Image Anal Stereol 2022;41:203-215
Original Research Paper
doi: 10.5566/ias.2804

DEEP LEARNING-BASED VECTOR MOSQUITOES CLASSIFICATION FOR PREVENTING INFECTIOUS DISEASES TRANSMISSION

MISAGH ASGARI[1], AREZOO SADEGHZADEH[1], MD BAHARUL ISLAM[✉,1,2,3], LAVDIE RADA[4] AND JAMES BOZEMAN[2]

[1]Department of Computer Engineering, Bahcesehir University, Yildiz, Ciragan Cd, Besiktas, 34349, Istanbul, Turkey, [2]College of Data Science and Engineering, American University of Malta, Triq Dom Mintoff, BML 1013, Malta, [3]Department of Computer Science & Engineering, Daffodil International University, DIU Road, Dhaka 1341, Bangladesh, [4]Department of Biomedical Engineering, Bahcesehir University, Yildiz, Ciragan Cd, Besiktas, 34349, Istanbul, Turkey
e-mail: misagh.asgari@bahcesehir.edu.tr, sadegzadeh.a@gmail.com, bislam.eng@gmail.com,lavdie.rada@eng.bau.edu.tr,james.bozeman@aum.edu.mt
(Received October 3, 2022; accepted November 15, 2022)

ABSTRACT

Healthcare systems worldwide are burdened by mosquitoes transmitting dangerous diseases. Conventional mosquito surveillance methods to alleviate these diseases are based on expert entomologists' manual examination of the morphological characteristics, which is time-consuming and unscalable. The lack of professional experts brings a high necessity for cheap and accurate automated alternatives for mosquito classification. This paper proposes an end-to-end deep Convolutional Neural Network (CNN) for mosquito species classification by taking advantage of both dropout layers and transfer learning to enhance performance accuracy. Dropout layers randomly disable the neurons of the neural network, mitigating co-adaptation and data overfitting. Transfer learning efficiently applies the extracted features from one dataset to others. Furthermore, a Region of Interest (ROI) visualization component is adopted to gain insight into the model learning. The generalization ability and feasibility of the proposed model are validated on four publicly available mosquito datasets. Experimental results on these datasets with an accuracy of 98.82%, 98.92%, 94.66%, and 98.40% demonstrate the superiority of our proposed system over the recent state-of-the-art approaches. The effectiveness of different numbers of dropout layers, their positions in the network, and their values are all investigated through ablation studies. Visualizing the model attention confirms that useful mosquito features are learned from insect legs and thorax through our model leading to optimistic predictions.

Keywords: Convolutional neural networks, mosquito classification, transfer learning, vector mosquitoes.

INTRODUCTION

Mosquitoes are tiny creatures with more than 3600 species, among which a few dozen are vectors of deadly diseases such as malaria, dengue, yellow fever, and chikungunya. These diseases cause over one billion infections and around one million deaths worldwide annually Organization *et al.* (2017); Omodior *et al.* (2018), making mosquitoes the deadliest animals in the world Gates (2014). *Aedes*, *Anopheles*, and *Culex* are considered as the most dangerous genera Roth *et al.* (2014); Kittichai *et al.* (2021a). These mosquitoes are found in almost every region of the world, and their female gender transmits the diseases by injecting infected saliva into human hosts. To prevent and minimize the distribution of mosquito-borne diseases and assist health authorities, it is beneficial to identify and classify the disease-spreading mosquitoes and monitor their populations. Vector control programs are fundamentally carried out using manual microscopic observations through which the insects are identified by morphological and dichotomous keys Rueda (2004); Park *et al.* (2016); Eritja *et al.* (2019). However, these conventional methods are required by a highly professional entomologist leading to a time-consuming, laborious, barely scalable, and costly process that is infeasible for practical implementations. Recently, it has become even more challenging due to the lack of experts compared to the extreme increment in mosquito diversity and populations Audisio (2017). Furthermore, the external morphological characteristics are susceptible to damage during sample acquisition, preservation, or transportation so mosquito identification is challenging even for professionals Mewara *et al.* (2018).

Molecular-based methods such as polymerase chain reaction Clapp (1996), ELISA, and DNA-barcoding Wang *et al.* (2012); Beebe (2018); Mee *et al.* (2021) are another practical alternative for identifying and classifying mosquitoes but are also impractical in real-world applications as they follow slow procedures, require expensive technical equipment, and are performed by molecular biology experts Kittichai *et al.* (2021a). These limitations inspired researchers to develop automated systems to classify mosquito species. The prior frequency-based automated mosquito surveillance systems analyzed wingbeat harmonics using acoustic recorders Jackson and Robert (2006); Silva *et al.* (2013); Arthur *et al.* (2014); Ouyang *et al.* (2015a); Mukundarajan *et al.* (2017). Despite their high classification accuracy, these devices suffered from limited storage memory, short-distance operational ranges, and dependence of the data acquisition procedures to the position of the recorder concerning the mosquitoes.

Consequently, vision-based methods have been extensively applied to deal with these challenges in the last five years. These approaches are generally divided into two main categories: 1) conventional machine learning (ML)-based approaches and 2) deep learning (DL)-based approaches. In conventional approaches, the handcrafted features are first extracted, then fed into the ML-based classifiers for mosquito identification Ouyang *et al.* (2015b); Reyes *et al.* (2016). Although they achieved successful performance, their accuracy is not still satisfactory for real-world applications. Furthermore, they have low generalization ability as the extracted features, so the model's performance is significantly affected when the dataset is changed. Thanks to recent advances in developing powerful GPUs and large-scale datasets, DL-based models such as convolutional neural networks (CNNs) have been extensively applied to numerous computer vision tasks, including mosquito identification and classification. However, they also suffer from some limitations restricting their applicability in real-world scenarios: i) they obtained satisfactory performance for the images acquired in a laboratory under environmental constraints Park *et al.* (2020); Goodwin *et al.* (2021), ii) even for those datasets, their accuracy still is not in the human expert level Motta *et al.* (2019), and iii) they have poor generalization capability as they are only validated on a limited number of datasets (mostly a single dataset) Park *et al.* (2020); Rustam *et al.* (2022).

To address these issues, a novel deep CNN (DCNN) model is proposed to simultaneously improve the generalization ability and the accuracy of mosquito classification benefiting from the strengths of both regularization layers and transfer learning. The proposed model can learn and extract the fine-grained features from the discriminant parts of the mosquitoes, similar to those used by entomologists in manual examinations. Visualization of these features based on the Grad-CAM Selvaraju *et al.* (2016) further proves the capabilities and effectiveness of the proposed mosquito classification system. Overall, the key contributions of the paper are summarized as follows:

- A novel end-to-end deep neural network is proposed for mosquito classification based on modifying the VGG16 architecture and applying transfer learning on the ImageNet Deng *et al.* (2009) dataset for taking advantage of features extracted from non-mosquito images. Consequently, the model can perform accurately for both small-scale and large-scale datasets.
- Inspired by the regularization technique, the original architecture of the pre-trained VGG16 is modified by adding two dropout layers which effectively increase the classification accuracy while mitigating model overfitting.
- Proper location and the optimal number and value of the dropout layers are selected through extensive ablation studies resulting in high classification accuracy. In addition to the promising quantitative results, the feasibility of the model is evaluated qualitatively through a visualization model based on the Grad-CAM algorithm.
- Assessing the performance of the proposed model on four different public datasets along with the combination of them, it is demonstrated that our modified model outperforms four pre-trained models under both controlled and uncontrolled environments with small inter-class and large intra-class variations leading to a high level of generalization.

The remainder of the paper is organized as follows. Recent related works are briefly reviewed. Then, the proposed model is presented in detail. The model's performance is evaluated and compared with the recent approaches through experiments besides the explanations regarding the employed datasets, experimental setup, and evaluation metrics. Ablation studies and failure cases for the proposed model are respectively discussed. Finally, we conclude the works with possible future research directions.

RELATED WORK

Due to the importance of automated mosquito identification and classification in monitoring the population of vector mosquitoes and controlling mosquito-borne diseases, researchers have developed numerous approaches using both audio and visual features. Among these studies, the recent competitive vision-based approaches using deep learning models

Table 1: Summary of the recent deep learning-based mosquito classification approaches, including brief details about their evaluated dataset, methodology, and the reported accuracy (ACC) (MA: Mosquito Alert).

Reference	Dataset	Availability	Methodology	ACC	Classes
Motta *et al.* (2019)	Own dataset (4,056 images)	×	GoogleNet	76.2%	6
Okayasu *et al.* (2019)	Own dataset (14,400 images)	×	ResNet50 + Aug.	95.5%	3
Park *et al.* (2020)	Own dataset (3,600 images)	✓	VGG16 + Aug. + Transfer learning	97.2%	5
Kittichai *et al.* (2021b)	Own dataset (10,564 images)	Upon request	YOLOv3 + Aug.	98.9%	15
Adhane *et al.* (2021)	MA multidisciplinary team (2017) (3,364 images)	✓	VGG-16	94.6%	2
Akter *et al.* (2021)	Web sources (3,600 images)	×	Custom CNN	93.0%	3
Goodwin *et al.* (2021)	Own dataset (2,696 images)	✓	Xception	97.0%	39
Rustam *et al.* (2022)	Pise *et al.* (2020) (1,404 images)	✓	VGG16 using RIFS	98.6%	2

are briefly presented in this section (as summarized in Table 1). Mota et al. Motta *et al.* (2019) applied the pre-trained GoogleNet, LeNet, and AlexNet models to classify both male and female genders of adult *Aedes aegypti*, *Aedes albopictus*, and *Culex quinquefasciatus* mosquitoes into 6 total classes. The highest testing accuracy of 76.2% was achieved by GoogleNet. The experiments were carried out on their own dataset with the images captured from dead mosquitoes in a laboratory under almost uniform lighting and environmental conditions. In Okayasu *et al.* (2019), Okayasu et al. compared the performance of the conventional ML techniques with that of DL-based models for mosquito classification. They extracted 12 different hand-crafted features from the images of their dataset and used them to train a Support Vector Machine (SVM) classifier. Their dataset was formed by the images of the living mosquitoes captured on plain white backgrounds. The performance of the SVM was compared with three DL models of AlexNet, VGG16, and ResNet, demonstrating that ResNet obtained the highest classification accuracy of 95.5% on the augmented dataset.

Park et al. Park *et al.* (2020) developed a new dataset from the images of *Aedes*, *Anopheles*, and *Culex* mosquitoes which were classified after augmentation based on applying transfer learning on VGG16, Resnet50, and SqueezeNet models. Considering three experimental schemes of employing no augmentation or fine-tuning, only fine-tuning, and both fine-tuning and augmentation, VGG16 outperformed the other models with an accuracy of 56.7%, 91.1%, and 97.2%, respectively. Kittichai et al. Kittichai *et al.* (2021b) developed another mosquito dataset with 15 different classes from both newborn and adult mosquitoes in two genders, male and female. They applied different YOLO-based models Redmon *et al.* (2015) for real-time classification, among which YOLO-v3 achieved the highest accuracy of 98.9% after enriching the dataset with augmentation.

Adhane et al. Adhane *et al.* (2021) presented a comparative study based on two DL-based models, i.e., VGG16 and ResNet50, for binary classification of tiger mosquitoes on the Mosquito Alert dataset. Applying the transfer learning on these pre-trained models, VGG16 surpassed ResNet50 with a validation accuracy of 94.6%. Furthermore, the regions used by the model to learn the features were visualized based on the explainable models. Although it was demonstrated that their model effectively learned features from discriminative morphological patterns of mosquitoes, the final accuracy is still insufficient for real-world scenarios. Moreover, it suffers from poor generalization as the performance was investigated only on a single dataset and for a binary classification task, i.e., tiger vs. non-tiger mosquitoes. At the same time, there are other genera of vector mosquitoes whose identification plays a significant role in efficiently and practically controlling mosquito-borne diseases. Akter et al. Akter *et al.* (2021) collected their dataset from different web sources. Their dataset was formed by 442 images which were increased to 3600 images by applying four types of augmentation. Proposing a custom CNN model with convolutional, pooling, and dropout layers, they achieved a classification accuracy of 70%, which was improved to 93% after augmentation outperforming the other models of VGG16, Random Forest, XGboost, and SVM. Goodwin et al. Goodwin *et al.* (2021) also developed a publicly available dataset with 2696 images in 39 classes and investigated the performance of the Xception on it, reporting an accuracy of 97%.

Recently, Rustam et al. Rustam *et al.* (2022) proposed a new feature selection method as RIFS, i.e., a combination of ROI- and wrapper-based feature selection methods, for binary classification of Aedes and Culex mosquitoes. Different ML- and DL-based classification models were applied, among which Extra Tree Classifier (ETC) (99.2% accuracy) and VGG16 (98.6% accuracy) achieved the best performance while the computational time and cost were reduced. Their implementations were all carried out on a single dataset developed by Pise et al. Pise *et al.* (2020), including mosquito images on various backgrounds. Although the existing approaches obtained satisfactory performance for mosquito classification, they still need to be improved to the expert-level performance. On the other hand, validating the performance mostly only on a single dataset challenges their generalization capability. Considering these issues, developing a high-quality mosquito classification system on various datasets captured in controlled and uncontrolled environments is highly demanded, providing effective preventive strategies and controlling the spread of the arboviruses.

PROPOSED METHOD

The main flowchart of the proposed mosquito classification system is demonstrated in Fig. 1. It comprises three main components: 1) feature extraction based on pre-trained VGG16, 2) classification module with fully-connected, dropout, and softmax layers, and 3) explainable model based on Grad-CAM for visualization. Each of these components and the applied techniques in the training process are explained in detail in the following subsections.

VGG16-BASED FEATURE EXTRACTOR BACKBONE

Before extracting the features from the input mosquito images, all the RGB images are resized into 224×224 pixels whose intensities are normalized, leading to the mean and standard deviation values of 0 and 1, respectively. These pre-processed images are fed into the pre-trained VGG16 model Simonyan and Zisserman (2014) to extract the rich discriminative features by minimizing the cross-entropy loss function. VGG16 is one of the successful vision model architectures. It extracts the feature maps from the input images by a total of 13 convolutional and five max-pooling layers arranged in 5 blocks. Two first blocks have similar structures formed by two convolutional layers with 3×3 kernel size and stride of 1, followed by a max-pooling layer with 2×2 pooling size and stride of 2. The only difference between these two blocks is the number of filters in the convolutional layers assigned as 64 and 128, respectively. In the last three blocks, the convolutional layers are increased to three with 256, 512, and 512 filters from the third to fifth blocks, respectively. Max-pooling layers are the same as the previous blocks. The activation function of all convolutional layers is the Rectified Linear Unit (ReLU) Agarap (2018). It introduces non-linearity to classification, makes the learning process faster, enhances the performance, and deals with the vanishing gradient issue while it has simple computation. Based on its main equation defined as $ReLU(x) = max(0, x)$, it returns back its input value if it is positive. Otherwise, its output is set to zero. The final extracted feature maps from the VGG16 architecture have a dimension of $7 \times 7 \times 512$. As training schemes of the CNN models are based on the feed-forward process, the shallower convolutional layers learn general patterns such as edges, corners, boundaries, etc. In contrast, the deeper ones learn more extensive patterns leading to the feature maps desired for a specific task Mayer *et al.* (2018). VGG16 has been initially pre-trained on ImageNet, a large dataset with over one billion images. Taking advantage of transfer learning, this pre-trained model and the extracted features are fine-tuned for mosquito classification. The main benefits of applying transfer learning instead of training the model from scratch are its faster and easier convergence, rich representations of the features, and obtaining high accuracy, even for small-scale datasets.

MODIFIED CLASSIFICATION BLOCK

Once the feature maps are extracted from the last block of the VGG16 model, they are flattened into a linear vector fed into the classification block for predicting the type of mosquito. Generally, the VGG16 model has approximately 138 million parameters, 90% of which are reserved for its classification block, emphasizing its importance. The original pre-trained VGG16 model has two fully-connected layers (with 4096 neurons) followed by a ReLU activation function and a Softmax layer. To further mitigate the effects of data scarcity, prevent the model from overfitting, and enhance the model's generalization, a dropout layer is added after each fully-connected layer as a regularization technique. Overfitting occurs when the trained model is too complex and performs poorly when confronted with new data. To address this issue, dropout layers with the value of 0.5 in our modified classification block randomly ignore 50% of the neurons and remove their contribution on forward pass during training Baldi and Sadowski (2013). Consequently, their weights are not updated on the backward pass, and the remaining neurons provide the desired representation from the input for the final prediction. In this case, the network learns multiple independent representations, leading to better generalization and less probability of overfitting.

The final layer is another fully connected layer with *Softmax* activation function whose number of neurons equals the number of classes. *Softmax* activation function is mathematically defined as follows:

$$Softmax(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \tag{1}$$

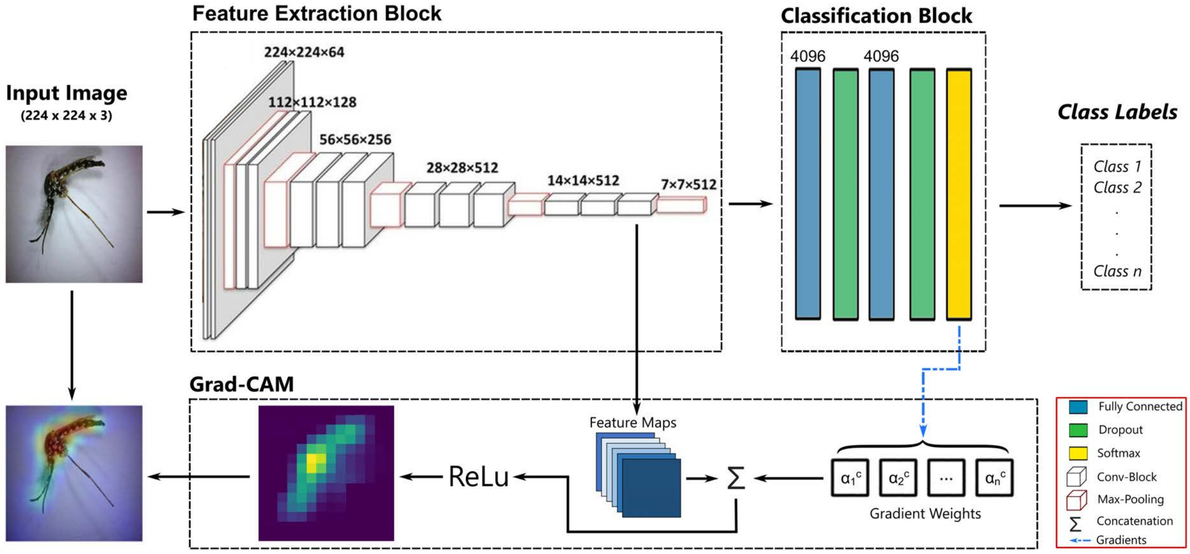

Fig. 1: *The main flowchart of the proposed model includes the convolutional blocks as feature extractors, fully connected and softmax layers as classifiers, and two dropout layers as the regularization techniques. The Grad-CAM component shows the model's focus on learning the discriminant features.*

Given a vector of numbers, i.e., x_i, *Softmax* converts it into a vector of normalized values as probabilities that sum to one. Each of these values in the output of the *Softmax* represents the probability score of membership for each class. The class with the highest probability score is considered the final prediction for the model. Considering mosquito recognition as a multi-class classification task, a categorical cross-entropy loss function is employed for training our model:

$$Loss = -\sum_{i=1}^{output\ size} y_i \,.\, log\ \hat{y}_i \tag{2}$$

where $\hat{y}_i$ and y_i are the predicted scalar probability and the corresponding target value.

GRAD-CAM-BASED EXPLAINABLE MODEL

Grad-CAM stands for Gradient-based Class Activation Maps, is a visualization algorithm that uses gradients of the predicted class (calculated by *Softmax*) as α_i concerning the extracted feature maps from the final convolution layer to generate heatmaps depicting ROI where the model focuses on it for final prediction Selvaraju *et al.* (2016). To this end, as illustrated in Fig. 1, for a given resized image of x with the height and width of H and W, the weighted sum of the alpha values is calculated as follows:

$$\alpha_k^c(i,j) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \frac{\partial \hat{y}^c}{\partial A_k(i,j)} \tag{3}$$

where $A_k(i,j)$ is the k-th activation unit in the last convolutional layer. The importance of different regions for the given class c is visualized by $\alpha_k^c(i,j)$. These regions are highlighted in the heatmap by applying the ReLU activation function as $Grad-CAM = ReLU(\sum_k \alpha_k^c A_k)$. Practically, these maps help to qualitatively assess a model to know if it effectively learns the important features from the morphological characteristics of the mosquitoes or if it just makes predictions from unrelated features of the image.

EXPERIMENTAL RESULTS

DATASETS

The proposed mosquito classification model is evaluated on four publicly available datasets captured in both controlled and uncontrolled environments with variations in the background and illumination conditions (shown in Fig. 2). The details of each dataset regarding the year of publication, number of classes and images, labels of the classes, and more information about the environmental conditions are all summarized in Table 2.

Park

This dataset was developed by Park et al. Park *et al.* (2020) in 2020, including a total of 4290 images from vector and non-vector mosquitoes. Vector mosquitoes have 5 sub-classes of *Aedes albopictus*, *Aedes vexans*, *Anopheles sinensis*, *Culex pipiens*, and

Table 2: Summarized details of four different mosquito datasets adopted to evaluate the performance and generalization of our proposed model.

Dataset	Classes	Size	Genus (Species)	Explanation
Park Park *et al.* (2020)	6	4,290	*Aedes* (2) *Anopheles* (1) *Culex* (2) Non-vectors	Dead mosquitoes Plain backgrounds The Non-vector class includes 5 different species
IEEE Pise *et al.* (2020)	2	1,404	*Aedes* *Culex*	Living mosquitoes Various backgrounds Some images were created by augmentation
Kaggle Isawasan (2020)	2	1,023	*Aedes* (2)	Dead mosquitoes Plain blue background
Goodwin Goodwin *et al.* (2021)	18	14,400	*Aedes* (10) *Anopheles* (4) *Coquilletiddia* (1) *Culex* (3)	Dead mosquitoes Plain backgrounds Only classes with over 100 images are adopted for our experiments Dataset is balanced through sampling or augmentation so that each class ends up having 800 images

Culex tritaeniorhynchus. All the images were captured from dead adult mosquitoes in the lab environment with a plain gray background and under stable lighting conditions. In our experiments, the images of all classes are utilized without enriching the quantity of the dataset by augmentation.

Fig. 2: *Sample images from four different mosquito datasets used to evaluate the performance and generalization of our proposed model.*

IEEE

This publicly available dataset was developed by Pise et al. Pise *et al.* (2020) for binary classification of the mosquitoes into two general classes of *Aedes* and *Culex* mosquitoes with overall 1404 images. All the images were captured in nature from alive mosquitoes, some of whom were fed with blood just before capturing the images leading to red stomachs. The number of images was enhanced by applying rotation as augmentation.

Kaggle

Isawasan et al. Isawasan (2020) provided this dataset for classifying two types of *Aedes* mosquitoes namely *Aedes albopictus* and *Aedes Anopheles*. The images were captured from dead mosquitoes in a laboratory environment with plain blue background. All images of both classes are employed in our experiments for binary classification.

Goodwin

This dataset was provided by Goodwin et al. Goodwin *et al.* (2021) with overall 6548 images in 67 classes. This is the first mosquito dataset with the highest number of classes and many varieties of mosquito species. However, the data distribution among the classes is highly imbalanced. To reach a good balance in this dataset, we use the classes with more than 100 images leading to only 18 out of 67 classes. The dataset is balanced by sampling or applying augmentation (i.e., random 0-360 degree rotation, random brightness, hue, contrast, and saturation variations in the ranges of 20%, 10%, 20%, and 20%, respectively) so that each class consists of 800 images resulting in overall 14400 images.

TRAINING DETAILS

All the experiments are conducted on a PC with windows 10 operating system, Intel Core i7-10700F CPU @ 2.90GHz, an Nvidia GeForce RTX 2080 GPU with 8 GB memory, and the TensorFlow framework with Keras deep learning API. The training process is carried out for 100 epochs, with a batch size of 16, and using an ADAM optimizer with an initial learning rate of $5e^{-6}$. *StepLR* scheduler decays this rate by a factor of 0.25 every 15 epochs. All the input images are resized to 224×224 pixels, and the performance is evaluated by applying the 5-fold cross-validation strategy and computing the average value of the results.

Table 3: Performance evaluation and comparison between our proposed model with four pre-trained models on the test set of four different datasets and their combination.

Model	Dataset	Accuracy (%)	Loss	Precision (%)	Recall (%)	F1-score (%)
MobileNet Howard *et al.* (2017)	Park	87.97	0.322	88	87	87
	IEEE	88.25	0.301	80	94	88
	Kaggle	74.70	0.591	77	73	75
	Goodwin	80.93	0.413	83	80	81
	Combined	79.43	0.482	82	87	85
VGG16 Simonyan and Zisserman (2014)	Park	96.93	0.098	97	97	97
	IEEE	93.14	0.121	94	93	93
	Kaggle	90.40	0.283	90	90	90
	Goodwin	94.33	0.251	93	96	94
	Combined	94.49	0.249	96	94	94
ResNet50 He *et al.* (2016)	Park	97.06	0.079	96	97	97
	IEEE	91.52	0.214	91	92	91
	Kaggle	92.04	0.198	92	92	92
	Goodwin	97.27	0.413	97	96	96
	Combined	95.86	0.231	98	94	95
VGG19 Simonyan and Zisserman (2014)	Park	97.51	0.088	97	97	97
	IEEE	96.55	0.103	97	96	97
	Kaggle	92.11	0.193	93	92	92
	Goodwin	94.94	0.232	96	97	97
	Combined	95.25	0.256	98	94	97
Ours	Park	**98.82**	0.074	99	99	99
	IEEE	**98.92**	0.068	99	99	99
	Kaggle	**94.66**	0.236	95	95	95
	Goodwin	**98.40**	0.082	98	99	98
	Combined	**97.55**	0.086	97	97	97

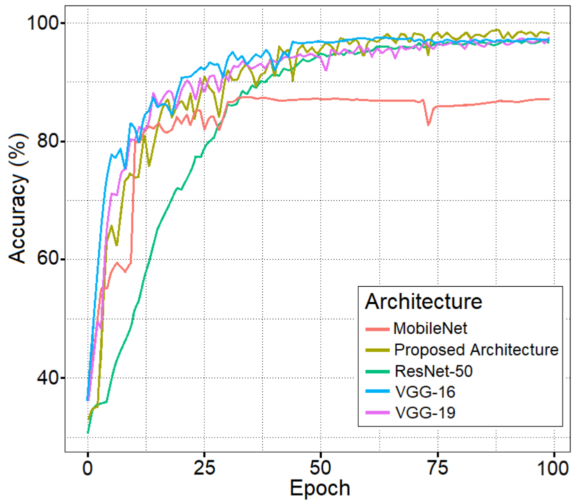

Fig. 3: *Accuracy curve of the proposed model for test set on Park dataset compared to those of four pre-trained models.*

EVALUATION METRICS

One of the critical metrics for evaluating the classification models is accuracy which is defined as $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$ where TP, TN, FP, and FN are respectively the number of true-positive, true-negative, false-positive, and false-negative samples obtained from the confusion matrix. In other words, accuracy is the fraction of right predictions over the total number of predictions. However, employing only accuracy is not enough for effectively evaluating the model. Consequently, three more evaluation metrics of *Precision*, *Recall/Sensitivity*, and *F1-Score* are utilized for assessing the performance of the model. *Precision* is calculated by the fraction of the true positive predictions over all the correct and incorrect positive predictions as $Precision = \frac{TP}{TP+FP}$. In *Recall/Sensitivity*, which is also referred to as True Positive Rate, the number of true positive predictions is divided by all the positive samples as $Recall/Sensitivity = \frac{TP}{FN+TP}$. *F1-Score* is another successful metric defined based on *Precision* and *Recall* as $F1-Score = \frac{2\times Precision\times Recall}{Precision+Recall}$.

PERFORMANCE ASSESSMENT

The performance of our proposed mosquito classification model is evaluated based on four metrics of accuracy, precision, recall, and F1-score on four publicly available datasets and their combination. These values, along with the loss values for the test set, are presented in Table 3 and compared with those of four pre-trained models, i.e., MobileNet Howard *et al.* (2017), VGG16 Simonyan and Zisserman (2014), ResNet50 He *et al.* (2016), and VGG19 Simonyan and Zisserman (2014). The bold values indicate the best accuracy values. Obtaining an accuracy of 98.82%, 98.92%, 94.66%, 98.40%, and 97.55% on the datasets of Park, IEEE, Kaggle, Goodwin, and their combination, demonstrates performance improvement in the ranges of 1.31-10.85%, 2.37-10.67%, 2.55-19.96%, 1.13-17.47%, 1.69-18.12%, respectively. In particular, by modifying the VGG16 architecture by adding two dropout layers, our model surpassed the

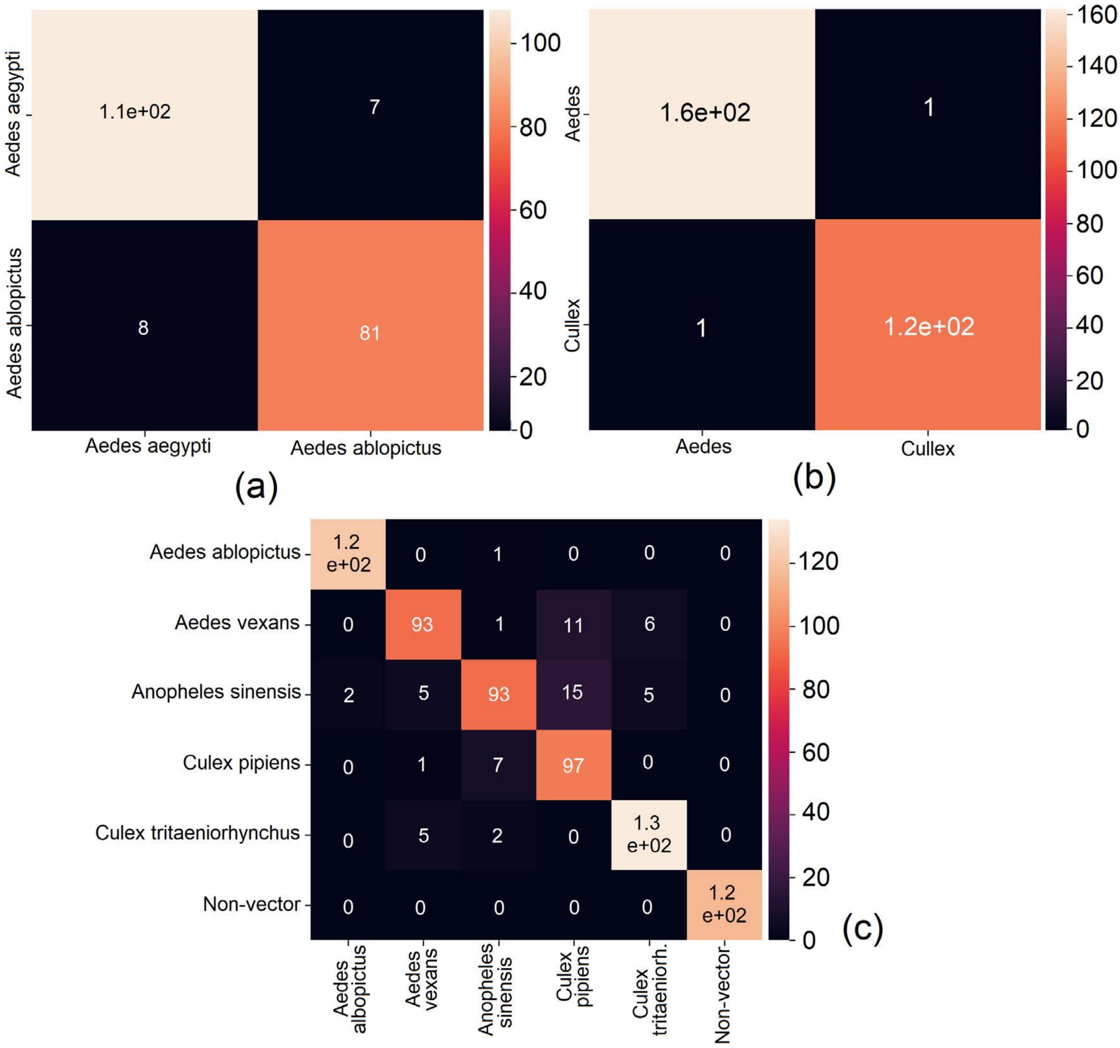

Fig. 4: *Confusion matrices on the test set for three datasets of (a) Kaggle, (b) IEEE, and (c) Park.*

original VGG16 model with an average accuracy enhancement of 3.81% in all datasets, highlighting the effectiveness of the dropout layers as the regularization technique. The superiority of our model is also proved through the loss values and the other three evaluation metrics. Comparing our model's learning curve (accuracy) with the other four pre-trained models on the Park dataset for 100 epochs in Fig. 3, our proposed model is dramatically improved over the MobileNet model after 25 epochs. The other models experienced huge fluctuations up to 40 epochs. Then, they plateau, and our model outperforms the others with at least 1.31% enhancement.

Confusion matrices for the proposed model with dropout layers are depicted in Fig. 4 for the Park, IEEE, and Kaggle datasets and Fig. 5 for the Goodwin dataset to further prove its feasibility. It is worth mentioning that the misclassified samples mainly belong to different species of the same genus whose morphological characteristics bear a striking resemblance. In the Park dataset, the highest number of misclassification occurs between the *Culex pipiens* and two other species of *Aedes vexans* and *Anopheles sinensis*.

In addition to the quantitative evaluation, the efficiency of our model is also evaluated qualitatively through an explainable model based on the Grad-CAM algorithm. As demonstrated in Fig. 6, the heatmaps generated from the final convolutional layer highlight the importance of the mosquitoes' thorax and legs for the model to learn the discriminative features. These regions that are used by the model for mosquito recognition are highly similar to those used by entomologists in the manual examination. These heatmaps give a real insight into the network, confirming its promising performance and capabilities.

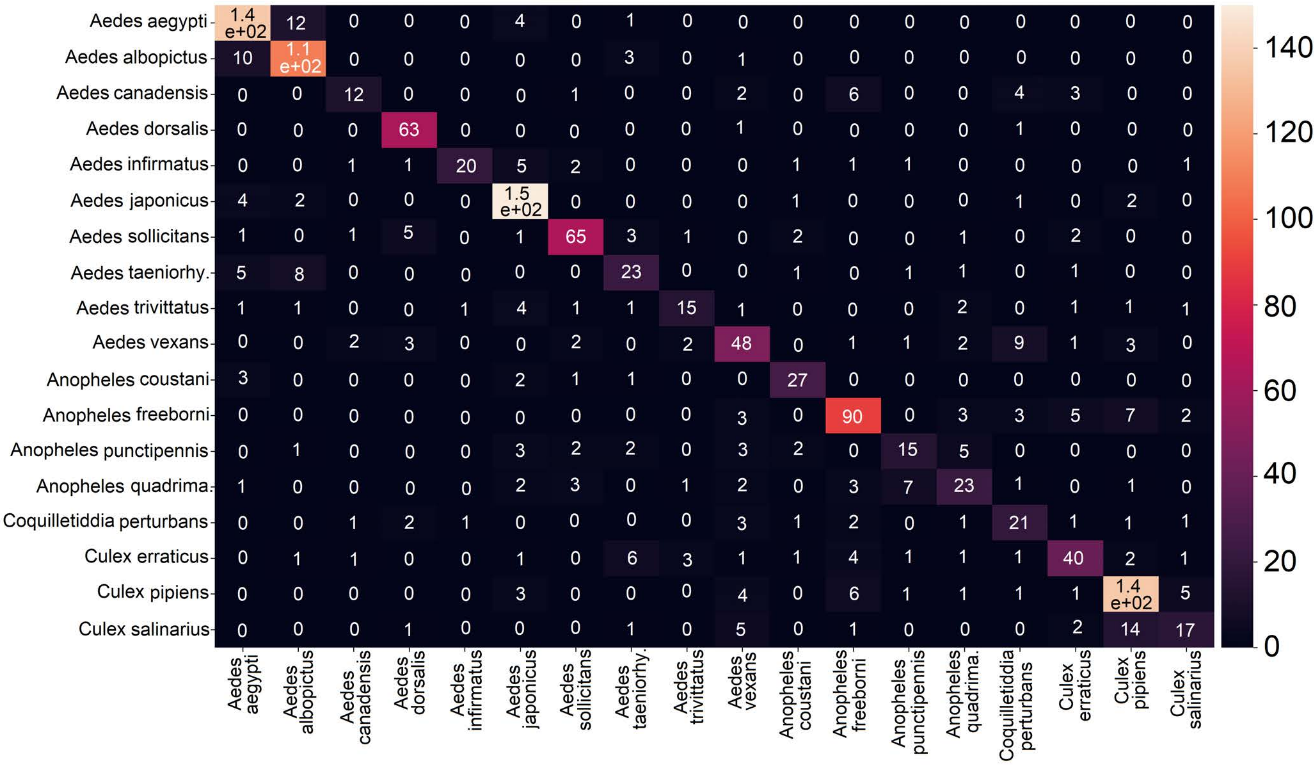

Fig. 5: *Confusion matrix on the test set for Goodwin dataset.*

COMPARISON WITH THE STATE-OF-THE-ART

The performance of the proposed method is compared with the state-of-the-art approaches in Table 4 on four different datasets in terms of classification accuracy. To have a fair comparison, we adopt the same data splits as those of compared methods and use their reported accuracy values except for the model proposed in Adhane *et al.* (2021), whose accuracy is obtained on the same datasets as ours based on their public available source code. Our proposed model outperforms the other approaches with an accuracy of 98.46%, 98.89%, 98.92%, and 94.66% on four datasets of Park, Goodwin, IEEE, and Kaggle, respectively. It is worth mentioning that, although our proposed system is closely related to the architecture in Adhane *et al.* (2021), it obtains superior performance with a significant margin (an increment of 2.05%, 4.07%, 2.26%, 3.33%, respectively) by applying Adam optimizer instead of SGD and adequately adjusting the hyperparameters of the dropout layers and training process. Adam is an extended version of SGD optimizer, and once its hyperparameters (i.e., learning rate and weight decay) are efficiently tuned, it obtains better performance than SGD. In addition to high accuracy, our model has superiority over the other approaches concerning generalization ability. As was mentioned, most existing approaches have poor generalization as they have been evaluated only on a single dataset. For instance, the performance of Adhane *et al.* (2021) was only evaluated on a subset of the Mosquito Alert dataset for binary classification of tiger and non-tiger mosquitoes. In contrast, our model evaluated binary and multi-class (6 and 18 classes) recognition tasks on four different datasets attaining competitive performance. Consequently, high generalization ability, as well as high classification accuracy, make our proposed system applicable to real-world vector control programs.

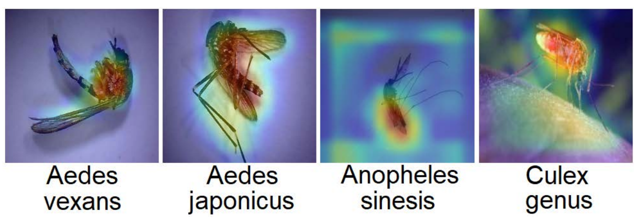

Fig. 6: *Heatmaps generated based on the Grad-CAM algorithm from the last convolutional layer of the model demonstrate the discriminant regions used to learn the features.*

ABLATION STUDY

To investigate the effectiveness of the dropout

Table 4: The performance comparison of the proposed model with the recent approaches on four different datasets.

Dataset	Approach	Classification accuracy (%)
Park	Transfer-learning + Dynamic augmentation Park *et al.* (2020)	97.19
	Transfer-learning + SGD Adhane *et al.* (2021)	96.41
	Proposed Model	**98.46**
Goodwin	Transfer-learning Goodwin *et al.* (2021)	97.04
	Transfer-learning + SGD Adhane *et al.* (2021)	94.82
	Proposed Model	**98.89**
IEEE	Transfer-learning + SGD Adhane *et al.* (2021)	96.66
	RIFS Rustam *et al.* (2022)	98.60
	Proposed Model	**98.92**
Kaggle	Transfer-learning + SGD Adhane *et al.* (2021)	91.33
	Proposed Model	**94.66**

Table 5: The results of ablation study on Goodwin dataset for different numbers of dropout layers and their locations (the bold values indicate two best results).

No. Dropout Layers / Dropout Locations	Accuracy (%)
No dropout layers	93.6
1× dropout layers / conv-block 1	93.3
2× dropout layers / conv-block 1, 2	93.5
3× dropout layers / conv-block 1-3	93.2
4× dropout layers / conv-block 1-4	93.1
5× dropout layers / conv-block 1-5	94.0
1× dropout layer / fc1	95.2
2× dropout layers / fc1 and fc2 (ours)	**98.4**
6× dropout layers / conv-block 1-5 and fc1	96.1
7× dropout layers / conv-block 1-5 and fc1, fc2	**98.5**

layers and their locations in the model architecture, the performance of the proposed model is evaluated by modifying the VGG16 model with different numbers of dropout layers in both feature extraction and classification modules. As presented in Table 5, adding dropout layers in the feature extraction module after each of the first four convolutional blocks degrades the model's performance. Adding one more dropout layer after the fifth convolutional layer improves the performance but still slightly. On the contrary, performance accuracy is significantly improved when the dropout layers are applied after fully connected layers in the classification module. The main reason behind this phenomenon is that almost 90% of the VGG16 parameters are in the classification module, so even small changes in this module can end up with big differences in performance. The performance of the original VGG16 model is enhanced by 1.6% and 4.8% by adding one and two dropout layers after fully connected layers proving the effectiveness of applying the regularization technique. Notably, this performance is slightly improved when dropout layers are simultaneously applied to the feature extraction module (i.e., from 98.4% to 98.5%). Consequently, fewer dropout layers with higher accuracy (i.e., the ones only applied to the classification block) are selected as the optimal case. It should be mentioned that all the dropout layers in this ablation study have the same values as 0.5.

The impact of dropout value as the prominent hyper-parameter of this layer is also investigated in Table 6. Good performance is achieved when the dropout value lies from 0.4 to 0.6. Within this range, the highest accuracy is achieved for 0.5, which is the optimal value for the dropout layers in our proposed model. It is worth mentioning that most of the neurons are ignored during training for weight updating when the dropout value is increased to 0.7 and more. In consequence, the performance dramatically degrades for the large dropout values.

FAILURE CASES

Although our modified architecture obtains successful performance in mosquito recognition, there are also some misclassification cases. The heatmaps of misclassified samples from different datasets are illustrated in Fig. 7. Model attention has carefully investigated for the wrong predictions where the model has failed to focus on the discriminative regions of the mosquitoes. Deeply analyzing these samples, we draw the inferences that three main conditions pose a challenge to our system and degrade its performance: 1) cluttered background, 2) dark shadows of mosquitoes on the background due to improper illumination, and 3) damaged or occluded morphological features such as legs and thorax.

Fig. 7: *Heatmaps generated from the last convolutional layer of the model for the misclassified samples of four datasets (the first and second row depict the original images and the heatmaps, respectively).*

Table 6: The results of ablation study on Goodwin dataset for different values of dropout layer.

Dropout Value	Accuracy (%)
No dropout layers	93.6
0.1	93.1
0.2	93.9
0.3	95.2
0.4	97.4
0.5	**98.4**
0.6	97.9
0.7	88.6
0.8	64.7
0.9	43.2

CONCLUSION

This paper tackled the problem of classifying vector mosquitoes by modifying the VGG16 model with dropout layers and taking advantage of transfer learning. The main focus of this paper was introducing an automated mosquito classification system with high accuracy and generalization ability. To this end, the performance of the proposed model has been evaluated on four publicly available datasets, and their combination proved its capabilities and feasibility. It outperformed the original VGG16 model and three other pre-trained models on all five datasets. An ablation study achieved a good trade-off between the number of dropout layers and the performance accuracy. Our proposed model surpassed the other existing approaches with an accuracy of 98.46%, 98.89%, 98.92%, and 94.66% on Park, Goodwin, IEEE, and Kaggle datasets, respectively. In addition to quantitative evaluation, the model's performance was assessed based on the Grad-CAM algorithm and visualizing the attention of the network for feature extraction. The generated heat maps confirmed that the model learned the data from the discriminative regions of the mosquitoes, which further supported the model's reliability. In future work, we plan to minimize the misclassification between different species of the same genus to improve the accuracy while pruning the model to reduce its computation cost. Providing a new complete dataset, physically capturing the images or artificially generating them based on generative adversarial networks, would greatly benefit the research community in this domain.

ACKNOWLEDGEMENT

This work is partially supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under the 2232 Outstanding Researchers program, Project No. 118C301, and the EU COST Grant No. CA16227.

DATA AVAILABILITY STAMEMENT

Data sharing does not apply to this article as no datasets were generated during the current study. The data supporting this study's findings are publicly available/requested from Park *et al.* (2020); Pise *et al.* (2020); Isawasan (2020); Goodwin *et al.* (2021).

CONFLICT OF INTERESTS STATEMENT

All authors declare that there are no known conflicts of interest.

REFERENCES

Adhane G, Dehshibi MM, Masip D (2021). A deep convolutional neural network for classification of iaedes albopictus/i mosquitoes. IEEE Access 9:72681–90.

Agarap AF (2018). Deep learning using rectified linear units (relu).

Akter M, Hossain MS, Ahmed TU, Andersson K (2021). Mosquito classification using convolutional neural network with data augmentation. In: Advances in Intelligent

Systems and Computing. Springer International Publishing, 865–79.

Arthur BJ, Emr KS, Wyttenbach RA, Hoy RR (2014). Mosquito (aedes aegypti) flight tones: Frequency, harmonicity, spherical spreading, and phase relationships. J Acoust Soc Am 135:933–41.

Audisio P (2017). Insect taxonomy, biodiversity research and the new taxonomic impediments. Fragmenta entomologica 49:121–4.

Baldi P, Sadowski PJ (2013). Understanding dropout. In: Burges C, Bottou L, Welling M, Ghahramani Z, Weinberger K, eds., Advances in Neural Information Processing Systems, vol. 26. Curran Associates, Inc.

Beebe NW (2018). Dna barcoding mosquitoes: advice for potential prospectors. Parasitology 145:622–33.

Clapp JP (1996). Species diagnostics protocols: PCR and other nucleic acid methods. Springer.

Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009). Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. Ieee.

Eritja R, Ruiz-Arrondo I, Delacour-Estrella S, Schaffner F, Álvarez-Chachero J, Bengoa M, Puig MÁ, Melero-Alcíbar R, Oltra A, Bartumeus F (2019). First detection of aedes japonicus in spain: an unexpected finding triggered by citizen science. Parasite Vector 12:1–9.

Gates B (2014). The deadliest animal in the world. Mosquito Week The Gates Notes LLC .

Goodwin A, Padmanabhan S, Hira S, Glancey M, Slinowsky M, Immidisetti R, Scavo L, Brey J, Sudhakar BMMS, Ford T, Heier C, Linton YM, Pecor DB, Caicedo-Quiroga L, Acharya S (2021). Mosquito species identification using convolutional neural networks with a multitiered ensemble model for novel species detection. Sci Rep UK 11.

He K, Zhang X, Ren S, Sun J (2016). Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition.

Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv170404861 .

Isawasan P (2020). Aedes mosquitos dataset.

Jackson JC, Robert D (2006). Nonlinear auditory mechanism enhances female sounds for male mosquitoes. P Natl Acad Sci 103:16734–9.

Kittichai V, Pengsakul T, Chumchuen K, Samung Y, Sriwichai P, Phatthamolrat N, Tongloy T, Jaksukam K, Chuwongin S, Boonsang S (2021a). Deep learning approaches for challenging species and gender identification of mosquito vectors. Sci Rep UK 11:1–14.

Kittichai V, Pengsakul T, Chumchuen K, Samung Y, Sriwichai P, Phatthamolrat N, Tongloy T, Jaksukam K, Chuwongin S, Boonsang S (2021b). Deep learning approaches for challenging species and gender identification of mosquito vectors. Sci Rep UK 11.

Mayer O, Bayar B, Stamm MC (2018). Learning unified deep-features for multiple forensic tasks. In: Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security. ACM.

Mee PT, Wong S, Brown K, Lynch SE (2021). Quantitative PCR assay for the detection of aedes vigilax in mosquito trap collections containing large numbers of morphologically similar species and phylogenetic analysis of specimens collected in victoria, australia. Parasite Vector 14.

Mewara A, Sharma M, Kaura T, Zaman K, Yadav R, Sehgal R (2018). Rapid identification of medically important mosquitoes by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Parasite Vector 11:1–9.

Motta D, Santos AÁB, Winkler I, Machado BAS, Pereira DADI, Cavalcanti AM, Fonseca EOL, Kirchner F, Badaró R (2019). Application of convolutional neural networks for classification of adult mosquitoes in the field. PLOS One 14:e0210829.

Mukundarajan H, Hol FJH, Castillo EA, Newby C, Prakash M (2017). Using mobile phones as acoustic sensors for high-throughput mosquito surveillance. elife 6:e27854.

multidisciplinary team A (2017). Mosquito alert: A citizen platform for studying and which transmit global diseases controlling mosquitos.

Okayasu K, Yoshida K, Fuchida M, Nakamura A (2019). Vision-based classification of mosquito species: Comparison of conventional and deep learning methods. Appl Sci 9:3935.

Omodior O, Luetke MC, Nelson EJ (2018). Mosquito-borne infectious disease, risk-perceptions, and personal protective behavior among us international travelers. Prev Med Rep 12:336–42.

Organization WH, UNICEF, *et al.* (2017). Global vector control response 2017-2030.

Ouyang TH, Yang EC, Jiang JA, Lin TT (2015a). Mosquito vector monitoring system based on optical wingbeat classification. COMPUT ELECTRON AGR 118:47–55.

Ouyang TH, Yang EC, Jiang JA, Lin TT (2015b). Mosquito vector monitoring system based on optical wingbeat classification. Comput Electron Agr 118:47–55.

Park J, Kim DI, Choi B, Kang W, Kwon HW (2020). Classification and morphological analysis of vector mosquitoes using deep convolutional neural networks. Sci Rep UK 10.

Park SI, Bisgin H, Ding H, Semey HG, Langley DA, Tong W, Xu J (2016). Species identification of food contaminating beetles by recognizing patterns in microscopic images of elytra fragments. PLoS One 11:e0157940.

Pise R, Aungmaneeporn M, Patil K, Chumchu P (2020). Image dataset of aedes and culex mosquito species.

Redmon J, Divvala S, Girshick R, Farhadi A (2015). You only look once: Unified, real-time object detection.

Reyes AMMDL, Reyes ACA, Torres JL, Padilla DA, Villaverde J (2016). Detection of aedes aegypti mosquito by digital image processing techniques and support vector machine. In: 2016 IEEE Region 10 Conference (TENCON). IEEE.

Roth A, Mercier A, Lepers C, Hoy D, Duituturaga S, Benyon E, Guillaumot L, Souares Y (2014). Concurrent outbreaks of dengue, chikungunya and zika virus infections–an unprecedented epidemic wave of mosquito-borne viruses in the pacific 2012–2014. Eurosurveillance 19:20929.

Rueda LM (2004). Pictorial keys for the identification of mosquitoes (diptera: Culicidae) associated with dengue virus transmission. Tech. rep., Walter Reed Army Inst Of Research Washington Dc Department Of Entomology.

Rustam F, Reshi AA, Aljedaani W, Alhossan A, Ishaq A, Shafi S, Lee E, Alrabiah Z, Alsuwailem H, Ahmad A, Rupapara V (2022). Vector mosquito image classification using novel RIFS feature selection and machine learning models for disease epidemiology. Saudi J Biol Sci 29:583–94.

Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2016). Grad-cam: Visual explanations from deep networks via gradient-based localization.

Silva DF, De Souza VM, Batista GE, Keogh E, Ellis DP (2013). Applying machine learning and audio analysis techniques to insect recognition in intelligent traps. In: 2013 12th International Conference on Machine Learning and Applications, vol. 1. IEEE.

Simonyan K, Zisserman A (2014). Very deep convolutional networks for large-scale image recognition. arXiv 14091556 .

Wang G, Li C, Guo X, Xing D, Dong Y, Wang Z, Zhang Y, Liu M, Zheng Z, Zhang H, Zhu X, Wu Z, Zhao T (2012). Identifying the main mosquito species in china based on DNA barcoding. PLoS One 7:e47051.