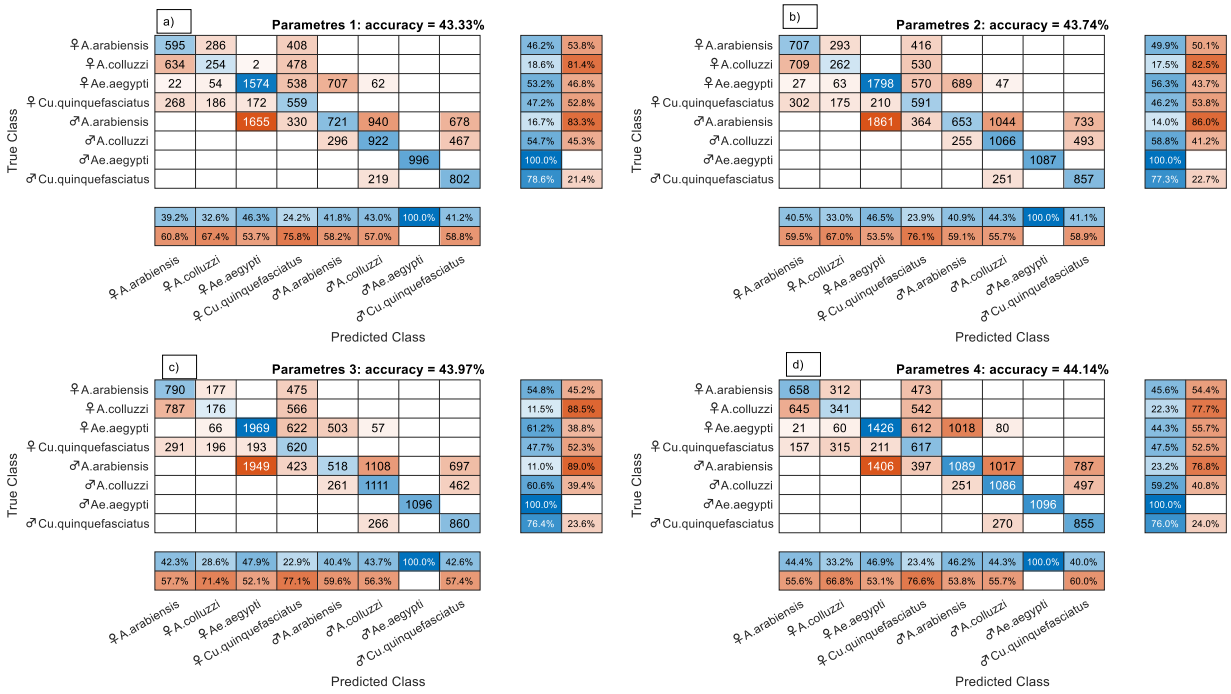

Fig. 7. Comparison of confusion matrix performances with different parameter combinations (a-d): Influence of Freq,

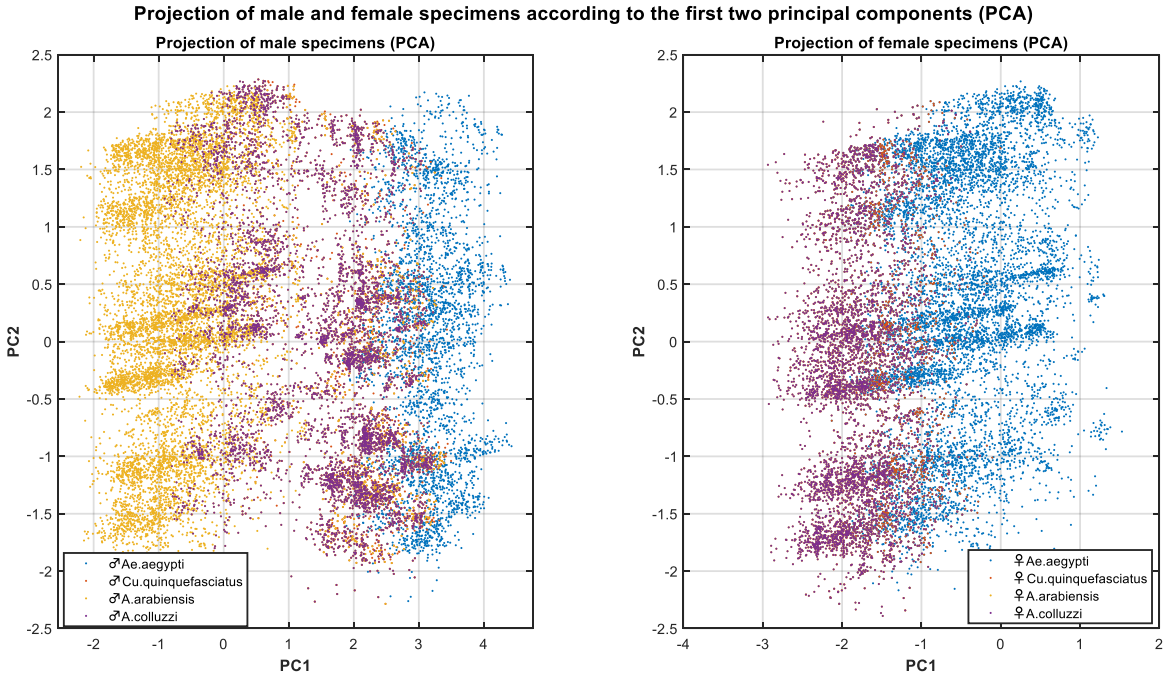

Fig. 8: Projection of male and female specimens based on the 10 main selected parameters (Freq, Harm1, Harm2, and MFCCs) in PCA space, showing high variability and significant overlaps between certain species**.**

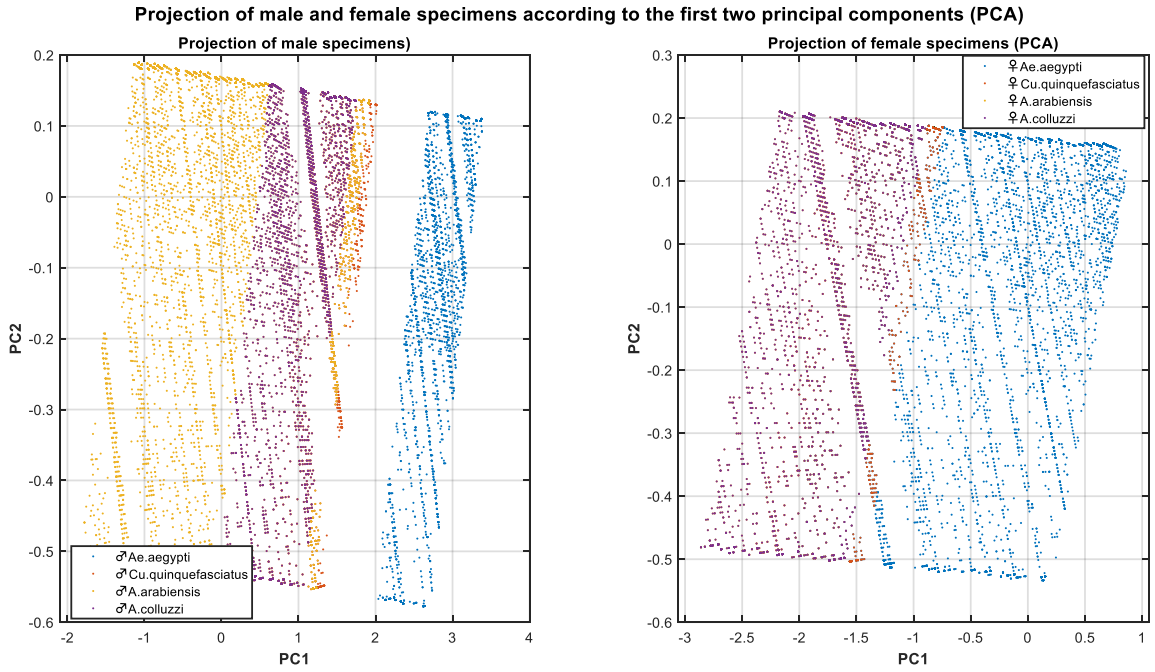

Fig. 9: Projection of male and female specimens based on the 2 main parameters (Freq and Harm1) in PCA space, providing better separation and reduced overlaps between species.

DISCUSSION AND CONCLUSION

Fig. 5 shows the probability density of wingbeat frequencies for male (a) and female (b) mosquitoes. This representation reveals overlapping frequency ranges for certain species, notably *Ae. aegypti* and *Cu. quinquefasciatus*. This overlap indicates that frequencies alone may be insufficient to differentiate mosquito specimens, especially when frequency ranges are similar or close. This limitation is more pronounced in the distribution of female specimens, where the overlap is more significant. These observations highlight the importance of supplementing the analysis with other discriminative parameters to overcome the ambiguity related to frequencies alone. Fig. 6 shows the relative importance of parameters using the Random Forest, Lasso, and Recursive Feature Elimination methods. It appears that wingbeat frequency (Freq) and the first two harmonics (Harm1 and Harm2) are the most influential parameters in species differentiation, representing 54.25%, 14.56%, and 11.60% importance, respectively, with the Random Forest method. While MFCC coefficients contribute less (around 1–2%), they still improve the model's performance. This importance distribution suggests that while frequencies and harmonics are primary parameters for classification, MFCCs may play a complementary role, especially in reducing classification errors due to frequency overlap observed in Fig. 5.

Table 1 compares the performance of classification models (SVM, k-NN, MLP, and Deep Learning) based on three distinct objectives:

- Objective 1: Use all available parameters, maximizing accuracy (up to 89.24% for MLP) but at the cost of high complexity.
- Objective 2: Parameter selection using Random Forest, Lasso, and RFE, simplifying the model while maintaining similar accuracy (89.24% for MLP), demonstrating the efficiency of parameter selection.
- Objective 3: Dimensionality reduction with PCA, which reduces model complexity and performance (83.15% for Deep Learning). Although this reduction simplifies the model, it results in information loss and reduces accuracy, making it less suitable for precise biological classification.

The results indicate that objective 2 represents the best balance between performance, complexity, and interpretability, particularly with the use of Random Forest for selecting the most discriminative parameters.

Fig. 7 compares confusion matrices for various parameter combinations, highlighting the impact of frequency (Freq) and the first harmonic (Harm1) on model accuracy. The combinations using only these two parameters (Fig. 3-d) achieve an accuracy of 44.14%, confirming that these two parameters alone are sufficient to distinguish many species.

The results presented in Fig. s 8 and 9 confirm that reducing the number of parameters improves the visualization and separation of mosquito species in PCA space. In Fig. 4, although 10 parameters were used, significant overlap is observed between specimens, notably between *A. arabiensis* and *A. coluzzii*. This overlap suggests that certain parameters, such as MFCC coefficients, do not substantially benefit mosquito

classification. This observation reinforces the idea that additional complexity from certain parameters does not always improve classification model performance.

In contrast, Fig. 9 illustrates that using two discriminative parameters—namely, the fundamental frequency (Freq) and the first harmonic (Harm1)—provides much better separation between species, particularly for mosquitoes like *Ae. aegypti* and *Cu. quinquefasciatus*. These results highlight the effectiveness of these two parameters for classification while limiting model complexity. This finding is supported by results from the confusion matrix (Fig. 7-d), where the use of Freq and Harm1 achieved the highest accuracy among the combinations tested (44.14%). This demonstrates that adding extra parameters is not necessary for improving classification performance, confirming the value of a minimalist approach.

These observations align with several previous studies emphasizing the importance of frequency- and harmonic-related parameters in insect classification. (Potamitis et al., 2015) demonstrated that the fundamental frequency and its harmonics are key indicators for distinguishing different insect species based on wingbeats. Similarly, (Potamitis & Rigakis, 2016) showed that fundamental frequencies and harmonic properties of wingbeats allow for distinguishing various insect species, such as *Cu. pipiens*, *molestus*, *A. gambiae*, and *Ae. albopictus*, using 2D optoacoustic sensors to record spectral properties.

When comparing our results to these studies, using simple parameters like the fundamental frequency and first harmonics constitutes a robust approach to mosquito species classification. Our results confirm this conclusion, showing that reducing the number of parameters can not only improve data visualization and classification accuracy but also prevent the risk of over-complexity. Thus, a streamlined approach based on essential discriminative parameters appears to be more effective than complex models that include redundant or uninformative variables.

The study (Vamsi et al., 2024), titled "Machine Learning-Based Classification of Mosquito Wing Beats Using Mel Spectrogram Images and Ensemble Modeling," takes a different approach, focusing on feature extraction from Mel spectrograms derived from mosquito wingbeats. The authors use convolution techniques to extract features, followed by ensemble models combining classifiers such as SVM, Random Forest, and Decision Tree. This model achieves an impressive accuracy of 95.05% for *Ae. albopictus*, demonstrating the effectiveness of advanced machine learning techniques for classification. However, although this approach offers very high accuracy, it relies on more complex feature extraction methods and similar models, which can increase computational complexity and associated implementation costs.

In comparison, our results show that it is possible to achieve reliable classification with much simpler parameters, such as the fundamental frequency and first harmonics, without significantly impacting accuracy. This minimalist approach can be particularly useful in applications where simplicity, fast execution, and low resource consumption are essential, especially for field monitoring systems or low-cost devices. The limitation of this study lies in the relatively low accuracy; we believe that adding other relevant parameters would improve this accuracy. While ensemble models and complex feature extraction techniques can offer remarkable accuracy, our study highlights the effectiveness of a reduced and optimized approach using key parameters for mosquito species classification. This method not only simplifies the models while enhancing data visualization in PCA space but also avoids over-complexity, which can hinder overall performance. Our results align with previous studies while providing a fresh perspective on the importance of simplicity and essential discriminative parameter selection to improve mosquito classification.

This study has demonstrated that reducing the number of parameters while retaining the most relevant ones, such as frequency and the first harmonic (Freq and Harm1), improves mosquito specimen identification while minimizing the complexity of classification models. The use of the Random Forest method for selecting discriminative parameters proved effective, and the integration of additional parameters, such as the degree of linear polarization, could be a promising avenue for further improving results, especially for species that are difficult to distinguish.

Acknowledgments

We thank the African Spectral Imaging Network (AFSIN) and Carla Puglia, coordinator of the International Science Program, Uppsala, for their continuous support.

Conflicts of interest

The authors declare no conflicts of interest.

REFERENCES

Benelli, G. (2016). Spread of Zika virus : The key role of mosquito vector control. *Asian Pacific Journal of Tropical Biomedicine*, *6*(6), 468-471.

Brydegaard, M., Gebru, A., & Svanberg, S. (2014). Super resolution laser radar with blinking atmospheric particles—-Application to interacting flying insects. *Progress In Electromagnetics Research*, *147*, 141-151.

Brydegaard, M., Kouakou, B., Jansson, S., Rydell, J., & Zoueu, J. (2021). High Dynamic Range in Entomological Scheimpflug Lidars. *IEEE Journal of Selected Topics in Quantum Electronics*, *27*(4), 1-11. IEEE Journal of Selected Topics in Quantum Electronics. https://doi.org/10.1109/JSTQE.2021.3062088

Brydegaard, M., Malmqvist, E., Jansson, S., Larsson, J., Török, S., & Zhao, G. (2017). The Scheimpflug lidar method. *Lidar Remote Sensing for Environmental Monitoring 2017*, *10406*, 104-120. https://doi.org/10.1117/12.2272939

Brydegaard, M., Merdasa, A., Gebru, A., Jayaweera, H., & Svanberg, S. (2016). Realistic Instrumentation Platform for Active and Passive Optical Remote Sensing. *Applied Spectroscopy*, *70*(2), 372-385. https://doi.org/10.1177/0003702815620564

Chen, H., Li, M., Månefjord, H., Travers, P., Salvador, J., Müller, L., Dreyer, D., Alison, J., Høye, T. T., Hu, G., & others. (2024). Lidar as a potential tool for monitoring migratory insects. *Iscience*, *27*(5).

Corbel, V., Akogbeto, M., Damien, G. B., Djenontin, A., Chandre, F., Rogier, C., Moiroux, N., Chabi, J., Banganna, B., & Padonou, G. G. (2012). Combination of malaria vector control interventions in pyrethroid resistance area in Benin : A cluster randomised controlled trial. *The Lancet infectious diseases*, *12*(8), 617-626.

Costantini, C., Gibson, G., Sagnon, N. F., Torre, A. D., Brady, J., & Coluzzi, M. (1996). Mosquito responses to carbon dioxide in B West African Sudan savanna village. *Medical and Veterinary Entomology*, *10*(3), 220-227. https://doi.org/10.1111/j.1365-2915.1996.tb00734.x

Dao, A., Yaro, A. S., Diallo, M., Timbiné, S., Huestis, D. L., Kassogué, Y., Traoré, A. I., Sanogo, Z. L., Samaké, D., & Lehmann, T. (2014). Signatures of aestivation and migration in Sahelian malaria mosquito populations. *Nature*, *516*(7531), 387-390.

de Cheveigné, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, *111*(4), 1917-1930. https://doi.org/10.1121/1.1458024

Ferguson, H. M., Dornhaus, A., Beeche, A., Borgemeister, C., Gottlieb, M., Mulla, M. S., Gimnig, J. E., Fish, D., & Killeen, G. F. (2010). Ecology : A prerequisite for malaria elimination and eradication. *PLoS medicine*, *7*(8), e1000303.

Gatton, M. L., Chitnis, N., Churcher, T., Donnelly, M. J., Ghani, A. C., Godfray, H. C. J., Gould, F., Hastings, I., Marshall, J., & Ranson, H. (2013). The importance of mosquito behavioural adaptations to malaria control in Africa. *Evolution*, *67*(4), 1218-1230.

Gebru, A., Jansson, S., Ignell, R., Kirkeby, C., Prangsma, J. C., & Brydegaard, M. (2018). Multiband modulation spectroscopy for the determination of sex and species of mosquitoes in flight. *Journal of biophotonics*, *11*(8), e201800014-e201800014.

Govella, N. J., Chaki, P. P., Geissbuhler, Y., Kannady, K., Okumu, F., Charlwood, J. D., Anderson, R. A., & Killeen, G. F. (2009). A new tent trap for sampling exophagic and endophagic members of the Anopheles gambiae complex. *Malaria Journal*, *8*(1), 157. https://doi.org/10.1186/1475-2875-8-157

Gubler, D. J. (1998). Resurgent vector-borne diseases as a global health problem. *Emerging infectious diseases*, *4*(3), 442.

Hagler, J. R., & Jackson, C. G. (2001). Methods for Marking Insects : Current Techniques and Future Prospects. *Annual Review of Entomology*, *46*(1), 511-543. https://doi.org/10.1146/annurev.ento.46.1.511

Hartfield, K. A., Landau, K. I., & Van Leeuwen, W. J. (2011). Fusion of high resolution aerial multispectral and LiDAR data : Land cover in the context of urban mosquito habitat. *Remote Sensing*, *3*(11), 2364-2383.

Jansson, S., Gebru, A., Ignell, R., Abbott, J., & Brydegaard, M. (2019). Correlation of mosquito wing-beat harmonics to aid in species classification and flight heading assessment. In A. Amelink & S. K. Nadkarni (Éds.), *Novel Biophotonics Techniques and Applications V* (p. 24). SPIE. https://doi.org/10.1117/12.2527224

Kenea, O., Balkew, M., Tekie, H., Gebre-Michael, T., Deressa, W., Loha, E., Lindtjørn, B., & Overgaard, H. J. (2017). Comparison of two adult mosquito sampling methods with human landing catches in south-central Ethiopia. *Malaria Journal*, *16*(1), 30. https://doi.org/10.1186/s12936-016-1668-9

Kirkeby, C., Wellenreuther, M., & Brydegaard, M. (2016). Observations of movement dynamics of flying insects using high resolution lidar. *Scientific Reports*, *6*, 29083. https://doi.org/10.1038/srep29083

Kouakou, B. K., Jansson, S., Jansson, S., Brydegaard, M., Brydegaard, M., & Zoueu, J. T. (2020). Entomological Scheimpflug lidar for estimating unique insect classes in-situ field test from Ivory Coast. *OSA Continuum*, *3*(9), 2362-2371. https://doi.org/10.1364/OSAC.387727

Li, Y., Wang, K., Quintero-Torres, R., Brick, R., Sokolov, A. V., & Scully, M. O. (2020). Insect flight velocity measurement with a CW near-IR Scheimpflug lidar system. *Optics Express*, *28*(15), 21891-21902-21891-21902.

Liu, Y.-N., Liu, Y.-J., Chen, Y.-C., Ma, H.-Y., & Lee, H.-Y. (2017). Enhancement of mosquito trapping efficiency by using pulse width modulated light emitting diodes. *Scientific reports*, *7*(1), 40074.

Maliti, D. V., Govella, N. J., Killeen, G. F., Mirzai, N.,

Johnson, P. C. D., Kreppel, K., & Ferguson, H. M. (2015). Development and evaluation of mosquito-electrocuting traps as alternatives to the human landing catch technique for sampling host-seeking malaria vectors. *Malaria Journal*, *14*(1), 502. https://doi.org/10.1186/s12936-015-1025-4

Malmqvist, E., Jansson, S., Török, S., & Brydegaard, M. (2016). Effective Parameterization of Laser Radar Observations of Atmospheric Fauna. *IEEE Journal of Selected Topics in Quantum Electronics*, *22*(3), 327-334. IEEE Journal of Selected Topics in Quantum Electronics. https://doi.org/10.1109/JSTQE.2015.2506616

Malmqvist, E., Jansson, S., Zhu, S., Li, W., Svanberg, K., Svanberg, S., Rydell, J., Song, Z., Bood, J., Brydegaard, M., & Åkesson, S. (2018). The bat–bird–bug battle : Daily flight activity of insects and their predators over a rice field revealed by high-resolution Scheimpflug Lidar. *Royal Society Open Science*, *5*(4), 172303. https://doi.org/10.1098/rsos.172303

Mboera, L. E. G., Knols, B. G. J., Braks, M. A. H., & Takken, W. (2000). Comparison of carbon dioxide-baited trapping systems for sampling outdoor mosquito populations in Tanzania. *Medical and Veterinary Entomology*, *14*(3), 257-263. https://doi.org/10.1046/j.1365-2915.2000.00239.x

Millette, T. L., Argow, B. A., Marcano, E., Hayward, C., Hopkinson, C. S., & Valentine, V. (2010). Salt marsh geomorphological analyses via integration of multitemporal multispectral remote sensing with LIDAR and GIS. *Journal of Coastal Research*, *26*(5), 809-816.

Murray, C. J., Rosenfeld, L. C., Lim, S. S., Andrews, K. G., Foreman, K. J., Haring, D., Fullman, N., Naghavi, M., Lozano, R., & Lopez, A. D. (2012). Global malaria mortality between 1980 and 2010 : A systematic analysis. *The Lancet*, *379*(9814), 413-431.

OMS. (2023). *Paludisme*. https://www.who.int/fr/news-room/fact-sheets/detail/malaria

Oppenheim, A. V., & Schafer, R. W. (2004). Dsp history - From frequency to quefrency : A history of the cepstrum. *IEEE Signal Processing Magazine*, *21*(5), 95-106. https://doi.org/10.1109/MSP.2004.1328092

Organization, W. H. (2015). *Global technical strategy for malaria 2016-2030*. World Health Organization. https://books.google.com/books?hl=fr&lr=&id=LV40DgAAQBAJ&oi=fnd&pg=PA1&dq=World+Health+Organization,+World+Malaria+Report+2016,+World+Health+Organization:+Geneva,+Switzerland&ots=kgqqwTCEFj&sig=UVF_0D9hGrnB-gGWAMM7mnZ_FsM

Organization, W. H. (2017). *Fourth meeting of the Vector Control Technical Expert Group (VCTEG) : Meeting report, 13-15 March 2017, Geneva, Switzerland*. World Health Organization. https://apps.who.int/iris/bitstream/handle/10665/255546/WHO-HTM-GMP-2017.11-eng.pdf

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. *Applied Acoustics*, *109*, 54-60. https://doi.org/10.1016/j.apacoust.2016.03.005

Potamitis, I., Rigakis, I., & Fysarakis, K. (2015). Insect biometrics : Optoacoustic signal processing and its applications to remote monitoring of McPhail type traps. *PloS one*, *10*(11), e0140474.

Repasky, K. S., Shaw, J. A., Scheppele, R., Melton, C., Carsten, J. L., & Spangler, L. H. (2006). Optical detection of honeybees by use of wing-beat modulation of scattered laser light for locating explosives and land mines. *Applied optics*, *45*(8), 1839-1843.

Silver, J. B. (2007). *Mosquito ecology : Field sampling methods*. springer science & business media. https://books.google.com/books?hl=fr&lr=&id=VM8MA4E_VT8C&oi=fnd&pg=PR15&dq=J.+B.+Silver,+Mosquito+Ecology:+Field+Sampling+Methods,+Springer:+Berlin,+Germany,+2007&ots=KF-1wFQMr5&sig=85_qyfvQIq8rWK_Li9pV_0K8ZxY

Vamsi, B., Al Bataineh, A., & Doppala, B. P. (2024). Machine Learning-Based Classification of Mosquito Wing Beats Using Mel Spectrogram Images and Ensemble Modeling. *Traitement Du Signal*, *41*(4), 2093-2101. https://doi.org/10.18280/ts.410437

Yohannes, M., & Boelee, E. (2012). Early biting rhythm in the afro-tropical vector of malaria, *Anopheles arabiensis* , and challenges for its control in Ethiopia. *Medical and Veterinary Entomology*, *26*(1), 103-105. https://doi.org/10.1111/j.1365-2915.2011.00955.x

Zheng, F., Zhang, G., & Song, Z. (2001). Comparison of different implementations of MFCC. *Journal of Computer Science and Technology*, *16*(6), 582-589. https://doi.org/10.1007/BF02943243