



OPEN Wing Interferential Patterns (WIPs) and machine learning for the classification of some *Aedes* species of medical interest

Arnaud Cannet^{1,9}, Camille Simon-Chane^{2,9}, Aymeric Histace^{2,9}, Mohammad Akhouni^{3,9}, Olivier Romain^{4,9}, Marc Souchaud^{2,9}, Pierre Jacob^{2,5}, Darian Sereno⁶, Louis-Clément Gouagna⁷, Philippe Bousset⁷, Françoise Mathieu-Daude^{7,8} & Denis Sereno^{6,7,9}✉

Hematophagous insects belonging to the *Aedes* genus are proven vectors of viral and filarial pathogens of medical interest. *Aedes albopictus* is an increasingly important vector because of its rapid worldwide expansion. In the context of global climate change and the emergence of zoonotic infectious diseases, identification tools with field application are required to strengthen efforts in the entomological survey of arthropods with medical interest. Large scales and proactive entomological surveys of *Aedes* mosquitoes need skilled technicians and/or costly technical equipment, further puzzled by the vast amount of named species. In this study, we developed an automatic classification system of *Aedes* species by taking advantage of the species-specific marker displayed by Wing Interferential Patterns. A database holding 494 photomicrographs of 24 *Aedes* spp. from which those documented with more than ten pictures have undergone a deep learning methodology to train a convolutional neural network and test its accuracy to classify samples at the genus, subgenus, and species taxonomic levels. We recorded an accuracy of 95% at the genus level and > 85% for two (*Ochlerotatus* and *Stegomyia*) out of three subgenera tested. Lastly, eight were accurately classified among the 10 *Aedes* sp. that have undergone a training process with an overall accuracy of > 70%. Altogether, these results demonstrate the potential of this methodology for *Aedes* species identification and will represent a tool for the future implementation of large-scale entomological surveys.

Pathogens (viruses, bacteria, parasites) transmitted by arthropods are among the most devastating infectious agents that scourge the human population worldwide. Hematophagous insects belonging to the *Aedes* genus are proven vectors of viral (chikungunya, Zika, dengue, Rift Valley fever, etc.) and filarial pathogens (*Brugia malayi*, *Wuchereria bancrofti*) of medical interest. The *Aedes* genus encompasses 79 subgenera, with more than 900 valid named species and 63 subspecies (<https://www.itis.gov/>). In 2000, a revision of the *Aedes* taxonomy was proposed, and the subgenus *Ochlerotatus* was “raised” to the level of the genus¹. Nevertheless, this remains a matter of debate², and a stable classification that considers utility and the current knowledge of the evolutionary relationship was proposed³. Still, much of the scientific literature refers to the subgenus *Ochlerotatus*. Out of 79 described subgenera, 15 contain named species with medical interests in transmitting viruses or parasites. Most *Aedes* species with medical interest fall into the *Ochlerotatus* (Lynch Arribalza, 1891), *Stegomyia* (Theobald, 1901), *Aedimorphus* (Meigen, 1818), or *Finlaya* (Theobald, 1903) subgenera. Among species with medical interest, only 2 disclose a worldwide distribution in 6 out of 7 continents: *Ae. aegypti* (Linnaeus 1762) and *Ae. albopictus*

¹Direction des affaires sanitaires et sociales de la Nouvelle-Calédonie, Nouméa, France. ²ETIS UMR 8051, Cergy Paris University, ENSEA, CNRS, 95000 Cergy, France. ³Parasitology-Mycology, Hôpital Avicenne, AP-HP, Bobigny, France. ⁴Cergy Paris University, Cergy, France. ⁵Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, 33400 Talence, France. ⁶InterTryp, Univ Montpellier, IRD-CIRAD, Infectiology Medical Entomology and One Health Research Group, Montpellier, France. ⁷MIVEGEC, Univ Montpellier, CNRS, IRD, Montpellier, France. ⁸Institut Louis Malardé, Tahiti, French Polynesia. ⁹These authors contributed equally: Arnaud Cannet, Camille Simon-Chane, Aymeric Histace, Mohammad Akhouni, Olivier Romain, Marc Souchaud and Denis Sereno. ✉email: denis.sereno@ird.fr

(Skuse, 1895). *Aedes albopictus*, a vector of public health importance, has encompassed a rapid change in its global distribution due to the worldwide trade in second-hand tires, often containing stagnant water, making it an ideal place for eggs and larvae, and its easiness to adapt to new environments, even in a temperate climate. This expansion offers opportunities for arboviruses (viruses transmitted by arthropods) to circulate in new areas, becoming a common cause of epidemics in *Ae. aegypti*-free countries. In addition to the mentioned species, at least seven others are indigenous mosquitoes on more than two continents (<https://www.worldatlas.com/continents> accessed on 01/24/2023).

For the entomological survey, identifying potential vectors belonging to the *Aedes* genus is typically based on intrinsic and extrinsic morphological criteria with morphological keys of determination as a tool. This is a time-consuming task that requires expertise and training. The discriminative morphological identification of adult *Ae. albopictus* is readily amenable by looking at bold black shiny scales and distinct silver-white scales on the palpus and tarsi⁴. Additionally, the dorsal scutum is black, with a distinguishing white stripe in the center, beginning at the dorsal surface of the head and continuing along the thorax. This character is nevertheless present in the specimen of the Scutellaris groups belonging to the *Stegomyia* subgenus, roughly 40 species. It cannot as perse be considered a diagnostic character. It is a medium-sized mosquito (2.0–10.0 mm; males are on average 20% smaller than females). Differential morphology of males from females includes the plumose antennae and modified mouthparts for nectar feeding. Dark scales cover the abdominal tergites. Legs are black with white basal scales on each tarsal segment. The abdomen narrows into a point characteristic of the genus *Aedes*. Adult *Ae. aegypti* can be recognized by white marks on the legs and a marking lyre on the thorax. In areas with multiple sympatric *Aedes* species with or without medical importance, these distinctions according to morphological characteristics are only sometimes easily discriminative. In addition, it is common for older adult specimens to have missed or damaged body parts or characters (e.g., scales, legs) that are essential for accurate identifications. Specimens with damage in critical regions for diagnostic characters separating vector species from closely related non-vector ones further puzzled entomological surveys.

Machine learning models, especially convolutional neural networks, can classify objects by identifying visible and non-visible features to the naked eye⁵. They are of choice and have been extensively used for insect identification involving whole insect image recognition. They demonstrate astonishing accuracy on a wide range of Arthropods^{6–8}, including Culicidae^{9–11}. Features related to animal behavior (e.g., flying and walking trajectories, postures, etc.)¹², allowing at-distance identification of alive specimens, have also been tested. Models based on insect morphology imaging of immobilized insects, an approach close to the entomological expertise deployed to identify insects that have inter-genus inter-species high morphological similarities, require a considerable number of data for training each Genus/species to learn the features and gain validation accuracy^{10,13,14}. Databases needed to train such models on whole insect recognition are filled with pictures of several poses, dorsal–ventral, etc., to collect taxonomic discriminant characters^{6,15–17}.

Wing Interference Patterns (WIPs) have received attention for their taxonomic potential^{18–20}. The thin-film interference occurring on the wings' transparent membrane allows the formation of a colored pattern. These WIPs significantly vary among specimens belonging to different species but moderately between representatives of the same species or between sexes. Unlike the angle-dependent iridescence effect of a flat film, the newton color series displayed is proportional to the thickness of the wing membrane at any given point, wing structures acting as diopters ensuring the WIPs appear essentially non-iridescent¹⁹. In previous papers, we have shown the value of WIPs for *Glossina* and *Anopheles* classification^{21,22}.

In the context of global climate change and the emergence of zoonotic infectious diseases, identification tools with field application are required to strengthen efforts in the entomological survey of arthropods with medical interest^{23,24}. Therefore, implementing new and affordable methods to accurately identify mosquitoes of the *Aedes* genus is a prerequisite for an entomological survey. To this aim, we have explored the accuracy and reliability of Wing Interferential Patterns (WIPs) to accurately identify *Aedes* specimens and classify them using a deep learning (DL) procedure.

Material and methods

Aedes selection and storage

The first WIPs reference collection of Culicidae gathers samples belonging to the *Aedes* genus using well-established *Aedes albopictus* laboratory breeds. Specimens were also selected in the ARIM collection (<https://arim.ird.fr/>) of IRD (Institut de Recherche pour le Développement). In addition, samples collected *in natura* whose identification was performed by trained entomologists at the time of their catch with available regional identification keys were also included in the database. The description of the samples used in this study is given in Table 1, and the worldwide distribution of samples included in the study is illustrated in Fig. 1. A complete list of the Diptera dataset used in this study and of species included in the database with their associated class is available in the supplementary data and the figshare server (Supdata: ID of species included in the study).

Image acquisition and database construction

The procedures applied to capture WIPs were the same as those described for *Glossina* sp. WIPs acquisition²¹. Wings were dissected and deposited on a glass slide, adding a small cover slide. The picture was taken with a Keyence™ VHX 1000 microscope, using the VH-Z20r camera and a Keyence VH K20 adapter, allowing an illumination incidence of 10°. The High Dynamic Range (HDR) function was used for all photos. All pictures were enlarged to a maximal occupancy, making the size of the wing not a discriminating parameter for insect species identification. The numerical parameters settled are the same as described and detailed in the publication on *Glossina* spp.²¹.

A <i>Aedes</i> spp. in our database	Med int*	Or	Year	N	Country code	Morphological identification by trained entomologists
<i>Aedes (Stegomyia) albopictus</i>	Yes	C-W	2015	239	250, 638	Ph Bousses
<i>Aedes (Stegomyia) aegypti</i>	Yes	W	2010	55	854, 266, 710	ND, Ph Bousses
<i>Aedes (Stegomyia) polynesiensis</i>	Yes	W	2016	27	876	F Mathieu-Daudé
<i>Aedes (Stegomyia) africanus</i>	Yes	W	2010	2	ND	JP Hervy
<i>Aedes (Stegomyia) opok</i>	No	W	ND	2	ND	JP Hervy
<i>Aedes (Stegomyia) luteocephalus</i>	Yes	W	1967	13	854	G Pichon, F Rodhain
<i>Aedes (Stegomyia) simpsoni</i>	No	W	ND	7	175	JP Hervy, ND
<i>Aedes (Ochlerotatus) detritus</i>	No	W	ND	22	504, 250	H Bailly-Choumara
<i>Aedes (Ochlerotatus) mariae</i>	No	W	1929	13	12	ND
<i>Aedes (Ochlerotatus) pullatus</i>	No	W	ND	15	ND	ND
<i>Aedes (Ochlerotatus) punctator</i>	No	W	1967	12	250	ND
<i>Aedes (Ochlerotatus) rusticus</i>	No	W	1967	8	250	ND
<i>Aedes (Ochlerotatus) dufouri</i>	No	W	2012	14	686	Ph Bousses
<i>Aedes (Ochlerotatus) cantans</i>	No	W	1960	9	250	H Bailly-Choumara
<i>Aedes (Neomelanicion) bolense</i>	No	W	1950	2	ND	ND
<i>Aedes (Neomelanicion) macintoshii</i>	Yes	W	1950	5	854	J Hamon
<i>Aedes (Neomelanicion) circumluteolus</i>	No	W	1955	5	ND	J Hamon
<i>Aedes (Aedimorphus) fowleri</i>	No	W	2012	3	638	Ph Bousses
<i>Aedes (Aedimorphus) irritans</i>	No	W	1986	7	686	JP Hervy
<i>Aedes (Diceromyia) furcifer</i>	Yes	W	1986	7	686	JP Hervy
<i>Aedes (Diceromyia) taylora</i>	Yes	W	1987	2	686	J Hamon
<i>Aedes (Finlaya) geniculatus</i>	No	W	2015	6	ND	G Le Goff
<i>Aedes (Finlaya) echinus</i>	No	W	1966	10	504	H Bailly-Choumara
<i>Aedes (Fredwardsius) vittatus</i>	Yes	W	1960	9	384	J Hamon
B <i>Aedes</i> spp. of medical interest not in our database						
<i>Aedes (Aedimorphus) cumminsii</i>						
<i>Aedes (Aedimorphus) dalzieli</i>						
<i>Aedes (Aedimorphus) dentatus</i>						
<i>Aedes (Aedimorphus) hirsutus</i>						
<i>Aedes (Aedimorphus) ochraceus</i>						
<i>Aedes (Aedimorphus) vexans</i>						
<i>Aedes (Catageomyia) minutus</i>						
<i>Aedes (Catageomyia) argenteopunctatus</i>						
<i>Aedes (Catageomyia) tarsalis</i>						
<i>Aedes (Downsiomyia) harinasutai</i>						
<i>Aedes (Downsiomyia) niveus</i>						
<i>Aedes (Elpeytonius) simulans</i>						
<i>Aedes (Finlaya) fijiensis</i>						
<i>Aedes (Finlaya) kochi</i>						
<i>Aedes (Finlaya) poicilius</i>						
<i>Aedes (Geogebraigius) atropalpus</i>						
<i>Aedes (Geogebraigius) epactius</i>						
<i>Aedes (Hulecoeteomyia) japonicus</i>						
<i>Aedes (Hulecoeteomyia) koreicus</i>						
<i>Aedes (Neomelanicion) palpalis</i>						
<i>Aedes (Ochlerotatus) angustivittatus</i>						
<i>Aedes (Ochlerotatus) atlanticus</i>						
<i>Aedes ((Ochlerotatus) canadensis</i>						
<i>Aedes (Ochlerotatus) cantator</i>						
<i>Aedes (Ochlerotatus) condolenscens</i>						
<i>Aedes (Ochlerotatus) dorsalis</i>						
<i>Aedes (Ochlerotatus) dupreei</i>						
<i>Aedes (Ochlerotatus) excrucians</i>						
<i>Aedes (Ochlerotatus) fitchii</i>						
<i>Aedes (Ochlerotatus) grossbecki</i>						
<i>Aedes (Ochlerotatus) infirmatus</i>						
Continued						

A <i>Aedes</i> spp. in our database	Med int*	Or	Year	N	Country code	Morphological identification by trained entomologists
<i>Aedes (Ochlerotatus) melanimon</i>						
<i>Aedes (Ochlerotatus) nigromaculis</i>						
<i>Aedes (Ochlerotatus) provocans</i>						
<i>Aedes (Ochlerotatus) scapularis</i>						
<i>Aedes (Ochlerotatus) sollicitans</i>						
<i>Aedes (Ochlerotatus) squamiger</i>						
<i>Aedes (Ochlerotatus) sticticus</i>						
<i>Aedes (Ochlerotatus) stimulans</i>						
<i>Aedes (Ochlerotatus) taeniorhynchus</i>						
<i>Aedes (Ochlerotatus) trivittatus</i>						
<i>Aedes (Ochlerotatus) vigilax</i>						
<i>Aedes (Polyleptomyia) albocephalus</i>						
<i>Aedes (Protomacleaya) triseratus</i>						
<i>Aedes (Rampomyia) notoscriptus</i>						
<i>Aedes (Stegomyia) scutellaris</i>						
<i>Aedes (Tanakiaius) togoi</i>						

Table 1. List of named *Aedes* species and description of samples included in the dataset (A) and of *Aedes* species having a medical interest but not included in our dataset (B). *Medical interest according to the WRBU database (https://wrbu.si.edu/vectorspecies?field_family_target_id=1194&title=&field_mt_products_tags_target_id=&field_pathogens_target_id=&field_geographic_locations_target_id=&items_per_page=30) and Wilkerson et al.³. & ISO 3166–1 country code available at (<https://www.atlas-monde.net/codes-iso/>). Or, the sample's origin; W, wild; C, colony; N, number of samples processed.

Collected dataset, image pre-processing, and dataset splitting for training/learning and validation

A rich annotated image dataset, including 494 pictures of 24 *Aedes* species, is available (figshare: <https://figshare.com/s/597f4e2f2d8206e09603>). Under-sampled *Aedes* species (less than ten samples/pictures) were discarded for training to prevent overfitting. All processed images were resized to 256 and 116 pixels for width and height, respectively. Pixel values were normalized within the (0, 1) range.

The dataset was then prepared for k-fold cross-validation, with k = 5 (Fig. 1). K-fold cross-validation is a classic approach to evaluate the robustness of a machine learning method, including Deep Learning ones. Mainly, for this study, the dataset was randomly shuffled and partitioned into k equal-size subsets with similar class distributions. A separately learned classifier was evaluated for each subgroup using kth of the whole dataset for validation and the remaining k–1 as training data (see Fig. 2 for illustration). This strategy allowed for measuring the mean accuracy of 5 distinct models and is the most accurate neural network performance estimation method²⁵.

This strategy allowed for measuring the mean accuracy of the five distinct classifiers. Among all existing machine learning methods, Deep Convolutional Neural Networks and their different architectures have shown in the last decade to be the most adapted for image classification. Compared to classic shallow methods (Support Vector Machine, Random Forest, and Boosting-based approaches for the main ones), they do not need hand-crafted features as input of the learning process: the selection of the best features is intrinsic to the method itself and is particularly well adapted to the WIPs scenario. A pipeline overview of the complete training procedure using CNN is shown in Fig. 3.

Training of the neural network (CNN)

The data used to develop the Deep Learning model includes Culicidae and non-Culicidae specimens²⁶. The methodology used was initially developed and tested on Tsetse flies²¹. Briefly, the original CNN architecture MobileNet²⁷, ResNet²⁸, and YOLOv2²⁹ architecture were deemed for automatically classifying *Aedes* sp. with the abovementioned dataset. Compared to classic Deep Learning, ours is more compact to cope with the specificity of our dataset in terms of size; therefore, thinner image recognition and classification architecture were developed to consider its reduced size. The first one is inspired by MobileNet, which takes advantage of depth-wise convolution²⁷. We propose to work with only one depth-wise convolution per layer of the CNN architecture to reduce the complexity and the number of extracted features. In addition, batch normalization was set to speed up and stabilize the training process³⁰.

In addition to this first compact CNN architecture based on MobileNet, two interconnected layers like VGG³¹ for YOLOv2 were applied with a DarkNet-19 architecture²⁹, as performed on tsetse flies samples²¹. As this architecture tends to over-fit the training set (meaning a lack of generalization of the performance when data other than the training data set are considered), we test two reduced architectures, i.e., using 1 or 2 scales less than the original network. For clarity, we called them DarkNet-9 (8 convolution layers and 1 classification layer) and DarkNet-14 (13 convolution layers and 1 classification layer). We also reproduced the ResNet18 architecture from He and collaborator²⁸ and trained it from random initialization. Even if this architecture seems too “deep”



Figure 1. Geographic distribution of samples included in the study using Google Looker Studio (<https://lookerstudio.google.com/overview>).

(may lead to overfitting) compared to our other architectures, the intrinsic properties of ResNet18, residual connections, allow convergence of the training procedure²¹.

We used a standard approach (shallow approach) based on extracting SURF descriptors (an efficient implementation of the classic SIFT descriptors), a Bag of Features (BoF) representation using a 4000 codewords dictionary, and an SVM with a standard polynomial kernel similar to it was proposed in Sereno et al.³². For each task, we only use one fully connected layer with the softmax activation to predict the probability that an image belongs to the correct class. We train our networks using Stochastic Gradient Descent (SGD) with a learning rate of 10^{-2} and a momentum of 0.9 for 30 epochs. The method was developed on a workstation with a quad-core CPU 3.0 GHz and 16Go RAM. Information on the training options, accuracy, and sensitivity, as well as the code, are available at <https://github.com/marcensea/diptera-wips/commit/12f39ab500a3f820cfb817c67ef25c580942301d>.

Results

Training and classification

We explored the training classifier accuracy on the *Aedes* dataset and on datasets of Culicidae that do not belong to the *Aedes* genus (non-*Aedes*) and from mosquitoes that do not belong to the Culicidae family (non-Culicidae) as negative samples. We trained the CNN on such a combination to improve the model’s accuracy. The database initially filled with a total of 494 pictures of *Aedes* spp. WIPs pictures. Since we cannot fill the database with all



Figure 2. Schematic representation of the dataset splitting for learning (red) and testing (orange) from Cannet et al.²¹.

Aedes species acting as primary vectors of viruses, parasites, or bacteria with a medical interest, we first focused on a restricted set of species (Table 1). Using this pictures-set, we ascertained the accuracy of the process to discriminate the *Aedes* genus (Meigen, 1818) from other Culicidae (Meigen, 1818), including Anophelinae (Grassi, 1900), Culicinae (Meigen, 1818), and non-Culicidae samples members (e.g., Psychodidae, Glossiniidae...). From our dataset and process, the automatic classification process accuracy is rather good and higher than 95% (Table 2).

We also evaluated the DL process's accuracy at the subgenus level (Table 3). The *Aedes* genus encompasses 79 subgenera, with only 14 gathering species documented to act as primary vectors of viruses, parasites, and bacteria of medical interest. As mentioned in the introductory section, the subgenus *Ochlerotatus* was "raised" to the level of the genus in 2000¹. Considering that a large number of the scientific literature refers to the subgenus *Ochlerotatus*, we refer to this taxonomic nomenclature to carry out our analysis³³. Our *Aedes* dataset does not represent the overall diversity at the subgenera level. Nevertheless, the computed accuracy for the identification process is high (> 85%) for the *Stegomyia* and *Ochlerotatus* subgenera, with a classification accuracy of 97.0 and 86.7%, respectively. Still, it fails to address the *Finlaya* subgenus accurately. Such low accuracy might be due to the small sample size available to train and test the process for *Finlaya*. Therefore, efforts in implementing the database with enough specimens representative of the *Finlaya* and other subgenera must be made to address our methodology's accuracy in addressing the classification process at the subgenus taxonomic level. Overall, our methods are the first based on deep learning with the versatility of automatically handling samples at the subgenus level, which might be of interest to speed and facilitate the identification of specimens of *Aedes* species of medical interest.

Finally, the reliability of the DL model to accurately classify WIPs pictures of 10 *Aedes* species was calculated, and the results are presented in Table 4. Variable level of accuracy is recorded, ranging from not accurate (0.00%) to perfect classification (100.00%). A perfect accuracy (100.00%) is achieved for 50% of *Aedes* species whose WIPs pictures are filled in the dataset. More than 75% accuracy in classification is recorded for 7 *Aedes* species, but the DL methods failed to assign 2 *Aedes* species correctly (*Ae. echinus* and *Ae. mariae*) (Table 4). Even if populations of *Ae. albopictus* bear a high variability independently of whether the populations are native or invasive^{34,35}, we record a high classification accuracy (Table 4). The DL approach based on WIPs can manage the diversity of specimens sampled in two distant geographic areas, France and La Réunion. *Aedes aegypti* exists in two subspecies or forms: the domestic ecotype, *Ae. aegypti aegypti* found outside Africa and the presumed ancestral *Ae. aegypti formosus* occurring in sub-Saharan Africa in sympatry with *Ae. aegypti aegypti* in some ecologies^{36,37}. Our samples of *Ae. aegypti* originating from sub-Saharan Africa were not identified at the subspecies level. They may be a mix of the two ecotypes with subtle variations in their WIPs. This might be the underlying factor triggering the lower accuracy (72.7%) of the identification process of *Ae. aegypti* in our dataset, as compared to the accuracy recorded for *Ae. albopictus* (95.8%) classification. This deserves to be investigated. Nevertheless, the overall calculated specific recognition of samples belonging to *Aedes* remains satisfactory.

Misclassified pictures

Inspection of the machine learning model weaknesses helps identify underlying problems. This can be performed via a review of the miss-predicted images. This will provide insights into what makes a photo hard to classify for the model. Our data set takes advantage of having to be filled with pictures of samples collected *in natura* or from laboratory colonies but also preserved for an extended period or damaged. This large panel of images taken with the same microscope under the same light condition, preparation process, and preprocessing methodologies will help to delineate factors (e.g., age, sample origin, sample preservation...) affecting the classification process. In Fig. 4, selected examples are presented. Deep learning models rely more on textures than shapes. A more extensive training set can avoid pitfalls linked to photo or sample quality, prevent confusing set-ups when taking pictures, and improve the accuracy of the automated classification. A guideline can be added to the application to advise participants to make good images of Culicidae samples.

Besides misclassified pictures, we generated activations maps pointing to which image region the architectures use, according to the feature extraction layers of the DCNN (Fig. 5).

These activation maps illustrate that the membrane and the vein structure are not targeted by the DCNN and that features activated concentrate in the lower part of the picture, particularly on the color pattern, when learning was successful Fig. 5 and high accuracy in the classification (Accuracy > 76%), while when classification

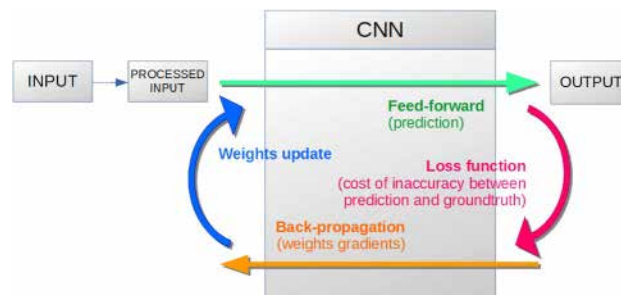


Figure 3. Schematic representation of the pipeline process developed for *Glossina* identification using the Convolutional Neural Network (CNN) approach from Cannet et al.²¹.

		Predicted		
		<i>Aedes</i> spp.	Other Culicidae	Non Culicidae
Truth	<i>Aedes</i> spp. N = 84	80 (95.2%)	4	0
	Other Culicidae N = 259	2	256 (98.8%)	1
	Non Culicidae N = 675	0	1	674 (99.9%)

Table 2. Test for accuracy of the DL (Deep Learning) process for the *Aedes* (Meigen, 1818) genus identification. N, number of pictures in the test subset. Accuracy data (%) in bold.

		Predicted			
		Subgenera	<i>Stegomyia</i>	<i>Ochlerotatus</i>	<i>Finlaya</i>
Truth	<i>Stegomyia</i> N=67	97.0%	0.0%	0.0%	3.0%
	<i>Ochlerotatus</i> N=15	0.0%	86.7%	0.0%	13.3%
	<i>Finlaya</i> N=2	50.0%	50.0%	0.0%	0.0%
	Non <i>Aedes</i> spp. N=934	0.0%	0.2%	0.0%	99.8%

Table 3. Test for accuracy (%) of the DL classification process at the subgenus taxonomic level. N, number of pictures in the test subset, Accuracy in %.

accuracy is low (30%), features extracted spread all over the wing. In all cases, neither the wing shape nor the venation are features targeted by DCNN.

Discussion

With robust vector surveillance, emerging vector-borne disease threats can be addressed. During an entomological survey, most routine identification involves skilled personnel that use diagnostic criteria, like the size, shape, or texture of specimens or the presence or absence of certain key features. Since the morphological identifications of medically important *Aedes* species is complex and meticulous, requiring highly skilled specialists, an alternative is to use computer vision that, like mosquito taxonomists, relies on visual (morphological) characters to assign a definitive taxon name. A framework based on a deep neural network was developed to extract mosquito anatomy from images. The process was tested on 23 Culicidae belonging to the *Aedes*, *Anopheles*, and *Culex* Genus⁹, but only 3 *Aedes* species (*Ae. aegypti*, *Ae. infirmatus*, and *Ae. taeniorhynchus*). Nevertheless, manipulating mosquitoes to acquire and extract morphological characters is complicated and damages specimens. A real-time pipeline was set but has only been tested for discrimination of *Ae. aegypti* and *Ae. albopictus*¹⁵. Finally, deep convolutional network approaches using whole-body pictures of field-caught mosquitoes have shown promise, demonstrating astonishing accuracy, even classifying unknown/unidentified species before further identification¹⁰. Nevertheless, we can anticipate that morphology-based methods might suffer from the same limitation encountered by taxonomists for cryptic species/subspecies delineations, for *Aedes* species belonging

				Predicted										
				1	2	3	4	5	6	7	8	9	10	11
Truth	Species	N	Class											
	<i>Ae. albopictus</i>	48	1	95.8	4.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	<i>Ae. aegypti</i>	11	2	0.0	72.7	0.0	0.0	0.0	0.0	0.0	9.1	0.0	0.0	18.2
	<i>Ae. detritus</i>	4	3	0.0	0.0	75.0	0.0	0.0	0.0	25.0	0.0	0.0	0.0	0.0
	<i>Ae. mariaae</i>	3	4	0.0	0.0	0.0	33.3	0.0	0.0	0.0	0.0	0.0	0.0	66.7
	<i>Ae. pullatus</i>	3	5	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	<i>Ae. punctator</i>	2	6	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
	<i>Ae. luteocephalus</i>	3	7	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
	<i>Ae. dufouri</i>	3	8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
	<i>Ae. echinus</i>	2	9	50.0	0.0	0.0	0.0	50.0	0.0	0.0	0.0	0.0	0.0	0.0
	<i>Ae. polynesiensis</i>	5	10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
	Non- <i>Aedes</i> spp.	934	11	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	99.8

Table 4. Accuracy in % of the DL process for *Aedes* spp. classification. Significant values are in [italics]. N, number of images in the test subset.

to the Neomelanoconion identification whose require examination of internal organs, and as well as the misidentification of damaged specimens. Besides the mentioned methods, some characteristics like flight tone and wing beat, which play a role in mosquito reproduction^{38–40}, were used for the automated identification assay of *Ae. albopictus* and *Ae. triseriatus*^{39,40}, and more recent development permits an “at a distance” discrimination between *Ae. albopictus*, *Ae. aegypti* and *Culex* sp.^{41,42}. Here, we prove the classification accuracy with WIPs is above 75% for 8 of the ten species investigated in our study.

Morphometry analysis of insects’ wings has demonstrated interest in vector mosquitoes and sibling or cryptic species^{43–45} identification. Nevertheless, this requires intact wings and extensive preprocessing to mount them without deformation. All that severely impedes its application for entomological survey purposes, but it is still attractive in taxonomic and populational studies. Our methodology would promise that *Aedes* species identification can be accurately performed with minimal training, cost, and time. In addition, wings are relatively robust insect organs that limit the risks of damaging them. In addition, they can be easily stored at room temperature and be carried with minimal caution.

Alternative methodologies not relying on morphological characteristics were implemented. These alternatives are DNA barcoding and protein profiling by MALDI-TOF, but they can also involve the wing beat features. MALDI-TOF has been applied to a restricted number of *Aedes* species at the egg stage, *i.e.*, *Ae. albopictus*, *Ae. atropalpus*, *Ae. cretinus*, *Ae. geniculatus*, *Ae. koreicus*, *Ae. phoeniciae*, and *Ae. triseriatus*, 4 have medical importance^{46,47}. It was also set up to identify the second larval stage to pupae of *Ae. albopictus* and *Ae. aegypti*⁴⁸ and on the fourth instar larvae⁴⁹ but with higher accuracy for late larval instar⁵⁰. Its efficiency for field-trapped or colony-reared insect adults is limited. It includes species having (*Ae. albopictus*, *Ae. vigilax*, *Ae. vexans*, *Ae. notoscriptus*, *Ae. polynesiensis*) or not (*Ae. fowleri*, and *Ae. dufouri*) a medical importance^{51–55}. However, the interoperability of the MALDI-TOF methodology requires a standardized procedure in the conservation of samples and the choice of the mosquito body part of adult specimens⁵⁶. For MS identification, standardization of procedures for preparation and reproducibility between instruments and homemade databases will be desirable^{57–60}. Contrary to MALDI-TOF or DNA-barcoding methodologies, our method is not designed to identify *Aedes* at larval stages. Nevertheless, a survey on adult mosquitoes is an alternative to larval surveys because it addresses the vector life stage responsible for transmission and thus adds information on vector density⁶¹. Infrared spectroscopy (NIR and MIR) can detect changes in the mosquito cuticle by quantifying light absorbed⁶². The discriminative capability of such methodology at the species level has been investigated on a restricted number of *Aedes* species having a medical interest (*Aedes aegypti*, *Ae. albopictus*, *Ae. japonicus*, and *Ae. triseriatus*)⁶³. The main interest of this methodology relies on its capacity to grade the age of natural *Anopheles* spp. populations when coupled with a Deep learning approach⁶⁴. DNA barcoding is also an alternative to morphological identification methods. A quest for “*Aedes*” as a keyword in the barcode of Life system (BOLD) database indicates 27,458 records with sequences representative of 328 species (<http://v4.boldsystems.org/> accessed on 21 of July 2022)⁶⁵. During proactive surveillance of invasive and/or medically and veterinary important species, genetic identification by PCR is costly, destructive, and time-consuming⁶⁶, a limitation that does not carry our method. Prompt and correct identification of exogenous species is required to monitor entry points under the International Health Regulations (IHR) to limit the dispersal and establishment of new vector species. In this case, even entomologists experienced in identifying local mosquito fauna may experience difficulty accurately identifying damaged specimens. We provide evidence that WIPs analysis coupled with Deep Learning for discriminative identification competency to accurately classify *Aedes* species at various taxonomic levels (genus, subgenus, and species). The competency to accurately classify *Aedes* samples at the subgenus level is of interest to medical entomologists and public health services since a restricted number of *Aedes* subgenus encompass species of medical interest. In this study, among *Aedes* species accurately classified (Accuracy > 75%), four are of medical interest (*Ae. albopictus*, *Ae. aegypti*, *Ae. luteocephalus*, and *Ae. polynesiensis*). Therefore, we can envision that a database of WIPs encompassing proven *Aedes* vectors at a global or regional scale would allow remote territories to address the vector status of a specimen.

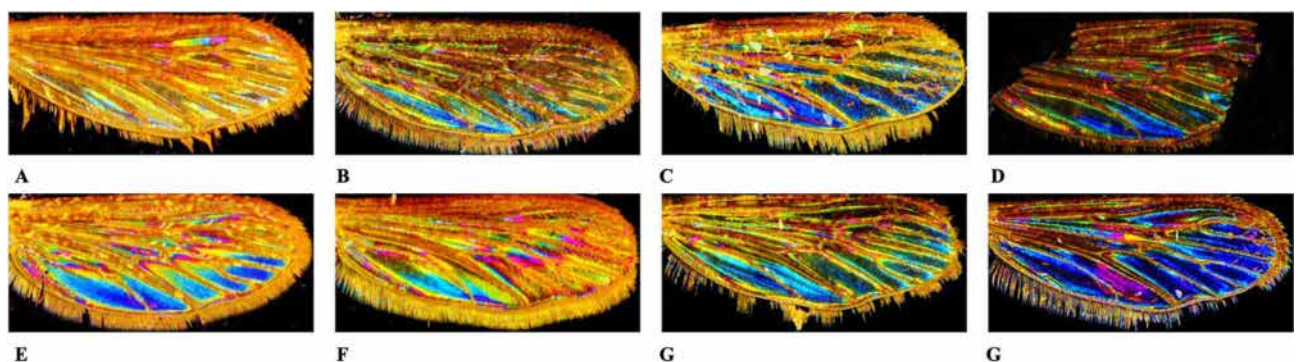


Figure 4. Selected examples of misclassified pictures. *Ae. albopictus* misidentified as *Ae. aegypti* (A), *Ae. albopictus* misidentified as *Ae. aegypti* (B), *Ae. aegypti* misidentified as *Ae. luteocephalus* (C), *Ae. aegypti* misidentified as *Ae. albopictus* (D), *Ae. mariaei* misidentified as *Cx. neavei* (E), *Ae. echinus* misidentified as *Ae. pullatus* (F), *Ae. polynesiensis* misidentified as *Ae. albopictus* (G), *Ae. polynesiensis* misidentified as *Cx. neavei* (H).

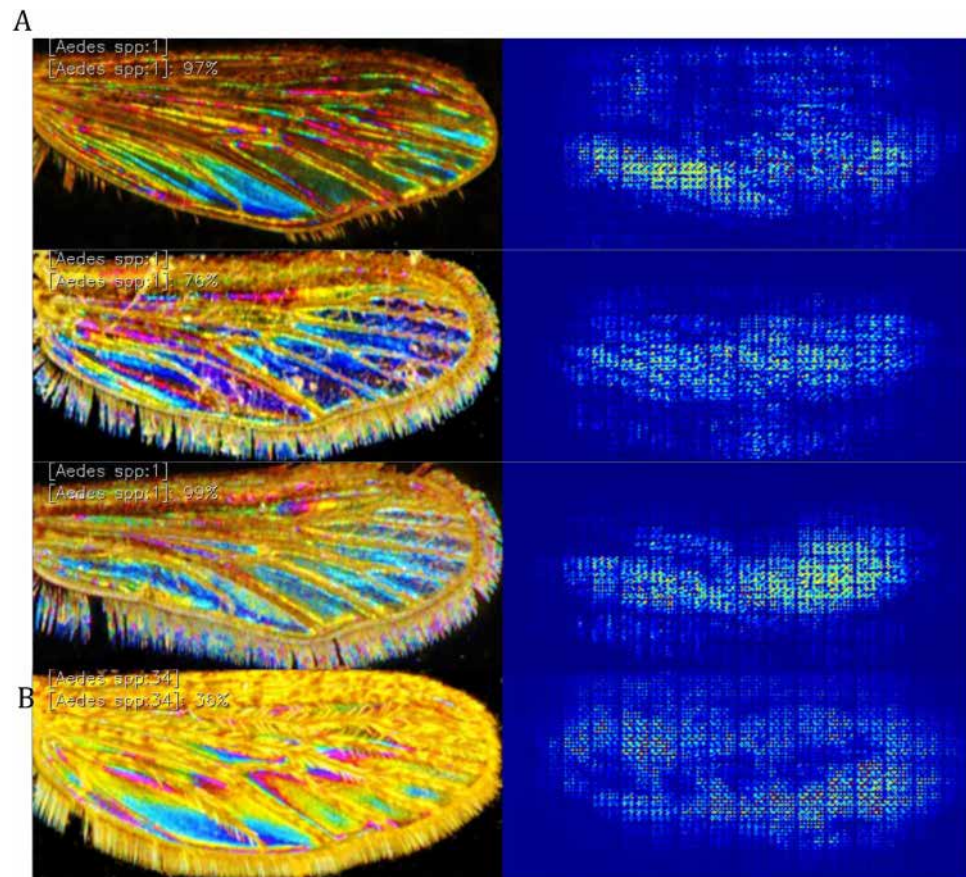


Figure 5. Activation maps for *Aedes* with a high confidence value (A) and low confidence value (B) on the prediction. Warm colors (yellow and red) of the activation point to wing regions where local features are essential for the decision.

Increasing the area of vector surveillance is of high interest, even at a continental scale. Community science relying on the participation of non-expert people was therefore applied for mosquitoes^{67,68}. Unfortunately, the quality of entomological data can be poor in terms of identification accuracy. Deep learning approaches have offered opportunities to overcome this limitation and have been developed to follow *Ae. albopictus* in Europe, from pictures taken by citizen¹¹. However, this methodology wasn't tested in high *Aedes* species diversity areas. Unfortunately, our method, as it stands, cannot be translated for this purpose because of the cost associated with microscopy and the need for high-quality pictures. Nevertheless, because wings are tiny and light, they can quickly be sent for identification purposes with this methodology. A global comparison between the various methods developed for *Aedes* identification is given in Table 5.

Current global vector surveillance systems are unstandardized, burdensome on public health systems, and threatened by the worldwide scarcity of entomologists. The Deep learning analysis of *Aedes* WIPs resulting in robust classification could empower non-experts to identify disease vectors accurately and rapidly in the field. This is of interest in areas where large-scale and/or proactive surveys must be performed, like the Pacific area, where more than 20,000 islands must be surveyed^{56,69}. This situation would make exogenous species introduction unnoticed. Nevertheless, the methods need now to be further investigated by evaluating, even qualitatively, whether the proposed approach could be usable in real-life scenarios regarding several important criteria: cost (including infrastructure, material, training, and cost associated with technically skilled personnel), computational resources, analyzing time, sample destructiveness, diversity, and taxonomic level (species, sibling species, cryptic species, sub-species) of the classification. All these will increase vector surveillance's programmatic capability and capacity with minimal training and cost.

Methods	Conservation procedure for identification	Technical cost*	Computational cost	Effort	Sample destruction	Currently available for <i>Aedes</i> sp.
DNA Barcoding	– 20 °C	High	Low	High	Yes	328/900 <i>Aedes</i> sp.
Protein Profiling (MALDI-TOF)	4 °C at best	High	Low	Low	Yes/part	7/900 <i>Aedes</i> sp.
Dichotomous keys	RT	High	Low	High	No/yes\$	Yes regional
Near Infrared spectroscopy	– 20 °C	Medium	Low	Low	No	4/900
WIPs + Deep Learning (ours)	RT	Low	Medium	Low	No	10/900 <i>Aedes</i> sp.

Table 5. Synthetic view of advantages and limitations of some identification methodologies for medically important arthropods in the framework of pro-active or large-scale entomological surveys. Number of *Aedes* species tested. *The technical cost includes sample preparation (DNA, protein extraction, slides, etc.) and expenses linked to the need for a skilled professional and the presence of dedicated infrastructure for running experiments.

Data availability and code availability

The whole database, including *Aedes* species WIPs, is available on <https://figshare.com/s/597f4e2f2d8206e09603>, and the code on <https://github.com/marcensea/diptera-wips.git>.

Received: 14 June 2023; Accepted: 13 October 2023

Published online: 17 October 2023

References

- Reinert, J. F. New classification for the composite genus *Aedes* (Diptera: Culicidae: Aedini), elevation of subgenus *Ochlerotatus* to generic rank, reclassification of the other subgenera, and notes on certain subgenera and species. *J. Am. Mosq. Control Assoc.* **16**, 175–188 (2000).
- Savage, H. M. & Strickman, D. The genus and subgenus categories within Culicidae and placement of *Ochlerotatus* as a subgenus of *Aedes*. *J. Am. Mosq. Control Assoc.* **20**, 208–214 (2004).
- Wilkerson, R. C. *et al.* Making mosquito taxonomy useful: A stable classification of tribe Aedini that balances utility with current knowledge of evolutionary relationships. *PLoS One* **10**, e0133602. <https://doi.org/10.1371/journal.pone.0133602> (2015).
- Hawley, W. A. The biology of *Aedes albopictus*. *J. Am. Mosq. Control Assoc. Suppl.* **1**, 1–39 (1988).
- Caramazza, P. *et al.* Neural network identification of people hidden from view with a single-pixel, single-photon detector. *Sci. Rep.* **8**, 11945. <https://doi.org/10.1038/s41598-018-30390-0> (2018).
- Abeywardhana, D. L., Dangalle, C. D., Nugaliyadde, A. & Mallawarachchi, Y. An ultra-specific image dataset for automated insect identification. *Multimedia Tools Appl.* **81**, 3223–3251. <https://doi.org/10.1007/s11042-021-11693-3> (2022).
- Ding, W. & Taylor, G. Automatic moth detection from trap images for pest management. *Comput. Electron. Agric.* **123**, 17–28. <https://doi.org/10.1016/j.compag.2016.02.003> (2016).
- Wang, J., Lin, C., Ji, L. & Liang, A. A new automatic identification system of insect images at the order level. *Knowl.-Based Syst.* **33**, 102–110. <https://doi.org/10.1016/j.knosys.2012.03.014> (2012).
- Minakshi, M., Bharti, P., Bhuiyan, T., Kariev, S. & Chellappan, S. A framework based on deep neural networks to extract anatomy of mosquitoes from images. *Sci. Rep.* **10**, 13059. <https://doi.org/10.1038/s41598-020-69964-2> (2020).
- Goodwin, A. *et al.* Mosquito species identification using convolutional neural networks with a multitiered ensemble model for novel species detection. *Sci. Rep.* **11**, 13656. <https://doi.org/10.1038/s41598-021-92891-9> (2021).
- Pataki, B. A. *et al.* Deep learning identification for citizen science surveillance of tiger mosquitoes. *Sci. Rep.* **11**, 4718. <https://doi.org/10.1038/s41598-021-83657-4> (2021).
- Tannous, M., Stefanini, C. & Romano, D. A Deep-Learning-Based detection approach for the identification of insect species of economic importance. *Insects* **14**, 148 (2023).
- Zhu, L.-Q. *et al.* Hybrid deep learning for automated lepidopteran insect image classification. *Oriental Insects* **51**, 79–91. <https://doi.org/10.1080/00305316.2016.1252805> (2017).
- Hansen, O. L. P. *et al.* Species-level image classification with convolutional neural network enables insect identification from habitus images. *Ecol. Evol.* **10**, 737–747. <https://doi.org/10.1002/ece3.5921> (2020).
- Ong, S.-Q., Ahmad, H., Nair, G., Isawasan, P. & Majid, A. H. A. Implementation of a deep learning model for automated classification of *Aedes aegypti* (Linnaeus) and *Aedes albopictus* (Skuse) in real time. *Sci. Rep.* **11**, 9908. <https://doi.org/10.1038/s41598-021-89365-3> (2021).
- Ong, S.-Q. & Ahmad, H. An annotated image dataset of medically and forensically important flies for deep learning model training. *Sci. Data* **9**, 510. <https://doi.org/10.1038/s41597-022-01627-5> (2022).
- Kittichai, V. *et al.* Automatic identification of medically important mosquitoes using embedded learning approach-based image-retrieval system. *Sci. Rep.* **13**, 10609. <https://doi.org/10.1038/s41598-023-37574-3> (2023).
- Buffington, L. M. & Sandler, J. R. The occurrence and phylogenetic implications of wing interference patterns in Cynipoidea (Insecta: Hymenoptera). *Invertebr. Syst.* **25**, 586–597 (2012).
- Shevtsova, E., Hansson, C., Janzen, D. H. & Kjærandsen, J. Stable structural color patterns displayed on transparent insect wings. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 668–673. <https://doi.org/10.1073/pnas.1017393108> (2011).
- Simon, E. Preliminary study of wing interference patterns (WIPs) in some species of soft scale (Hemiptera, Sternorrhyncha, Coccoidea, Coccidae). *Zookeys* **269–281**, 2013. <https://doi.org/10.3897/zookeys.319.4219> (2013).
- Cannet, A. *et al.* Wing interferential patterns (WIPs) and machine learning, a step toward automatized tsetse (*Glossina* spp.) identification. *Sci. Rep.* **12**, 20086. <https://doi.org/10.1038/s41598-022-24522-w> (2022).
- Cannet, A. *et al.* Deep learning and wing interferential patterns identify *Anopheles* species and discriminate amongst Gambiae complex species. *Sci. Rep.* **13**, 13895. <https://doi.org/10.1038/s41598-023-41114-4> (2023).
- Reed, S. Biodiversity. Pushing DAISY. *Science* **328**, 1628–1629. <https://doi.org/10.1126/science.328.5986.1628> (2010).
- MacLeod, N., Benfield, M. & Culverhouse, P. Time to automate identification. *Nature* **467**, 154–155. <https://doi.org/10.1038/467154a> (2010).
- Sohil, F., Sohali, M. U. & Shabbir, J. An introduction to statistical learning with applications in R. *Stat. Theory Relat. Fields* **6**, 87–87. <https://doi.org/10.1080/24754269.2021.1980261> (2022).

26. Sereno, D. *et al.* Listing and pictures of Diptera WIPs. *Sci. Rep.* <https://doi.org/10.6084/m9.figshare.22083050.v1> (2023).
27. Howard, A. G. *et al.* MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv:abs/1704.04861* (2017).
28. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778 (2016).
29. Redmon, J. & Farhadi, A. YOLO9000: Better, faster, stronger. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 6517–6525 (2017).
30. Ioffe, S. & Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv:abs/1502.03167* (2015).
31. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
32. Sereno, D., Cannet, A., Akhouni, M., Romain, O. & Histace, A. Système et procédé d'identification automatisée de diptères hématophages. France PCT/FR15/000229. patent (2015).
33. Soares, I. M. N., Polonio, J. C., Zequi, J. A. C. & Goliás, H. C. Molecular techniques for the taxonomy of *Aedes* Meigen, 1818 (Culicidae: Aedini): A review of studies from 2010 to 2021. *Acta Trop.* **236**, 106694. <https://doi.org/10.1016/j.actatropica.2022.106694> (2022).
34. Paupy, C. *et al.* Comparative role of *Aedes albopictus* and *Aedes aegypti* in the emergence of Dengue and Chikungunya in Central Africa. *Vector-Borne Zoonot. Dis.* **10**, 259–266. <https://doi.org/10.1089/vbz.2009.0005> (2010).
35. Goubert, C., Minard, G., Vieira, C. & Boulesteix, M. Population genetics of the Asian tiger mosquito *Aedes albopictus*, an invasive vector of human diseases. *Heredity* **117**, 125–134. <https://doi.org/10.1038/hdy.2016.35> (2016).
36. Tchouassi, D. P., Agha, S. B., Villinger, J., Sang, R. & Torto, B. The distinctive bionomics of *Aedes aegypti* populations in Africa. *Curr. Opin. Insect Sci.* **54**, 100986. <https://doi.org/10.1016/j.cois.2022.100986> (2022).
37. Futami, K. *et al.* Geographical distribution of *Aedes aegypti aegypti* and *Aedes aegypti formosus* (Diptera: Culicidae) in Kenya and environmental factors related to their relative abundance. *J. Med. Entomol.* **57**, 772–779 (2020).
38. Sotavalta, O. Flight-tone and wing-stroke frequency of insects and the dynamics of insect flight. *Nature* **170**, 1057–1058. <https://doi.org/10.1038/1701057a0> (1952).
39. Moore, A., Miller, J. R., Tabashnik, B. E. & Gage, S. H. Automated identification of flying insects by analysis of wingbeat frequencies. *J. Econ. Entomol.* **79**, 1703–1706. <https://doi.org/10.1093/jee/79.6.1703> (1986).
40. Moore, A. Artificial neural network trained to identify mosquitoes in flight. *J. Insect Behav.* **4**, 391–396. <https://doi.org/10.1007/BF01048285> (1991).
41. Rydhmer, K. *et al.* Automating insect monitoring using unsupervised near-infrared sensors. *Sci. Rep.* **12**, 2603. <https://doi.org/10.1038/s41598-022-06439-6> (2022).
42. Genoud, A. P., Basistyy, R., Williams, G. M. & Thomas, B. P. Optical remote sensing for monitoring flying mosquitoes, gender identification and discussion on species identification. *Appl. Phys. B* **124**, 3. <https://doi.org/10.1007/s00340-018-6917-x> (2018).
43. Wilke, A. B. *et al.* Morphometric wing characters as a tool for mosquito identification. *PLoS One* **11**, e0161643. <https://doi.org/10.1371/journal.pone.0161643> (2016).
44. Dujardin, J. P. *et al.* Outline-based morphometrics, an overlooked method in arthropod studies?. *Infect. Genet. Evol.* **28**, 704–714. <https://doi.org/10.1016/j.meegid.2014.07.035> (2014).
45. Martinet, J. P. *et al.* Wing Morphometrics of *Aedes* mosquitoes from North-Eastern France. *Insects* **12**, 896. <https://doi.org/10.3390/insects12040341> (2021).
46. Schaffner, F., Kaufmann, C., Pflüger, V. & Mathis, A. Rapid protein profiling facilitates surveillance of invasive mosquito species. *Parasit. Vectors* **7**, 142. <https://doi.org/10.1186/1756-3305-7-142> (2014).
47. Suter, T. *et al.* First report of the invasive mosquito species *Aedes koreicus* in the Swiss-Italian border region. *Parasit. Vectors* **8**, 402. <https://doi.org/10.1186/s13071-015-1010-3> (2015).
48. Dieme, C. *et al.* Accurate identification of Culicidae at aquatic developmental stages by MALDI-TOF MS profiling. *Parasit. Vectors* **7**, 544. <https://doi.org/10.1186/s13071-014-0544-0> (2014).
49. Nebbak, A. & Almeras, L. Identification of *Aedes* mosquitoes by MALDI-TOF MS biotyping using protein signatures from larval and pupal exuviae. *Parasit. Vectors* **13**, 161. <https://doi.org/10.1186/s13071-020-04029-x> (2020).
50. Nebbak, A. *et al.* Field application of MALDI-TOF MS on mosquito larvae identification. *Parasitology* **145**, 677–687. <https://doi.org/10.1017/s0031182017001354> (2018).
51. Abdellahoum, Z. *et al.* Identification of Algerian field-caught mosquito vectors by MALDI-TOF MS. *Vet. Parasitol. Reg. Stud. Rep.* **31**, 100735. <https://doi.org/10.1016/j.vprsr.2022.100735> (2022).
52. Tandina, F. *et al.* Using MALDI-TOF MS to identify mosquitoes collected in Mali and their blood meals. *Parasitology* **145**, 1170–1182. <https://doi.org/10.1017/s0031182018000070> (2018).
53. Yssouf, A. *et al.* Matrix-assisted laser desorption/ionization–time of flight mass spectrometry: An emerging tool for the rapid identification of mosquito vectors. *PLoS One* **8**, e72380. <https://doi.org/10.1371/journal.pone.0072380> (2013).
54. Huynh, L. N. *et al.* MALDI-TOF mass spectrometry identification of mosquitoes collected in Vietnam. *Parasit. Vectors* **15**, 39. <https://doi.org/10.1186/s13071-022-05149-2> (2022).
55. Fall, F. K., Laroche, M., Bossin, H., Musso, D. & Parola, P. Performance of MALDI-TOF mass spectrometry to determine the sex of mosquitoes and identify specific colonies from French Polynesia. *Am. J. Trop. Med. Hyg.* **104**, 1907–1916. <https://doi.org/10.4269/ajtmh.20-0031> (2021).
56. Rakotonirina, A. *et al.* MALDI-TOF MS: An effective tool for a global surveillance of dengue vector species. *PLoS One* **17**, e0276488. <https://doi.org/10.1371/journal.pone.0276488> (2022).
57. Rakotonirina, A. *et al.* MALDI-TOF MS: Optimization for future uses in entomological surveillance and identification of mosquitoes from New Caledonia. *Parasit. Vectors* **13**, 359. <https://doi.org/10.1186/s13071-020-04234-8> (2020).
58. Bamou, R. *et al.* Enhanced procedures for mosquito identification by MALDI-TOF MS. *Parasit. Vectors* **15**, 240. <https://doi.org/10.1186/s13071-022-05361-0> (2022).
59. Vega-Rúa, A. *et al.* Improvement of mosquito identification by MALDI-TOF MS biotyping using protein signatures from two body parts. *Parasit. Vectors* **11**, 574. <https://doi.org/10.1186/s13071-018-3157-1> (2018).
60. Kostrzewa, M. & Maier, T. In *MALDI-TOF and Tandem MS for Clinical Microbiology* 39–54 (2017).
61. Leandro, A. S. *et al.* Citywide integrated *Aedes aegypti* mosquito surveillance as early warning system for arbovirus transmission, Brazil. *Emerg. Infect. Dis.* **28**, 701–706. <https://doi.org/10.3201/eid2804.211547> (2022).
62. Mayagaya, V. S. *et al.* Non-destructive determination of age and species of *Anopheles gambiae s.l.* using near-infrared spectroscopy. *Am. J. Trop. Med. Hyg.* **81**, 622–630. <https://doi.org/10.4269/ajtmh.2009.09-0192> (2009).
63. Srout, L., Byrd, B. D. & Huffman, S. W. Classification of mosquitoes with infrared spectroscopy and partial least squares-discriminant analysis. *Appl. Spectrosc.* **74**, 900–912. <https://doi.org/10.1177/0003702820915729> (2020).
64. Siria, D. J. *et al.* Rapid age-grading and species identification of natural mosquitoes for malaria surveillance. *Nat. Commun.* **13**, 1501. <https://doi.org/10.1038/s41467-022-28980-8> (2022).
65. Savolainen, V., Cowan, R. S., Vogler, A. P., Roderick, G. K. & Lane, R. Towards writing the encyclopedia of life: An introduction to DNA barcoding. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **360**, 1805–1811. <https://doi.org/10.1098/rstb.2005.1730> (2005).
66. Schaffner, F. *et al.* Development of guidelines for the surveillance of invasive mosquitoes in Europe. *Parasit. Vector* **6**, 209 (2013).

67. Braz Sousa, L. *et al.* Citizen science and smartphone e-entomology enables low-cost upscaling of mosquito surveillance. *Sci. Total Environ.* **704**, 135349. <https://doi.org/10.1016/j.scitotenv.2019.135349> (2020).
68. Jordan, R. C., Sorensen, A. E. & Ladeau, S. Citizen science as a tool for mosquito control. *J. Am. Mosq. Control Assoc.* **33**, 241–245. <https://doi.org/10.2987/17-6644r.1> (2017).
69. Anderson, L., Sopoaga, F. & Jack, S. *Aedes* mosquito control and surveillance in the Pacific. *Pac. Health Diag.* **21**, 226–232. <https://doi.org/10.26635/phd.2020.623> (2020).

Acknowledgements

We thank Pr. P. Marty and P. Delaunay (CHU Nice) for gaining access to the microscopic facility of the CHU. Dr. D. Fontenille (UMR MIVEGEC, Montpellier, France) for his support and fruitfully scientific discussions on medical entomology aspects. Mr JP Commes, former CEO of 2CSI, for his enthusiasm and advice on the digital aspect of the project.

Author contributions

Conceptualisation D.S., A.C., M.A., A.H., C.S.C., O.R., M.S. Data acquisition D.S., A.C., M.S., A.H., D.S. Database construction D.S., D.S., A.H., P.J., M.S., O.R. Sample collection and arthropod management P.B., F.M.D., L.C.G. Project management D.S., A.H., C.S.C. Writing first draft A.H., D.S., A.C. Writing and editing D.S., A.C., M.A., C.S.C., P.B., A.H., F.M.D., L.C.G.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-44945-3>.

Correspondence and requests for materials should be addressed to D.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023