JITMM Proceedings
www.proceedings.jitmm.com
JITMM Proceedings 12 (2024) 36-50

DEEP LEARNING TECHNOLOGY FOR FIELD-BASE MOSQUITO VECTOR IDENTIFICATION

Songpol Eiamsamang[1], Santhad Chuwongin[2], Peeraphon Promma[2], Yudthana Samung[1], Atiporn Saeung[3], Kittipong Chaisiri[4], Theeraphap Chareonviriyaphap[5], Patchara Sriwichai[1]

[1]*Department of Medical Entomology, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand*
[2]*Center of Industrial Robot and Automation (CiRA), College of Advanced Manufacturing Innovation, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand*
[3]*Center of Insect Vector study, Department of Parasitology, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand*
[4]*Department of Helminthology, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand*
[5]*Department of Entomology, Faculty of Agriculture, Kasetsart University, Bangkok, Thailand*

ABSTRACT

The incorporation of deep learning enhances the morphological identification process have been improved. However, there has been limited investigation into the model's accuracy concerning species complexity. This study aimed to enhance mosquito identification using multi-stage deep learning on the CiRA CORE platform. An image collection of 175,000 pictures from 7 mosquito species obtained from the laboratory and field strains (*Ae. aegypti, Ae. albopictus, Anopheles minimus, An. harrisoni, An. dirus, An. maculatus, and Culex quinquefasciatus*). Hundred mosquito samples from each species was photographed with minimum 10 different view sides. The images were subsequently put through the augmentation process 25 times to increase the numbers of images to 25,000 images per species. Model evaluation, based on 173 observations, showed that the model achieved 90±5% accuracy in distinguishing among genus levels (*Aedes, Anopheles*, and *Culex* groups). Optimization processes demonstrated the model's accuracy in *Ae. aegypti* 99%, *Ae. albopictus* 99%, *An. minimus* 94%, *An. harrisoni* 86%, *An. dirus* 98%, *An. maculatus* 98%, and *Cx. quinquefasciatus* 98%. We further evaluated the mosquito identification efficacy between AI and public health officers using 30 unknown images among the seven total mosquito species. Results showed no significant difference in species classification between the AI system and public health officers (*P* value > 0.05). Surprisingly, at the species complex identification level, the AI system demonstrated a significant 90% accuracy advantage over public health officers (*P* value < 0.05). This AI system represents an optional tool to support vector surveillance in the local public health officers, enabling faster and more accurate mosquito monitoring.

Keywords: Artificial Intelligence, Mosquitoes Identification, Deep Learning

Correspondence: Patchara Sriwichai, Department of Medical Entomology, Faculty of Tropical Medicine, Mahidol University, 420/6 Ratchawithi Road, Ratchathewi Bangkok 10400, Thailand
Email: patchara.sri@mahidol.ac.th;

INTRODUCTION

Mosquitoes are a major global problem vectors of many diseases (e.g., Malaria, Dengue, West Nile Fever, and most recently Zika Fever). Over 100,000 people worldwide die from mosquito-borne diseases every year (WHO malaria report 2018). Moreover, the increasing globalization that leads to expanding the habitats of many vector species. The surveillance of mosquitoes relies on capture by the variation of several collecting methods such as human-baited, animal-baited or artificial-baited trap, which requires regular manual inspection, dedicated personnel, making large-scale monitoring difficult and expensive (Motta *et al*, 2019). Moreover, mosquito classification is the classical important method for medical investigation of vector-borne disease that can determine the relationship between vector and disease including epidemiology distribution, and the potential of transmission diseases. In general, the morphological-base microscopic identification is a practical routine process which is inexpensive and requires minimal tools and/or equipment. Although it requires well training and needs a high level of expertise, the species complex is still limited.

Deep learning machines have been applied to entomological work in many publications, such as the geometric morphometrics technique (Jame Rohlf and Bookstein, 1990) and the Buzz of wingbeats detection (Li *et al*, 2005; Tai-Hsien *et al*, 2015; Kiskin *et al*, 2017) which were applied by mosquito's wing vein and wing beat sound. Moreover, visualization deep learning method have been reported with successful application of the innovation of deep learning framework to use in mosquito morphological identification (Huang *et al*, 2018; Minakshi *et al*, 2020; Joshi and Miller, 2021). Improvement of the accurate toolkit for visualization identification (Mulchandani *et al*, 2019) can narrow the gap of the entomological field.

Currently, entomological identification have been developed by artificial intelligence technology for classification of images (Orlando *et al*, 2015; Goodwin 2021). However, the optimization of accuracy and apply in the practical field remains challenge. This research aimed to optimize conditions of artificial intelligence technology for mosquito vectors classification for increasing the accuracy and efficiency. The trained model could be further applied in practical field applications.

MATERIALS AND METHODS

Ethics Statement

Animal and Human ethics documents was submitted to the Ethics Committee (EC) at the Faculty of Tropical Medicine on January 2023. The certificate number: FTM-ACUC 002/2023.

Sample collections

Laboratory specimens from the Insectary, Department of Medical Entomology, Faculty of Tropical Medicine, Mahidol University, were used in the initial phase for training on complete morphology characteristics. Field samples, collected from cow bait traps and larvae collection technique, focused on *Anopheles* mosquitoes in districts across Thailand (Tha Song Yang District, Tak province; Saiyok district, Kanchanaburi province; and Suanphueng district, Ratchaburi province). One hundred mosquitoes per species was targeted in the collection. Adult mosquito specimens were kept in 1.5 mL microtubes with silica beads. Larval collections were transferred to the laboratory of the Department of Medical Entomology, Faculty of Tropical Medicine, Mahidol University, and maintained until emerging into the adult stage. Mosquitoes were determined to species by morphological characters (Rattanarithikul *et al*, 2006). For long-term preservation, the mosquitoes were stored at -20°C in microtube boxes. The dataset was created to classify among the difference species of *Ae. albopictus, Ae. aegypti, Ae. albopictus, An. minimus, An. dirus, An. maculatus, Cx. quinquefasciatus*. All mosquito pictures were separated into 3 groups (deep training group, validate group and testing group). All visualized data were labeled by a professional entomologist under the CiRA CORE program.

Species complex confirmation

One of the common problems in mosquito classification is complex species that have the same physical characteristics but clearly different DNA. *Anopheles minimus* is the candidate complex species of malaria vector available in Thailand (Taai *et al*., 2017). After morphology identification, every *An. minimus* A and *An. harrisoni* specimen were cut in the middle leg for confirmation through the modify PCR (Polymerase Chain Reaction) method using the AS-PCR assay based on ITS2 rDNA sequences (Taai *et al*, 2017). Genomic DNA was extracted from individual adult females using the DNeasy® Blood and Tissue Kit (Qiagen, Hilden, Germany), and the isolated DNA underwent sequential PCR procedures. The ITS2 region was amplified using the universal forward primer ITS2A (5′-TGT GAA CTG CAG GAC ACA T-3′) and the specific reverse primers MIA (5′-CCC GTG CGA CTT GAC GA-3′ for *An. minimus*) and MIC (5′-GTT CAT TCA GCA ACA TCA GT-3′ for *An. harrisoni*). PCR was carried out in 25 µL volumes containing 0.5 U of Taq DNA polymerase, 1× Taq buffer, 2.0 mM of MgCl2, 0.2 mM of each dNTP, 0.25 µM of each primer, and 1 µL of the extracted DNA. The amplification profile comprised initial denaturation at 94 °C for 2 min, 30 cycles at 94 °C for 30 s, 45 °C for 30 s, and 72 °C for 40 s, followed by a final extension at 72 °C for 5 min. The amplified products were electrophoresed on a 1.5% agarose gel. Additionally, PCR products were sequenced using the BigDye® Terminator Cycle Sequencing Kit and analyzed with the 3130 genetic analyzer for species confirmation.

Dataset

The dataset comprises images sourced from the training group, including *Ae. aegypti, Ae. albopictus, Anopheles minimus, An. harrisoni, An. dirus, An. maculatus,* and *Culex quinquefasciatus*. Each species' sample images were meticulously prepared in 10 different views, encompassing whole body, right and left-lateral view, dorsal view, ventral view, as well as specific parts such as head, head+thorax, wing, leg, and abdomen. To enhance diversity, these images underwent augmentation, resulting in 25,000 variations per species, as depicted in Figure 1. The augmentation process involved techniques such as image rotation, vertical and horizontal flips, and adjustments for sharpness, highlights, and shadows. Captured using a Nikon automatic high-resolution camera and a smartphone, these images were trained into the AI system, accommodating both high-resolution and mobile pictures for practical field applications in mosquito data collection. The high-resolution camera operated at 2048 x 1365 pixels, while smartphone pictures were captured at 1920 x 1080 pixels, directly under a stereomicroscope.

Dataset sample size comparison

The purpose of the experiment was to determine the relationship between sample size and accuracy in the context of species identification. The variation of mosquito images among 40, 60, 80, and 100 images were prepared and compared the percentages of accuracy and precision levels by The Receiver Operating Characteristic (ROC) curve of Deep classification.

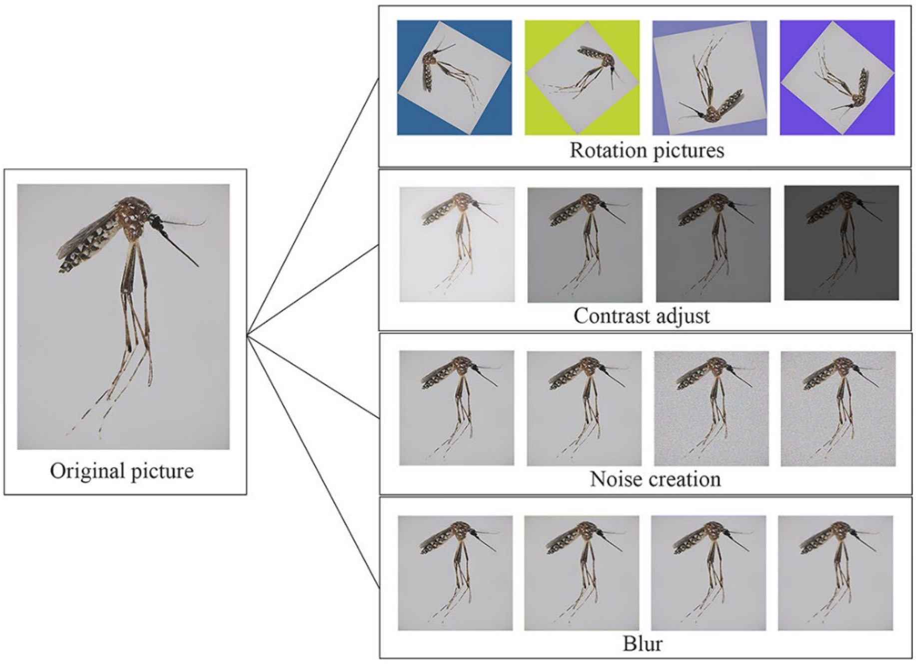


Figure 1 Picture augmentation technique for mosquito dataset.

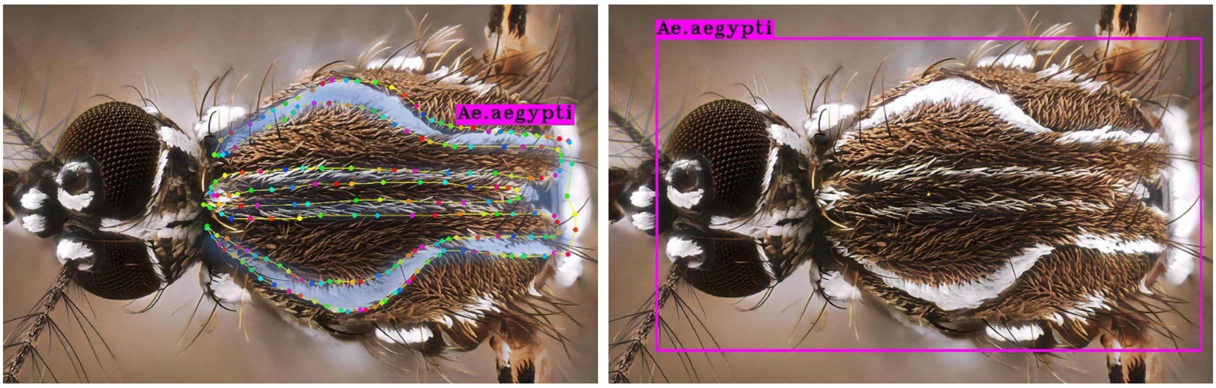


Figure 2 **Deep learning techniques.** In Deep detection (left), the labeling area specifically targets the identification of *Ae. aegypti* mosquitoes. Conversely, in Deep Classification (right), identification of *Ae. aegypti* is achieved through a non-specific point on mosquito pictures.

Network selection

The Deep learning model operated on CIRA CORE Software. This software is developed from the Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang. For GPU processing by nVidia DIGITS software with a NVIDIA Geforce RTX 2070 super. Software were processed on the Linux Ubuntu operating system LTS Distribution 16.04. In this research, the modified-learning methods were trained on two-stage YOLO versions. Recently, an experiment indicated that the accuracy rate of the two-stage version was significantly higher than the one-stage YOLO version. Project conducted with personal computer contain with CPU: Intel Core i5 8500, Ram: 16Gb, And GPU: RTX 2070 super 8Gb running on Linux Ubuntu 16.03 operation.

Deep learning techniques

The CiRA CORE software offers two main deep training methods for image recognition: Deep classification and Deep detection. Deep detection, illustrated in Figure 2 (left), involves labeling the focus area to “extract” entomological characteristics of mosquitoes, from genus to species level, using techniques like YOLO V3 model. Conversely, Deep classification, depicted in Figure 2 (right), focuses on mass datasets of mosquito pictures, using the Classif-Function in CIRA CORE to classify multiple pictures per species into groups like "Non-extraction morphology."

Approach model to Entomological taxonomy logic

Our objective was to develop an artificial intelligence model capable of species identification comparable to that of an entomologist. Illustrated in Figure 3, the workflow outlines the multi-layer models employed for mosquito identification. The dataset comprises images exhibiting distinct morphological features among various mosquito species. Initially, a decision is made to classify an image as a "mosquito". Subsequently, focus shifts to the morphological features of each organ. The dataset is annotated to identify four specific regions: 1) the upper region encompassing all organs on the head and thorax, 2) the entirety of the wings, 3) the abdominal boundary extending from the thorax to the final segment, and 4) the legs. The last model undergoes extensive training for accurate identification of mosquito species.

Mosquito identification comparison between human and deep learning

The efficacy of mosquito identification was assessed through a comparison between human and AI capabilities in identifying mosquito species from a total of 30 pictures. Twenty entomologists specializing in field surveillance, selected from the Bureau of Communicable Diseases division and the Department of Disease Control of Tak province, Ratchaburi province, and Kanchanaburi province, participated in the evaluation. The 30 unknown specimens pictured (Table1), focusing on mosquito vectors in Thailand, were separated into two groups for identification. The first group comprised basic identification by genus and species (images 1-16), while the second group consisted of specimens requiring more careful identification at the cryptic level (images 17-30).

Data analysis and statistical method

The ROC curve illustrates the performance of a binary classifier across different decision thresholds, comparing True Positive Rate (the rate of correctly identified positives) to False Positive Rate (the rate of incorrectly identified negatives) at various settings. An ideal classifier achieves TPR=1 and FPR=0. This curve offers a visual representation of the trade-off between TPR and FPR, facilitating comparison of classifier effectiveness. Widely utilized in machine learning, particularly in fields such as medical diagnosis, where the consequences of missing a positive condition or incorrectly identifying a negative one is significant, the ROC curve was generated by plotting the cumulative distribution function of the detection probability against that of the false-alarm probability. This analysis provided insights into the model's performance across different discrimination thresholds.

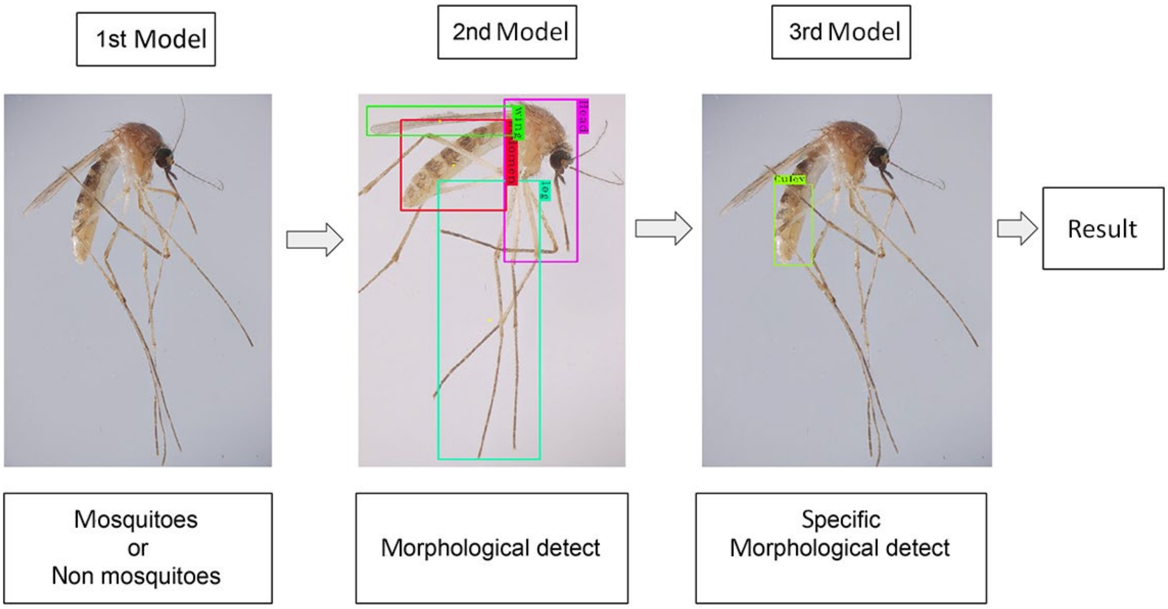


Figure 3 Workflow of a deep learning model with three layers for mosquito identification.

Table 1 Mosquitoes species for identification comparing between public health officers and Deep learning

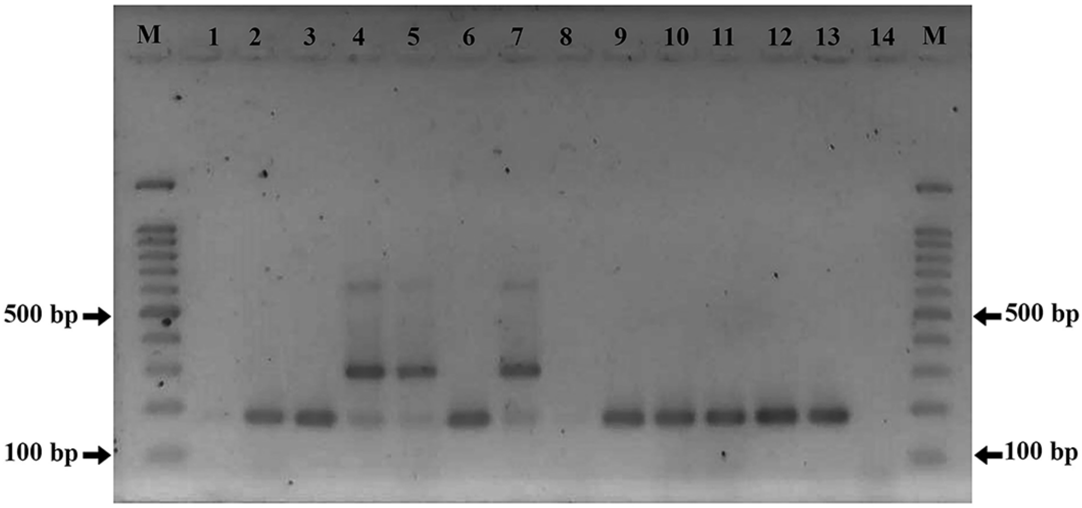
No.	Mosquitoes species	No.	Mosquitoes species
1	*Aedes albopictus*	16	*Anopheles maculatus*
2	*Culex quinquefasciatus*	17	*Anopheles harrisoni*
3	*Aedes albopictus*	18	*Anopheles dirus*
4	*Culex quinquefasciatus*	19	*Anopheles harrisoni*
5	*Aedes aegypti*	20	*Anopheles maculatus*
6	*Aedes aegypti*	21	*Anopheles harrisoni*
7	*Aedes albopictus*	22	*Anopheles dirus*
8	*Anopheles dirus*	23	*Anopheles harrisoni*
9	*Anopheles maculatus*	24	*Anopheles harrisoni*
10	*Anopheles minimus*	25	*Anopheles harrisoni*
11	*Aedes albopictus*	26	*Anopheles minimus*
12	*Aedes aegypti*	27	*Anopheles harrisoni*
13	*Anopheles minimus*	28	*Anopheles minimus*
14	*Anopheles dirus*	29	*Anopheles minimus*
15	*Anopheles minimus*	30	*Anopheles harrisoni*

Table 2 The number of total mosquito specimens for model training dataset

Species	Laboratory Specimen	Fields Specimen	Total
Ae. albopictus	100	100	200
Ae. aegypti	100	100	200
An. minimus	100	100	200
An. harrisoni		24	24
An. dirus	100	100	200
An. maculatus		200	200
Cx. quinquefasciatus	100	100	200
Total	500	724	1224

A total of 1,220 natural early-stage Anopheles larvae (L1-L2) were collected from field sites. We could obtain 29 immerging adults. Molecular techniques, illustrated in Figure 4, were employed to identify complex species. Among 29 adults morphologically identified as *An. harrisoni*, molecular analysis confirmed 24 specimens as *An. harrisoni* and 4 as *An. minimus* A, while 1 species remained unidentified.

Figure 4 Species complex identification by PCR for the identification of *An. minimus* A and *An. harrisoni*. In Lanes 4-5 and 7, *An. minimus* A is represented by a 310 bp band. Lanes 1-3, 6 and 8-13 show *An. harrisoni* with a 180 bp band. Lane M displays a 100 bp ladder for size comparison.



RESULTS

Sample collection and species identification

A total of 100 specimens each of *Ae. albopictus, Ae. aegypti, An. minimus, An. dirus*, and *Cx. quinquefasciatus* were collected from fields, along with an additional 100 specimens from laboratory sources, for training and identification to mosquito. Additionally, 24 specimens of *An. harrisonai* from Kanchanaburi and 100 specimens of *An. minimus* A from Tha Song Yang, Tak Province, were included in the dataset (Table2).

Dataset sample size comparison

To ascertain and enhance sample size's impact on the accuracy of species identification, our results revealed a positive correlation between dataset numbers and model accuracy. Specifically, varying sample sizes of 40, 60, 80, and 100 images exhibited mean accuracies of 81.4%, 87.3%, 89.0%, and 93.3%, respectively. These findings underscored the significance of sample size in model training and suggested the potential for further accuracy improvement through larger datasets.

Picture augmentation

Based on our experiments, the accuracy of mosquito identification using the Deep classification model ranged from 66% to 99% (Figure 5-left). We observed a positive correlation between increasing picture augmentations and improved accuracy, particularly beneficial for identifying *An. harisoni* and *An. minimus*. This phenomenon can be attributed to their complex morphological similarities.

Multilayer model application

The initial layer employs the deep detection method to identify the overall mosquito object, achieving an average precision of up to 98%. Subsequently, the second layer of the network model evaluated separate mosquito morphological parts—head, abdomen, wing, and leg—utilizing deep detection methods. The mean Average Precision (mAP) for this stage was 85% (ranging from 72% to 93%). In the third layer, responsible for species identification through detection of species-specific characteristics, the system demonstrates the ability to identify mosquito components with high accuracy: up to 90% for the head (ranging from 77% to 90%), 94% for wings (ranging from 89% to 98%), 93% for abdomen (ranging from 87% to 99%), and 90% for legs (ranging from 71% to 92%). Deep detection techniques were compared at this stage, with the following average precision scores for species identification as follows: *Cx. quinquefasciatus*: 95.40%, *Ae. aegypti*: 97.92%, *Ae. albopictus*: 99.00%, *An. minimus*: 96.19%, *An. harisoni*: 83.54%, *An. maculatus*: 93.73%, *An. dirus*: 96.71%. These results are illustrated in the precision-recall graph in Figure 5-right.

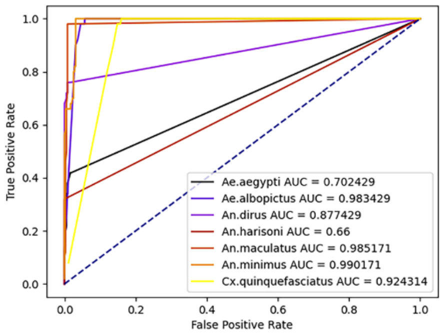
Species complex training and evaluation

In our study, we used *An. minimus* group by specific to *An. harrisoni* and *An. minimus* A as the samples for AI training model and evaluation from total 24 *An. harrisoni* while 100 *An. minimus* from Thasong Yang district, Tak province. The samples were imaged and used for training by Deep classification model and deep detection. The result showed that species complex can be used “Deep detection” technique greater than “Deep classification” at 66% (Figure 5-left) and 86% (Figure 5-right) respectively.

Comparison of image-based mosquito identification between humans and Deep learning technologies

Table 3 compares the identification scores of 30 unknown mosquito images across seven species between human participants (20 public health officers) and the Deep Learning model. The mosquito identification images were divided into two groups: 16 randomly selected pictures for basic skill assessment and 14 unknown pictures representing species at a complex level, potentially requiring a dichotomous key. Identification scores were obtained from 20 human participants and 20 instances of Deep Learning. Overall, human identification scores ranged from 43.33% to 76.67%, whereas the Deep Learning model achieved identification rates ranging from 66.67% to 83.33%. Interestingly, no significant differences in species classification were observed between the AI system and public health officers for the total of 30 unknown images (*P* value > 0.05). Surprisingly, in the species complex identification level (images 17-30), the AI system exhibited a statistically significant 99% accuracy advantage over public health officers (*P* value < 0.001). The raw data indicates a higher incidence of incorrect identifications by humans compared to the Deep Learning model in the blue box area (Table3).

Figure 5 The multiple layer identification by Deep classification showed the Receiver Operating Characteristic (ROC) curve of Deep classification for mosquito's species identification (left). The Precision recall Graph of first layer model that work for classify mosquitoes from picture by deep detection (right).



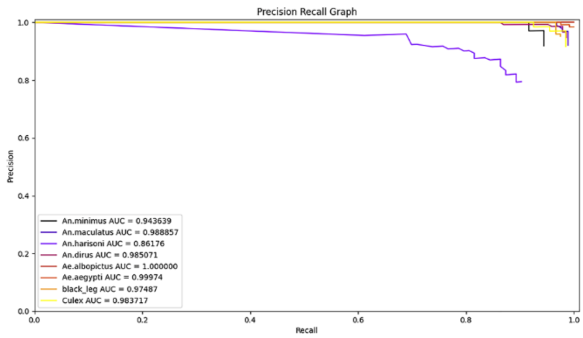


Table 3 The mosquito identification scores of 30 mosquito unknown images by 20 public health staffs and the Deep Learning model.

Occupation	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	Score	Identification score
Entomologist1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✓	✓	✗	18	60.00%
Entomologist2	✓	✗	✗	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✗	✓	✗	✗	✗	✓	✗	✓	✗	✗	17	56.67%
Entomologist3	✓	✓	✗	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	15	50.00%
Entomologist4	✓	✓	✗	✓	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓	✗	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	13	43.33%
Public health officer1	✓	✗	✗	✗	✓	✗	✗	✓	✓	✓	✓	✓	✓	✗	✓	✓	✗	✓	✗	✗	✗	✓	✗	✗	✗	✓	✗	✓	✓	✗	15	50.00%
Public health officer2	✓	✓	✗	✓	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✗	✗	✗	✗	✓	✗	✗	15	50.00%
Public health officer3	✓	✗	✗	✗	✓	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✗	✗	✓	✗	✗	✗	✓	✗	✗	✗	✓	✗	✗	✓	✗	14	46.67%
Entomologist5	✓	✓	✗	✗	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✗	✓	✗	✗	✗	✗	✗	✓	✗	✗	16	53.33%
Entomologist6	✓	✓	✗	✗	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✗	✓	✗	✗	✗	✓	✗	✓	✗	✗	18	60.00%
Public health officer4	✓	✓	✗	✓	✓	✓	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓	✗	✓	✗	✓	✗	✓	✗	✗	✗	✓	✗	✓	✓	✗	19	63.33%
Entomologist7	✓	✓	✗	✗	✓	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✗	✓	✗	✗	✗	✓	✗	✓	✗	✗	17	56.67%
Public health officer5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✗	✓	✗	✗	✗	✓	✗	✓	✓	✗	22	73.33%
Entomologist8	✓	✓	✗	✗	✓	✓	✗	✓	✓	✓	✓	✗	✓	✗	✓	✓	✗	✗	✗	✓	✗	✓	✗	✗	✗	✓	✗	✓	✓	✗	16	53.33%
Entomologist9	✓	✓	✗	✓	✓	✓	✗	✓	✓	✗	✓	✗	✓	✓	✓	✓	✗	✓	✗	✓	✗	✓	✗	✗	✗	✓	✗	✓	✗	✗	17	56.67%
Public health officer6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✗	✓	✗	✗	✗	✓	✗	✓	✓	✗	22	73.33%
Public health officer7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✓	✓	✓	✗	23	76.67%
Entomologist10	✓	✓	✓	✗	✓	✓	✗	✓	✓	✓	✗	✗	✓	✓	✓	✓	✗	✓	✗	✓	✗	✓	✗	✗	✗	✓	✗	✓	✓	✗	18	60.00%
Public health officer8	✓	✓	✗	✓	✗	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✗	✓	✗	✗	✗	✓	✗	✓	✓	✗	19	63.33%
Public health officer9	✓	✗	✗	✓	✓	✓	✗	✓	✓	✓	✗	✓	✓	✓	✓	✗	✗	✓	✗	✓	✗	✓	✗	✗	✗	✓	✗	✓	✓	✗	17	56.67%
Entomologist11	✓	✗	✓	✓	✓	✓	✗	✓	✓	✓	✗	✗	✓	✓	✓	✓	✗	✓	✗	✓	✗	✓	✗	✗	✗	✓	✗	✓	✓	✗	18	60.00%
Deep Learning1	✓	✓	✗	✓	✓	✓	✗	✓	✓	✓	✓	✗	✓	✗	✗	✓	✓	✗	✗	✗	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	20	66.67%
Deep Learning2	✗	✓	✓	✓	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓	✗	✓	✓	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	24	80.00%

Table 3 The mosquito identification scores of 30 mosquito unknown images by 20 public health staffs and the Deep Learning model. (Continue)

Deep Learning	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**Score**	Identification score
Deep Learning3	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	X	✓	X	✓	X	X	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	23	76.67%
Deep Learning4	✓	✓	✓	✓	X	✓	✓	X	✓	✓	✓	X	✓	✓	✓	✓	✓	X	✓	✓	✓	X	✓	✓	✓	✓	X	✓	✓	✓	24	80.00%
Deep Learning5	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	X	✓	X	X	✓	✓	✓	✓	X	✓	X	✓	X	✓	✓	X	✓	✓	✓	21	70.00%
Deep Learning6	X	✓	✓	✓	X	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	25	83.33%
Deep Learning7	✓	✓	X	✓	✓	✓	X	✓	✓	X	✓	✓	✓	X	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	21	70.00%
Deep Learning8	X	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	X	✓	✓	✓	✓	✓	X	✓	X	X	✓	✓	✓	23	76.67%
Deep Learning9	✓	✓	X	✓	✓	✓	X	✓	✓	X	✓	X	✓	X	X	X	X	X	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	20	66.67%
Deep Learning10	X	✓	✓	✓	X	✓	✓	X	✓	X	✓	X	X	✓	X	✓	✓	X	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	X	✓	20	66.67%
Deep Learning11	✓	✓	X	✓	✓	✓	X	✓	X	✓	✓	X	✓	X	X	✓	✓	X	✓	X	✓	✓	✓	X	✓	✓	X	✓	✓	✓	20	66.67%
Deep Learning12	X	X	X	X	X	✓	✓	✓	✓	✓	✓	X	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	23	76.67%
Deep Learning13	✓	✓	X	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	X	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	25	83.33%
Deep Learning14	X	✓	✓	✓	X	✓	✓	✓	X	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	X	24	80.00%
Deep Learning15	✓	✓	X	✓	✓	✓	X	X	✓	X	✓	X	X	X	X	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	21	70.00%
Deep Learning16	X	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	X	✓	✓	X	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	25	83.33%
Deep Learning17	✓	✓	X	✓	✓	✓	X	✓	✓	✓	✓	X	✓	X	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	25	83.33%
Deep Learning18	X	✓	✓	X	X	✓	✓	✓	X	✓	✓	✓	✓	✓	X	✓	✓	X	✓	✓	X	✓	✓	✓	X	✓	✓	✓	✓	✓	22	73.33%
Deep Learning19	✓	✓	X	✓	✓	✓	X	✓	✓	✓	✓	X	✓	X	X	✓	✓	X	X	X	✓	✓	✓	X	✓	✓	✓	✓	X	✓	20	66.67%
Deep Learning20	X	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	X	✓	✓	X	✓	X	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	24	**80.00%**

DISCUSSION

Data set training and optimization

This study was to assemble a comprehensive dataset of medically significant mosquito species found in Thailand among *Ae. albopictus, Ae. aegypti, An. minimus, An. harrisoni, An. dirus, An. maculatus*, and *Cx. quinquefasciatus*. This dataset served for training a computer-based system capable of more accurately identifying mosquitoes. Additionally, we targeted to develop a user-friendly application tailored to professionals engaged in mosquito surveillance, thereby enhancing the precision and efficiency of mosquito identification processes. Notably, AI systems typically face difficulties in distinguishing between closely related species that are easily discernible to the human eye (Lorenz *et al*, 2018, Kittichai *et al*, 2021). Our research was further to address these challenges by optimizing and exploring the relationship between dataset size and accuracy in mosquito identification. We observed a significant positive correlation between dataset size and the accuracy of AI models. As the dataset's specimen count increased, the model's proficiency in identifying mosquito species improved—a principle aligned with machine learning fundamentals wherein larger and more diverse datasets enhance model performance. Moreover, the precision score, a key metric for evaluating model performance, showed a positive correlation with dataset size. A precision score nearing 1 indicated heightened accuracy, implying fewer false positive identifications. This finding highlights the importance of dataset size in achieving precise and accurate mosquito species classification. It is consistent to a study on deep learning-based organ auto-segmentation for head-and-neck patients, which observed improved accuracy with increasing dataset size (Orlando *et al*, 2021). Furthermore, the optimization of datasets with ample image quantities provided compelling evidence of the positive relationship between dataset size and accuracy in mosquito species identification. This emphasizes the importance of both data quality and quantity in training AI models for species identification.

Concerning image resolution and complexity, our evaluation revealed their impact on the accuracy of AI models for mosquito species identification. Two groups of datasets were created: one comprising high-resolution plain mosquito pictures and the other featuring a blend of high and low-resolution images. Contrary to expectations, the group containing low-resolution images exhibited superior performance, challenging the conventional belief that models trained solely on high-resolution images yield greater accuracy. Recent research was found no statistically significant difference between datasets containing high and low-quality images (Fang *et al*, 2021). In real-world situations, mobile devices often capture mosquito images, which may lack the high resolution of laboratory images. The unexpected efficacy of the AI model on low-resolution images underscores the importance of adaptability. This revelation has practical implications for fieldwork, reducing reliance on high-end equipment and enhancing accessibility to mosquito surveillance tools. Specifically, designed tools for microscopes or mobile devices can assist entomologists in real-world scenarios, alleviating workload burdens and augmenting accuracy.

Deep learning approach to entomology work

The traditional linear logic of AI models for mosquito species identification were demonstrated here as well as the explored artificial intelligence techniques for classifying living organisms, emphasizing the gradual evolution of artificial neural network systems as practical tools (Bartoń & Barton, 2019). Unlike human cognition, which relies on experience, theory, and knowledge, AI operates on logic and probabilities. This prompted an exploration into infusing AI with more human-like logic, particularly that of entomologists who use standardized morphological pictorial keys. Existing AI models, appearing as "linear robots," raised concerns about their efficiency in emulating entomologists' decision-making processes. Heat maps of Table 3 showed that AI models often relied on image features rather than the fundamental characteristics used by human experts for identification. Conducting multilayer deep learning, we aimed to train the computers analysis information in a manner similar to human cognitive processes that was inspired by a successful study on spotting diseases in potatoes (Rashid *et al*, 2021), achieving over 80% accuracy. The comparison of the accurate species identification between expert entomologists and an Artificial Intelligence (AI) identification platform was achieved an accuracy range of 67% to 87%. This experiment underscores the persistent challenge posed by species complexes. Although experts can identify the subtle differences between closely related species, field officers might overlook these nuanced distinctions. Such misidentifications can significantly impact vector surveillance efforts, leading to incomplete investigations and suboptimal disease management strategies. Similar challenges related to species complexes have also been observed in studies involving Anopheles mosquitoes. One particular study focused on analyzing the wing vein patterns of the Gambiae complex, which includes *An. gambiae, An. arabiensis*, and *An. coluzzii* (Cannet *et al*, 2023). Moreover, the study results have demonstrated that discerning between these closely related species remains highly challenge for future research. The cryptic patterns of their wing veins create problem to species identified, emphasizing the requirement for additional research in this topic. In practical scenarios, confirmation continues to rely on molecular techniques, running parallel to the development of innovative methods for collaborative advancement. The capacity of artificial intelligence to classify insects is still restricted; unlike the work of experienced entomologists, it cannot perform effectively when presented with unclear specimen. However, artificial intelligence has one benefit over human memory and actually accurate.

In this experiment, it was pointed out that the classification of mosquitoes as the main disease vectors in Thailand using artificial intelligence. It can be done and even working at the complex species level can produce positive results. However, hardware limitations for processing are still the main problem in local application. In the future, if computing hardware is invented with good performance and a price that is accessible to officials or the general public, this research will help reduce the workload of entomologists in the future.

CONCLUSIONS

Our study revealed no significant difference in performance between the Deep Detection AI model and human participants. Surprisingly, the Deep Classification AI model exhibited significantly better performance than humans in identifying mosquito species complexes. These findings suggest that the Deep Classification AI model holds promise for enhancing the accuracy of mosquito identification, while the Deep Detection AI model may not offer substantial advantages over human identification.

ACKNOWLEDGEMENTS

This study was supported by grants (U19AI089672 and D43TW006571) from the National Institutes of Health, USA. We thank the field data collection team and entomology team from Mahidol University and the staffs of the Department of Public Health in Tha Song Yang District, Tak Province for their dedication to collect samples and follow up the study participants. We would like to acknowledge with mosquito collection of *An. harrisoni* by Mr. Suntorn Pimnon from Department of Entomology Kasetsart University.

REFERENCES

Cannet A, Simon-Chane C, Akhoundi M, *et al*, Deep learning and wing interferential patterns identify Anopheles species and discriminate amongst Gambiae complex species. *Sci Rep*. 2023 Aug 25;13(1):13895.

CIRA CORE | AI Platform by CIRA Robotics. (n.d.). https://www.cira-ai.com/

Fang Y, Wang J, Ou X, Ying H, Hu C, Zhang Z, Hu W. The impact of training sample size on deep learning-based organ auto-segmentation for head-and-neck patients. *Phys Med Biol*. 2021:14;66(18): 34450599.

Goodwin A, Padmanabhan S, Hira S, Glancey M, Slinowsky M, Immidisetti R. Mosquito species identification using convolutional neural networks with a multitiered ensemble model for novel species detection. *Scientific Reports*. 2021;11(1):13656.

Huang L, Hong M, Luo C, Mahajan S, and Chen L. A Vector Mosquitoes Classification System Based on Edge Computing and Deep Learning. Conference on Technologies and Applications of Artificial Intelligence (TAAI), Taichung, 2018.

Jame Rohlf F, Bookstein F. Proceedings of the Michigan morphometrics workshop. 1990:pp. 61-74.

Joshi A, Miller C. Review of machine learning techniques for mosquito control in urban environments. *Ecological Informatics*. 2021;61:101241.

Kiskin I, Orozco BP, Windebank T, *et al*, Mosquito detection with neural networks: the buzz of deep learning. *arXiv preprint* 2017:1705.05180.

Kittichai V, Pengsakul T, Chumchuen K, *et al*, Deep learning approaches for challenging species and gender identification of mosquito vectors. *Sci Rep*. 2021 Mar 1;11(1):4838.

Li Z, Shen Z, Yao Q. Automated Identification of Mosquito (Diptera: Culicidae) Wingbeat Waveform by Artificial Neural Network. 2005.

Lorenz C, Ferraudo AS, Suesdek L. Artificial Neural Network applied as a methodology of mosquito species identification. *Acta Trop*. 2015:152;165-169.

Minakshi M, Bharti P, Bhuiyan T, Kariev S, and Chellappa, S. A framework based on deep neural networks to extract anatomy of mosquitoes from images. Scientific Reports. 2020:10(1).

Motta D, Santos AÁB, Winkler I, *et al*, Application of convolutional neural networks for classification of adult mosquitoes in the field. *PLoS One*. 2019 Jan 14;14(1):e0210829.

Mulchandani P, Siddiqui M, Kanani P. Real-Time Mosquito Species Identification using Deep Learning Techniques. *International Journal of Engineering and Advanced Technology*. 2019: 9(2);2000–2003.

Orlando N, Gyacskov I, Gillies DJ, *et al*, Effect of dataset size, image quality, and image type on deep learning-based automatic prostate segmentation in 3D ultrasound. *Phys Med Biol*. 2022 Mar 29;67(7). doi: 10.1088/1361-6560/ac5a93. PMID: 35240585.

Rattanarithikul R, Harrison BA, Harbach RE, Panthusiri P, Coleman RE, Panthusiri P. Illustrated keys to the mosquitoes of Thailand. IV. Anopheles. *Southeast Asian J Trop Med Public Health*. 2006;37: 17262930.

Rashid J, Khan I, Ali G, Almotiri SH, AlGhamdi MA, Masood K. Multi-Level Deep Learning Model for Potato Leaf Disease Recognition. *Electronics*. 2021;10(17):2064.

Taai K, Harbach R E, Aupalee K, *et al*. An effective method for the identification and separation of *Anopheles minimus*, the primary malaria vector in Thailand, and its sister species *Anopheles harrisoni*, with a comparison of their mating behaviors. *Parasites Vectors*. 2017:10;97.

Tai-Hsien EY, Joeair J, Tate L. Mosquito vector monitoring system based on optical wingbeat classification. *Computers and Electronics in Agriculture*. 2015.

World Health Organization. World Malaria Report 2018 ISBN 978 92 4 156565 3, Available from: www.who.int/malaria, accessed on [20 December 2023].