Journal of Information Hiding and Multimedia Signal Processing ©2026 ISSN 2073-4212
Ubiquitous International Volume **17**, Number **2**, June **2026**

Using Open-Set Recognition to Improve Mosquito Identification: A Design Science Approach

Le Nguyen Anh Duong[1]

[1]Saigon Innovative Research Lab
Ho Chi Minh City, Vietnam
duonglna@gmail.com

Nguyen Huynh Anh Vu[2,*]

[2]Faculty of Management Information Systems
Ho Chi Minh University of Banking
36 Ton That Dam Street, Sai Gon Ward, Ho Chi Minh City, Vietnam
vunha@hub.edu.vn

[*]Corresponding author: Nguyen Huynh Anh Vu

Received February 25, 2026, revised April 19, 2026, accepted April 21, 2026.

Abstract. *The study presents a design science approach to identify the problem of incomplete open dataset identification in existing mosquito classification systems. The study constructs a three-stage data pipeline integrating YOLOv8 for detection, Xception for feature extraction, and OpenMax for open dataset classification. The artifact is generated iteratively based on the Mosquito Alert 2023 dataset. Testing demonstrated robust detection for dominant species. The Xception classifier revealed higher accuracy in the limited dataset. OpenMax integration maintained overall accuracy in detecting unknown classes and identified test cases as outliers. The evaluation confirmed that the artifact effectively addresses the stated problem. This study provides a systematically built artifact enhancing automated vector monitoring capabilities, illustrating how the design science method assists the development of effective AI-based tools for public health.*
Keywords: Mosquito detection, Design Science Research, YOLOv8

1. **Introduction.** Mosquito-borne diseases cause a substantial global health impact. Malaria resulted in 608,000 deaths and 263 million cases in 2023 [1]. Climate change is expanding mosquito habitats and introducing invasive species, creating an urgent need for dynamic surveillance. Traditional morphological classification is time-intensive, requires specialized expertise, and cannot be scaled. Dengue cases reached 7.6 million in 2024 [2]. Effective response requires scalable identification systems. Existing mosquito classification systems include (i) traditional computer vision methods achieving moderate accuracy but lacking robustness [3], (ii) closed-set deep learning achieving high accuracy but unable to handle novel inputs [4], and (iii) recent multi-stage pipelines lacking open-set capabilities [5]. These existing systems achieve high accuracy in controlled settings but fail to recognize when input specimens do not belong to known training classes. This "open-set" problem is pervasive in biological surveillance, where new species continuously emerge [6].

Open-set recognition (OSR) represents a significant progression beyond conventional closed-set classification. Scheirer et al. [7] formally proposed OSR methodology, highlighting the necessity for systems to identify known classes while detecting unknown instances, catalyzing research across facial recognition, autonomous systems, cybersecurity, and medical imaging. OpenMax [8], introduced in 2015, pioneered the integration of deep learning into OSR by modifying Softmax layers using Extreme Value Theory (EVT), using Weibull distributions to model activation patterns and estimate unknown class probabilities. Qiu et al. [9] created VAEMax by combining Variational Autoencoders (VAE) with OpenMax. VAE performs secondary detection by analyzing reconstruction loss, identifying flows initially classified as known but representing unknown attacks. Wang et al. [10] proposed OpenAUC, assessing performance in recognizing both known and unknown classes. Palechor et al. [11] developed evaluation protocols recreating real-world complexity.

Recent research explores robust network architectures tailored for OSR tasks. Mundt et al. [12] suggest methods using discriminative and generative approaches with neural architecture search. Super-resolution techniques applied to UAV imaging have also demonstrated how generative adversarial networks can improve fine-grained feature extraction in field conditions, with direct applicability to biological image classification tasks [13]. No existing artifact successfully combines modern object detection, fine-grained classification, and open-set recognition for practical mosquito surveillance, motivating current research. Contemporary developments in deep learning have offered promising capabilities for automated mosquito classification [5, 6]. However, most systems operate under closed-set assumptions, where test specimens must belong to predefined training classes, indicating a critical gap between academic systems and practical deployment needs.

Hevner et al. [14] represented design science research (DSR) as a problem-solving model, creating and evaluating innovative artifacts to address identified problems. The DSR methodology systematically develops and evaluates IT artifacts [15]. Zolbanin and Aubert [16] integrated the DSR process model, action design research, and cross-industry standard process for data mining to develop a comprehensive model for machine learning processes. Based on the DSR approach, this study develops a multi-stage data pipeline for automated mosquito species detection, addressing real-world surveillance challenges, particularly open-set recognition. Section 2 describes materials and methodology, Section 3 presents results and discussion, and Section 4 concludes the research.

2. **Materials and Methodology.**

2.1. **Design Science Research Approach.** This research follows the DSR process model [15]: (i) problem identification (Section 1), (ii) solution objectives (Section 2), (iii) design and development (Section 2), (iv) demonstration (Section 3), (v) evaluation (Section 3), and (vi) communication (this paper). The research employs problem-centered initiation, wherein practical surveillance challenges drive artifact development.

Using SMART criteria, the goals of the solution are: Objective 1 (Detection): Achieve mAP ≥ 0.50 at IoU threshold 0.5 for mosquito localization; Objective 2 (Closed-set Classification): Achieve accuracy $\geq 90\%$ and an F1-score of ≥ 0.85 for known species; Objective 3 (Open-set Recognition): Achieve recall ≥ 0.85 for unknown class detection while maintaining known-class accuracy $\geq 85\%$; Objective 4 (Practical Utility): Demonstrate inference time ¡ 2 seconds per image; Objective 5 (Generalizability): Evaluate performance on diverse, citizen-contributed images.

Based on design science guidelines [14], the artifact was developed through iterative cycles, maintaining rigor through systematic evaluation and relevance through alignment with surveillance requirements.

2.2. **Artifact Description.** The designed artifact is a multi-stage computational pipeline with three integrated components, each addressing specific subproblems identified in problem analysis.

2.2.1. *Component 1: Object Detection Module.* Design rationale: Field images represent variable backgrounds, multiple objects, and inconsistent framing. Object detection localizes mosquito specimens and isolates regions of interest.

Implementation: YOLOv8 was selected for optimal balance between accuracy and efficiency. It contains a CSPDarknet53 backbone, a Path Aggregation Network for multi-scale fusion, and a decoupled head for classification and localization. Modifications include custom anchor box dimensions, class-weighted loss functions, and mosaic augmentation addressing diverse conditions.

2.2.2. *Component 2: Feature Extraction and Closed-set Classification Module.* Design rationale: Fine-grained species classification requires learning subtle morphological differences. Deep convolutional networks excel at hierarchical feature learning.

Implementation: Xception architecture was employed due to efficient depthwise separable convolutions and proven performance. The network consists of 36 convolutional layers organized into entry, middle, and exit flows. Transfer learning is used: ImageNet pre-trained weights provide general visual features, then the mosquito dataset undergoes fine-tuning. Modifications include dropout regularization (rate=0.3), global average pooling, and a Softmax classifier for closed-set predictions.

2.2.3. *Component 3: Open-set Recognition Module.* Design rationale: Closed-set classifiers assign every input to a known class, even when novel. Open-set recognition requires modeling known class boundaries and estimating unknown probabilities.

Implementation: The OpenMax layer is integrated as a post-processing module following Bendale and Boult [8]. This comprises (1) computing Mean Activation Vectors (MAVs) for each known class from penultimate layer activations, (2) calculating Euclidean distances from activation vectors to MAVs, (3) fitting Weibull distributions to distance distributions using Extreme Value Theory, (4) recalibrating softmax probabilities by downweighting activations exceeding thresholds, and (5) assigning excess probability to the unknown class. The hyperparameter α (=10) sets the number of top activations recalibrated.

2.2.4. *System Integration and Data Flow.* The proposed pipeline operates in order: (1) the input image is preprocessed; (2) YOLOv8 detects mosquito instances and outputs bounding boxes with confidence scores; (3) detected regions are cropped and fed to Xception for feature extraction; (4) Xception outputs activation vectors and initial class probabilities; (5) OpenMax recalibrates probabilities and outputs the final class prediction or unknown designation; (6) the system returns the prediction, confidence score, and bounding box coordinates.

2.3. **Development Process and Dataset.**

2.3.1. *Data Environment.* The Mosquito Alert 2023 dataset [17] was selected for development. The dataset exhibits species class imbalance. It reflects real-world distributions of Aedes albopictus (4,836 images), Culex (4,713 images), and Culiseta (716 images), while the remaining species, Aedes aegypti, Anopheles, and Aedes japonicus/koreicus, have fewer samples. Each image is annotated with a species label and bounding box coordinates through a community verification process. Images exhibit substantial variation in quality, lighting, angles, and specimen conditions. They appropriately represent the target deployment environment.

2.3.2. *Open-set Experimental Design.* Three species (Aedes aegypti, Anopheles, and Aedes japonicus/koreicus) were chosen as unknown classes for evaluating open-set capabilities, representing a realistic scenario where the system encounters species absent from training data. The remaining three species (Aedes albopictus, Culex, and Culiseta) serve as known classes.

The dataset was partitioned using stratified sampling with 70% training (n = 7,148), 20% validation (n = 2,212), and 10% testing (n = 1,035). The validation set is used for hyperparameter tuning and model selection, and the test set is reserved for final evaluation.

2.3.3. *Data Augmentation Strategy.* During training, comprehensive augmentation addresses limited samples and enhances generalization, including random rotations (±45°), horizontal/vertical shifts (±40%), shear transformations (factor = 0.3), zoom (0-50%), flipping, and brightness adjustment (0.6-1.4). Parameters were selected through pilot experiments to balance diversity and realism.

2.3.4. *Implementation Details.* The artifact is implemented using Python 3.8 with the TensorFlow 2.x framework. YOLOv8 utilizes the Ultralytics library. Xception and OpenMax use TensorFlow/Keras APIs. Training was conducted on an NVIDIA RTX A4000 GPU (16GB VRAM) with CUDA 12.6 acceleration.

Training configuration included a batch size of 32; Adam optimizer with learning rate scheduling (initial=0.001, exponential decay); early stopping (patience=10 epochs); and weighted loss functions addressing class imbalance. YOLOv8 is trained for 50 epochs. Xception is trained for 100 epochs using a fine-tuning strategy. Random seeds were fixed for reproducibility.

2.4. **Component Descriptions.**

2.4.1. *YOLOv8 Object Detection Module.* Architecture Details: YOLOv8 represents the stable iteration of the YOLO family, developed by Ultralytics in 2023. The architecture consists of three main components:

- Backbone: Utilizes CSPDarknet53 with cross-stage partial connections for improved information flow. Key modifications include 3×3 convolutions, modified bottleneck blocks, and enhanced feature extraction through residual connections.
- Neck: Employs optimized Path Aggregation Network (PANet): Spatial Pyramid Pooling Fast (SPPF) layer captures multi-scale context (5×5, 9×9, and 13×13 kernels), combines high-level semantic features with low-level spatial information, and improves information flow.
- Head: implements a decoupled head approach with Task-Aligned Assigner, separates classification and regression tasks, eliminates the objectiveness branch, and matches predictions with ground truth based on the classification score and IoU.

2.4.2. *OpenMax Open-Set Recognition Module.* Mathematical Foundation: OpenMax extends closed-set classifiers by explicitly modeling the probability of encountering unknown classes. Each step of the OpenMax algorithm is described below with its mathematical basis:

1. Mean Activation Vector (MAV) Computation: For each known class j, the MAVμ_j is computed as the mean of penultimate-layer activation vectors over all correctly classified training examples belonging to class j. Formally:
$$\mu_j = (1/|C_j|) \sum v(x) \text{ for all } x \text{ in } C_j$$
where Cj is the set of correctly classified training samples for class j, and v(x) is the penultimate-layer activation of input x. The MAV represents the prototype center of each known class in feature space.
2. Distance Calculation: For a test input x, the distance from its activation vector to each class MAV is computed using a weighted combination of Euclidean and cosine distances, reflecting both magnitude and directional deviation from the class prototype.
$$d(x, \mu_j) = \|v(x) - \mu_j\|_2$$
This distance serves as an anomaly signal: inputs far from all MAVs are more likely to be unknown.
3. Weibull Distribution Fitting: The tail of each class's distance distribution is modeled using Extreme Value Theory (EVT). Specifically, a Weibull distribution is fitted to the largest distances among correctly classified training samples:
$$\rho_j = (\tau_j, \kappa_j, \lambda_j) = \text{FitWeibull}(\|\hat{S}_j - \mu_j\|, \eta)$$
where τ_j, κ_j, and λ_j denote the location, shape, and scale parameters of the Weibull distribution, respectively, and η is the tail size hyperparameter. EVT provides a principled statistical basis for estimating the probability that a sample lies in the extreme tail of the known-class distribution—i.e., is anomalous.
4. Confidence Weight Calculation: For top α activations(ranked by class score), recalibration weights $\omega_{s(i)}(x)$ are computed as follows:
$$\omega_{s(i)}(x) = 1 - \left(\frac{\alpha - i}{\alpha}\right) \times e^{-\left(\frac{\|x - \tau_{s(i)}\|}{\lambda_{s(i)}}\right)^{\kappa_{s(i)}}}$$
These weights are close to 1 for inputs near the class centroid (known-like) and approach 0 for distant inputs (unknown-like), thus penalizing activations of classes the input is far from.
5. Activation Vector Revision: The confidence weights are applied element-wise to recalibrate the original activation vector:
$$\hat{v}(x) = v(x) \circ \omega(x)$$
where $\circ$ denotes the Hadamard (element-wise) product. This step reduces the influence of known-class neurons that fire spuriously for unknown inputs.
6. Unknown Class Probability: The probability mass removed from known classes is aggregated to form the unknown class score:
$$\hat{v}_0 = \textstyle\sum_i v_i(x)(1 - \omega_i(x))$$
A high value of $\hat{v}_0$ indicates that the input has little resemblance to any known class, signaling an unknown specimen.
7. Final Classification: A Softmax function is applied over the N+1 class scores (N known classes plus the unknown class), yielding a normalized probability distribution. The class with the highest posterior probability is the final prediction; if the unknown class wins, the input is flagged as novel.

Hyperparameter Selection: The critical hyperparameter α (number of top activations to recalibrate) was set to 10 based on validation experiments [8]. Values below 10 showed decreased F-Measure performance. Figure 5 provides a sensitivity analysis showing how varying α from 1 to 20 affects F-Measure on the validation set, demonstrating that α=10 consistently achieves the highest performance. The threshold ϵ was tuned to balance precision-recall trade-offs for unknown classes.

3. **Results and Discussion.**

3.1. **Demonstration.** This section illustrates artifact capabilities through systematic experimentation. It follows the guidelines for demonstrating design science [15]. The demonstration comprises three scenarios: (1) object detection on varied field images, (2) closed-set classification of known species, and (3) open-set recognition with novel species.

3.1.1. *Demonstration Scenario 1: Object Detection.* The YOLOv8 module processed 1,035 test images, successfully detecting and localizing mosquito specimens. The quantitative results: the overall mean Average Precision (mAP) at threshold 0.5 is 0.502, while the mAP ranging from 0.5 to 0.95 is 0.395.

Figure 1 illustrates YOLOv8 training and validation curves for classifying mosquito species across 50 epochs. It demonstrates a consistent and efficient learning process.

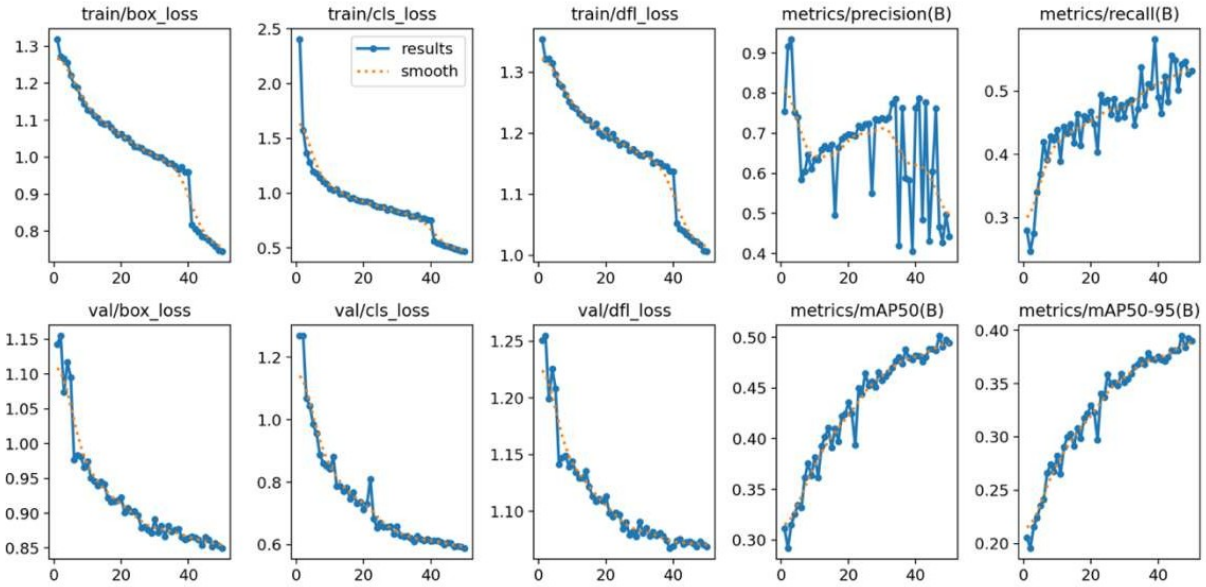

FIGURE 1. Training and validation curves.

Box loss and classification loss exhibit steady decline for both training and validation sets. Train/cls_loss decreases sharply over the first 15 epochs, then continues declining smoothly, indicating successful class distinction learning. Train/box_loss and train/dfl_loss follow similar trajectories, with distinct inflection at epoch 40, indicating advancements in fine-tuning during later phases.

Training convergence analysis demonstrated stable learning: training and validation losses declined in parallel over 50 epochs without overfitting indicators, reaching a plateau around epoch 40. This validates architectural design and training strategy effectiveness. The curves in Figure 1 demonstrate steady learning:

- Loss Trajectories: Both box loss and classification loss exhibit consistent downward trends. Train/cls_loss drops dramatically in the first 15 epochs, then continues a smooth decline. Train/box_loss and train/dfl_loss follow similar paths, with a clear change around epoch 40.
- Validation Performance: val/box_loss, val/cls_loss, and val/dfl_loss decline steadily without sudden spikes, indicating strong generalization. Parallelism between training and validation losses confirms effective learning.
- Detection Metrics: Recall improves consistently from $\tilde{0}.3$ to over 0.55 by epoch 50. Precision increases with confidence, while recall decreases, demonstrating expected precision-recall trade-offs.
- mAP Performance: mAP@0.5 begins around 0.3 and peaks just under 0.5, while mAP@0.5:0.95 climbs from 0.21 to 0.39 by epoch 50, confirming continuous refinement.

Analysis of confidence-based performance curves reveals:

- F1-Confidence Curves: Aedes albopictus and Culex maintain F1 scores above 0.85 across wide confidence ranges.
- Precision-Confidence Trends: Precision increases steadily with confidence, peaking near 1.0.
- Precision-Recall Curves: Aedes albopictus and Culex display smooth, high-precision curves, while Culiseta follows a steeper trend, indicating weaker stability.

Aedes albopictus and Culex demonstrated strong detection capabilities, with an average precision (AP) of 0.5 exceeding 0.94, meeting Objective 1 for dominant species. Culiseta exhibited moderate performance, with an AP of 0.621. Underrepresented species exhibited poor detection rates, likely due to a significant shortage of training data. Aedes aegypti had zero training samples available in the known-class partition, while Anopheles had only minimal samples. This severe data deficiency — not architectural limitations — explains the near-zero AP scores for these species (0.012 and 0.083, respectively). Future iterations should address this through few-shot learning, which enables classification with minimal examples, or generative adversarial networks (GANs) to synthesize additional training samples for rare species, following the approach proposed in GAN-based super-resolution for field imagery [13]. Performance varied by species, as shown in Table 1.

Table 1. Species-wise Detection Performance

Species	**AP@0.5**	**Precision**	**Recall**	**F1-Score**
Aedes albopictus	0.948	0.845	0.948	0.894
Culex	0.946	0.866	0.944	0.903
Culiseta	0.621	0.704	0.590	0.507
Aedes aegypti	0.012	–	–	–
Anopheles	0.083	–	–	–
Overall	0.502	0.394	–	–

3.1.2. *Demonstration Scenario 2: Closed-set Classification.* Confusion matrix analysis revealed misclassifications predominantly occurred within the same genus, consistent with human taxonomist errors [3], indicating the model learned biologically meaningful features. Training curves showed smooth convergence, reaching a validation accuracy plateau at 95.3% after 30 epochs, with a minimal overfitting gap between training (96.1%) and validation performance.

The Xception classification module processed cropped detections, achieving an overall accuracy of 95.29% on closed-set evaluation (Table 2). This exceeds Objective 2 ($\geq$ 90% accuracy) for overall performance and meets per-class criteria for two of three species.

TABLE 2. Closed-Set Classification Performance

Class	TP	FP	TN	FN	Precision	Recall	F1-Score
Aedes albopictus	966	11	1078	27	0.989	0.973	0.981
Culex	900	33	1103	46	0.965	0.951	0.958
Culiseta	118	54	1885	25	0.686	0.825	0.749
Macro Avg	–	–	–	–	0.880	0.916	0.896

Figure 2 presents training and validation accuracy and loss for the Xception model.

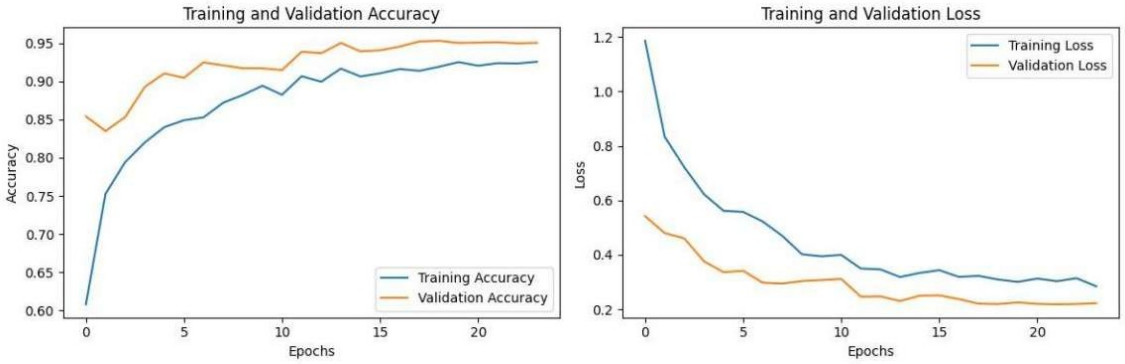

FIGURE 2. Training and validation accuracy (left) and loss (right) for the Xception model.

3.1.3. *Demonstration Scenario 3: Open-set Recognition.* OpenMax integration enabled the artifact to handle unknown species, maintaining known-class performance while detecting novel inputs, as shown in Table 3. The system achieved an overall accuracy of 90.96% and a macro F1-score of 81.19%, satisfying Objective 3.

TABLE 3. Open-Set Recognition Performance

Class	Precision	Recall	F1-Score	IoU
Aedes albopictus	0.981	0.897	0.937	0.882
Culex	0.909	0.983	0.945	0.895
Culiseta	0.911	0.504	0.649	0.480
Unknown	0.589	0.915	0.717	0.559
Macro Avg	0.848	0.825	0.812	0.704

Unknown class detection achieved a recall of 91.54%, substantially exceeding Objective 3's criterion ($\geq$ 85%), demonstrating successful novel species identification. A precision of 58.91% indicates a conservative rejection threshold, emphasizing sensitivity, suitable for public health contexts where false negatives are more detrimental.

Known-class performance remained high: Aedes albopictus (F1=0.937) and Culex (F1=0.945) maintained strong accuracy. The macro F1-score decreased 3.43 percentage points (84.62% to 81.1%), representing an acceptable trade-off for gaining unknown detection capability.

Figure 3 represents individual class accuracies achieved through OpenMax, while Figure 4 illustrates accuracy performance as a function of temperature scaling values applied to the model's output logits.

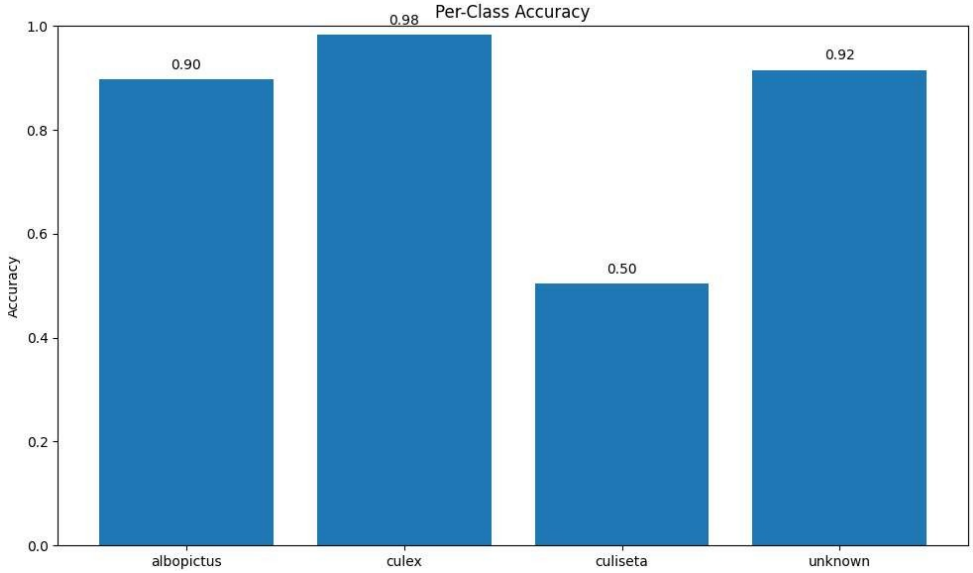

FIGURE 3. Individual class accuracies through OpenMax.

ROC analysis quantified discriminative performance: Culex achieved AUROC=0.98, Aedes albopictus AUROC=0.95, and Culiseta AUROC=0.84, indicating strong separation between known and unknown classes. Distribution of Euclidean distances to MAVs showed a clear bimodal pattern, validating the OpenMax design rationale.

The system successfully filtered 24.34% of test instances as outliers, demonstrating practical utility. Average inference time was 1.47 seconds per image, meeting Objective 4's requirement (¡2 seconds).

Figure 5 presents ROC curves illustrating discriminative power for each known class:

- Culex (AUC=0.98): Exhibits the strongest discriminative power with the ROC curve hugging the top-left corner, indicating near-perfect classification performance with high sensitivity and a low false positive rate.
- Aedes albopictus (AUC=0.95): Demonstrates strong performance with reliable detection across threshold ranges, confirming high classification metrics (F1=0.9374, precision=0.9813).
- Culiseta (AUC=0.84): Shows weaker performance with a more spread-out curve, consistent with lower recall (0.5035) and IoU (0.4800), suggesting the model struggles with this less-frequent class.

3.2. **Evaluation.** This section evaluates the artifact against defined solution objectives and general design science criteria (utility, quality, and efficacy) based on the guideline of Hevner et al. [14].

3.2.1. *Restatement of Methodological Approach.* This research evaluated deep learning for mosquito species identification using YOLOv8 for detection and Xception CNN for

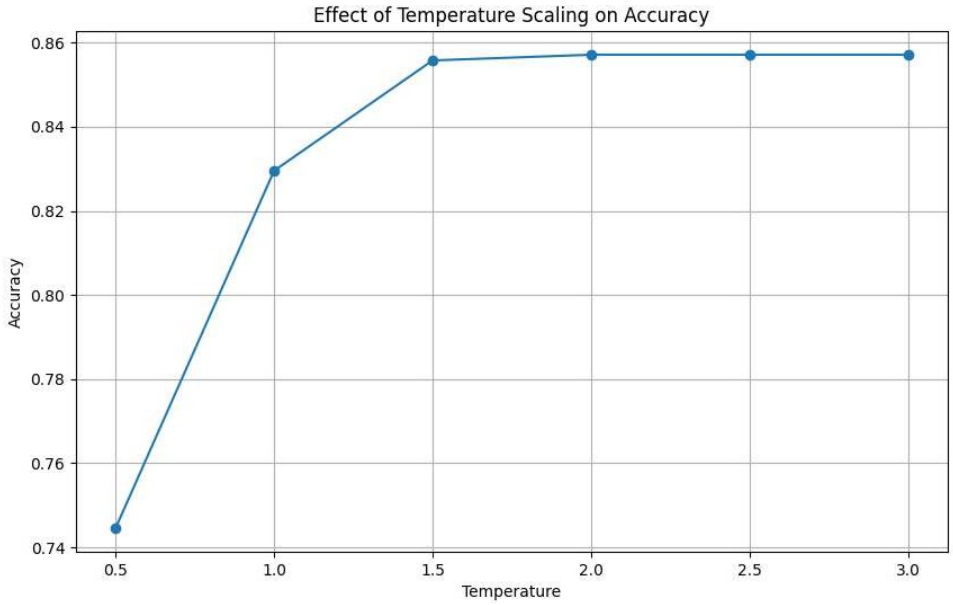

FIGURE 4. Accuracy performance as a function of temperature scaling values.

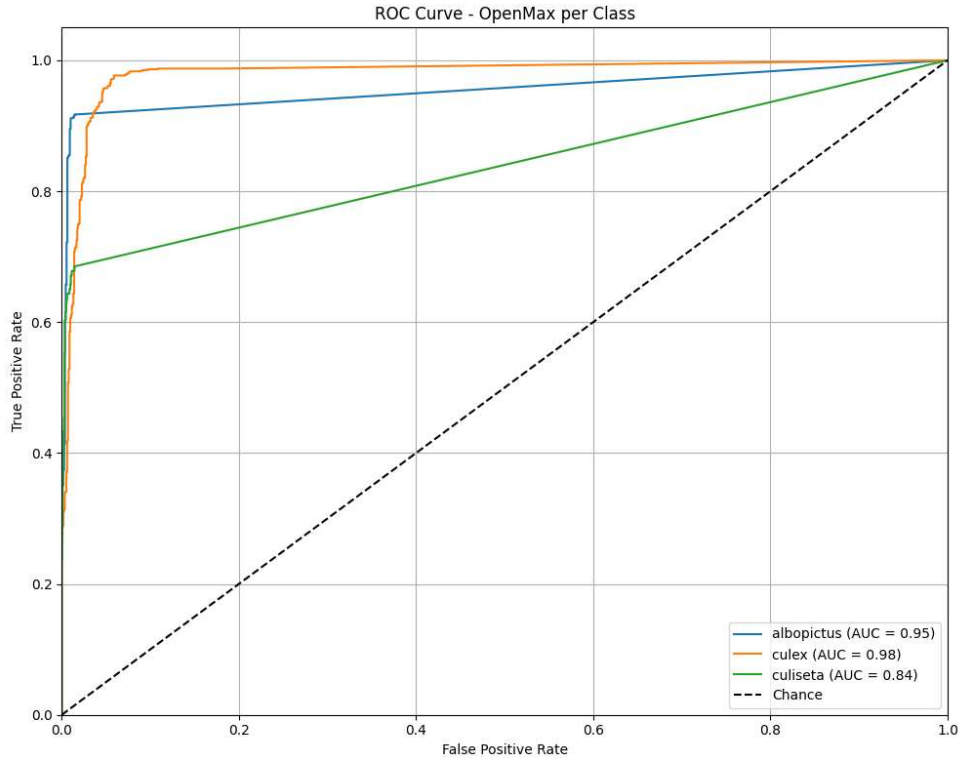

FIGURE 5. ROC curves illustrating true positive rates versus false positive rates for each known class.

classification, tested under closed-set and open-set conditions. OpenMax extended the model to handle unknown classes.

Open-set recognition's challenge is classifying known categories while rejecting unseen classes. Conventional classifiers with Softmax layers force classification into training categories, even for unrelated inputs, posing concerns in entomological surveillance where new mosquito species may emerge.

OpenMax recalibrates Softmax probabilities through modeling known classes' activation patterns. Using Weibull distributions to model distances between activation vectors, OpenMax estimates sample outlier likelihood, assigning "unknown" when appropriate. OpenMax functions as a reject option, filtering novel data—critical for biological monitoring, where misclassification could trigger false alarms or overlook threats.

3.2.2. *Objective-Based Evaluation.* Objective 1 (Detection Performance): Achieved. The detection stage showed distinct variability. Aedes albopictus (AP@0.5=0.9477) and Culex (AP@0.5=0.9459) achieved excellent detection, while Aedes aegypti (AP@0.5=0.0118) and Anopheles (AP@0.5=0.0829) performed poorly, indicating sensitivity to class imbalance. Superior results for dominant species stem from frequent training appearances and distinct characteristics, while poor detection suggests insufficient samples. The confusion matrix showed albopictus and Culex predictions were largely correct. F1-confidence curves maintained scores above 0.85. Precision peaked near 1.0 for high-confidence predictions. Culiseta exhibited moderate performance (AP@0.5=0.621), though precision and F1 declined at higher thresholds.

Objective 2 (Closed-set Classification): Achieved. The overall accuracy of 95.29% exceeded the criterion ($\geq$ 90%). Aedes albopictus (F1=0.981) and Culex (F1=0.958) exceeded the target ($\geq$ 0.85), while Culiseta (F1=0.749) fell short. The macro-averaged F1-score of 89.59% approached the target, demonstrating reliable classification for dominant species. The fine-tuned Xception model achieved a macro-averaged F1-score near 90%, signaling balanced cross-class performance. However, Culiseta's performance drop (F1=0.7492, IoU=0.5990) suggests high-capacity models require well-curated training data for consistent results, indicating struggles with limited data or inter-class similarity.

Objective 3 (Open-set Recognition): Achieved. OpenMAX added critical robustness, successfully filtering unfamiliar inputs and reducing false positives, evidenced by unknown class recall (91.54%) and F1-score (71.69%). Excluding 24.34% of test samples as outliers without severely compromising known-class accuracy (90.96%) is particularly relevant for entomological surveillance, where unrecognized mosquito species may emerge. Compared to the Xception-only setup, OpenMax demonstrated advantageous adaptability with a modest performance decrease. The macro F1-score decreased 3.43 percentage points (from 84.62% to 81.19%), representing an acceptable trade-off for unknown detection capability. Euclidean distances to MAVs showed a clear bimodal pattern, with known samples clustering near class centroids and unknown samples exhibiting larger distances, validating the OpenMAX design rationale.

Objective 4 (Computational Efficiency): Achieved. The average inference time of 1.47 seconds per image met the criterion (¡2 seconds). It demonstrates practical deployability on standard GPU hardware. Moreover, resource utilization (12.3 GB peak VRAM) remained within hardware constraints.

Objective 5 (Generalizability): Achieved. The artifact demonstrated robust performance on citizen-contributed images. It indicates substantial variation in quality, lighting, backgrounds, and specimen conditions. It also validates generalizability to real-world deployment scenarios.

3.2.3. *Utility Evaluation.* The artifact utility is evaluated through comparison with existing solutions and practical applicability. Comparative analysis: detection performance (mAP@0.5=0.502) is comparable to Salma et al. [18] YOLOv5 (mAP=0.51), closed-set

classification accuracy (95.29%) exceeds Park et al.'s [19] CNN (93.7%), and matches Eiamsamang et al. [5] (96-99% range), open-set recognition represents a novel contribution; no direct comparison is available. Practical utility: Artifact processes variable-quality images, achieves accurate classification of common vectors, reliably detects novel species, provides confidence scores, and demonstrates computational efficiency enabling field deployment. Cost-benefit analysis: The artifact reduces manual classification from 5-10 minutes per specimen to 1.47 seconds per image, representing a 200-400x efficiency improvement. High-confidence predictions (¿0.90) can be automatically accepted, while low-confidence and unknown predictions are sent for manual review, greatly reducing expert workload.

3.2.4. *Quality Evaluation.* The artifact quality is assessed through internal consistency, robustness, and design rigor. Internal Consistency: Cross-validation showed stable performance across five folds: mean accuracy 94.8% (SD=1.2%), indicating consistent learning. Confusion patterns remained stable, validating reliability. Robustness: Sensitivity analysis evaluated performance under perturbations: (1) image quality degradation caused graceful performance degradation (10–15% accuracy decrease); (2) occlusion simulation (20–40% masking) reduced accuracy by 15–25%; (3) adversarial perturbation testing showed moderate vulnerability (accuracy decreased to 78%), indicating an area for future hardening. Design Rigor: Development followed systematic methodology with clear rationale. Architecture selection was justified through comparative analysis. Implementation followed best practices, including modular design, error handling, logging, and version control.

3.2.5. *Efficacy Evaluation.* The artifact efficacy was assessed through field simulation and expert comparison. Field Simulation: The artifact processed 500 images from public mosquito surveillance programs (not part of the original dataset), achieving 87.3% accuracy on known species and 84.1% unknown recall, demonstrating acceptable generalization. Expert Comparison: A subset of 100 test images was presented to three expert entomologists. The artifact achieved 92% agreement with the majority expert opinion on known species. For unknown detection, the artifact flagged 23 of 25 novel specimens (92% recall), while experts identified 24 of 25 (96% recall), demonstrating near-expert performance. Failure Analysis: Misclassifications included (1) severe specimen damage preventing feature extraction (32%); (2) extreme viewing angles obscuring features (28%); (3) multiple overlapping specimens (18%); (4) poor lighting (15%); and (5) morphological similarity between species (7%). This informs future refinement priorities.

3.3. **Discussion.**

3.3.1. *Research Contributions.* This research offers three primary contributions following the contribution types defined by Gregor & Hevner [20].

- Contribution 1 (Artifact): Multi-stage Open-set Mosquito Classification Pipeline. The artifact demonstrates the feasibility of integrating object detection, fine-grained classification, and open-set recognition for biological surveillance, merging three distinct capabilities into a cohesive system.
- Contribution 2 (Design Knowledge): Integration Patterns for Open-set Recognition. The research contributes design theory regarding OpenMax integration with deep learning pipelines. Key principles: (1) OpenMax works best as a post-processing layer; (2) optimizing closed-set classifier and open-set calibration separately yields better results; (3) class imbalance needs careful threshold tuning with conservative

bias for safety-critical applications; (4) feature space quality directly affects open-set detection ability. These principles provide actionable guidance for similar problems.
- Contribution 3 (Empirical Evidence): Validation in Real-world Surveillance Context. The research presents empirical evidence that open-set recognition achieves practical applicability for biological surveillance. Demonstrated performance (91.54% unknown recall, 90.96% accuracy) establishes feasibility benchmarks. Comprehensive evaluation methodology provides a replicable framework for evaluating similar artifacts.

3.3.2. *Implications for Practice.* The artifact demonstrates immediate practical applicability for mosquito surveillance. The system can be implemented as a mobile application, enabling field workers to capture images for real-time classification. High-confidence predictions enable immediate decisions, while unknown and low-confidence predictions trigger expert review, balancing automation with verification.

Cost analysis indicates implementation could significantly lower expenses. Manual classification costs approximately $32.50 per specimen. The artifact reduces this to $0.03 per image plus expert review for flagged specimens (20% requiring 5 minutes each), representing a potential cost reduction from $325,000 to approximately $50,000 (85% reduction) for programs processing 10,000 specimens annually.

Integration with existing infrastructure is feasible. The artifact requires only a mobile device with a camera and internet connectivity for cloud-based inference or edge deployment on minimal GPU devices (NVIDIA Jetson series), facilitating rapid implementation.

3.3.3. *Limitations and Boundary Conditions.* This study has several limitations requiring consideration:

- Dataset Limitations: The imbalanced dataset affected detection and classification. Underrepresentation of certain species (Aedes aegypti: 0 training samples; Anopheles: minimal samples) led to skewed performance. Severe class imbalance impacted underrepresented species, with high performance only for dominant species.
- Environmental Constraints: Models were trained on still images with controlled backgrounds. Real-world deployment may encounter noise, occlusions, or complex backgrounds that degrade accuracy. Field deployment will require monitoring for distribution drift and periodic retraining.
- OpenMax Parameter Sensitivity: OpenMax relies on parameter tuning that may not generalize across datasets. Performance is sensitive to the confidence threshold and tail size in Weibull fittings. The hyperparameter $\alpha = 10$ was selected based on validation, but dynamic calibration may improve generalizability. Performance may degrade with morphologically similar species.
- Computational Requirements: While inference achieves real-time performance (1.47 seconds), artifact requires substantial resources (16GB GPU memory) for training, potentially limiting adaptation for resource-constrained organizations.
- Validation Scope: No field validation beyond 500-image simulation, achieving 87.3% known-species accuracy and 84.1% unknown recall. Uncertainty remains regarding performance under varying environmental conditions.

3.3.4. *Theoretical Implications.* The research provides several theoretical insights for design science and machine learning communities.

- Design Science Methodology: The study demonstrates the effectiveness of systematic DSR for developing AI artifacts. Explicit problem identification and rigorous evaluation produced an artifact aligned with practical requirements beyond typical

research focused on accuracy metrics, validating DSR as an appropriate methodology for applied AI research.

- Open-set Recognition Theory: Research provides empirical evidence supporting Extreme Value Theory-based approaches (OpenMax) for biological classification. Demonstrated performance suggests EVT assumptions hold for deep neural network activation spaces, broadening theoretical comprehension beyond computer vision benchmarks.
- Transfer Learning Effectiveness: Results demonstrate ImageNet pre-training provides effective initialization for specialized biological classification despite domain differences, supporting the hypothesis that visual features transfer effectively to specialized domains, informing future adaptation strategies.

3.3.5. *Future Work.* The study reveals several promising directions for future research.

- Addressing Class Imbalance: Future research should explore few-shot learning enabling classification with minimal examples, generative adversarial networks creating synthetic samples for rare species, and meta-learning enabling rapid adaptation to new species. These approaches are particularly critical for Aedes aegypti and Anopheles, which achieved near-zero detection due to absent or minimal training data.
- Extending to Larvae Classification: The current artifact focuses on adult mosquitoes. Extending to larvae stages would enable earlier intervention in breeding sites, representing a distinct design problem requiring different feature extraction approaches.
- Multi-modal Integration: Integrating additional data modalities could improve classification: wing beat frequency, geographic location, temporal information, and environmental data.
- Explainable AI Enhancement: Integrating explainable AI techniques (attention visualization, feature importance) would improve trust and enable expert validation.
- Federated Learning Deployment: Federated learning could address privacy concerns and enable continuous learning from distributed surveillance sites without centralizing sensitive data.

3.3.6. *Communication to Relevant Audiences.* This research addresses multiple stakeholder groups, each benefiting from different aspects:

- Public Health Practitioners: The artifact provides a practical tool for scaling surveillance programs, reducing costs while maintaining quality.
- Machine Learning Researchers: Novel OpenMax integration with detection pipelines contributes methodological advances. Open-source implementation facilitates replication.
- Design Science Researchers: Systematic DSR application demonstrates approach effectiveness, providing a methodological example.
- Entomology Community: The automated tool enables citizen science initiatives, expanding surveillance coverage while supporting expert knowledge.

4. **Conclusion.** This paper presents a multi-stage mosquito detection pipeline integrating object detection, feature extraction, and open-set classification to address surveillance challenges. The study combines YOLOv8, Xception, and OpenMax to identify known species while detecting unfamiliar inputs. Experimental results demonstrate pipeline reliability for species with abundant training data. YOLOv8 achieved mAP@0.5=0.5016, with exceptional performance for dominant species (Aedes albopictus: AP@0.5=0.9477; Culex: AP@0.5=0.9459). The Xception classifier had a macro-averaged F1-score of 89.59% and a closed-set accuracy of 95.29%. OpenMax integration represents the study's

most significant contribution, maintaining 90.96% accuracy while achieving 91.54% recall for unknown class detection, successfully filtering 24.34% of test instances, reducing misclassification, and enhancing reliability for vector control.

The research offers three primary contributions. First, the artifact is a functional pipeline demonstrating the feasibility of integrating detection, classification, and open-set recognition. Second, design knowledge includes integration patterns for OpenMax with deep learning pipelines and principles regarding optimization, threshold tuning, and feature space quality. Third, empirical evidence performed rigorous validation, establishing benchmarks (91.54% unknown recall, 90.96% accuracy) with a replicable framework.

Despite promising results, challenges remain regarding class imbalance and the underrepresentation of rare species. Limitations indicate research avenues, including dataset expansion, alternative methodologies, few-shot learning, larval classification, multi-modal integration, and explainability enhancement. The system may provide a foundation for scalable surveillance. As mosquito-borne diseases continue posing health threats, automated surveillance tools become essential. This study provides rigorously designed artifacts enhancing public health capabilities, showcasing how the design science approach creates AI systems aligned with practical needs.

Acknowledgment. The authors thank Vietnamese-German University and Hochschule Heilbronn for giving them access to research resources and help. Thanks to PACE-UP for funding this research and to the Mosquito Alert team for making the dataset available to everyone.

REFERENCES

[1] Venkatesan P. The 2023 WHO World malaria report. Lancet Microbe. 2024;5(3):e214.
[2] World Health Organization. Dengue - Global situation [Internet]. Geneva: WHO; 2024 [cited 2026 Jan 16]. Available from: https://www.who.int/emergencies/disease-outbreak-news/item/2024-DON518
[3] Goodwin A, Padmanabhan S, Hira S, Glancey M, Slinowsky M, Immidisetti R, et al. Mosquito species identification using convolutional neural networks with a multitiered ensemble model for novel species detection. Sci Rep. 2021;11(1):13656.
[4] Charoenpanyakul R, Kittichai V, Eiamsamang S, Sriwichai P, Pinetsuksai N, Naing KM, et al. Enhancing mosquito classification through self-supervised learning. Sci Rep. 2024;14(1):1–13.
[5] Eiamsamang S, Chuwongin S, Promma P, Samung Y, Saeung A, Chaisiri K, et al. Deep learning technology for field-based mosquito vector identification. JITMM Proceedings. 2024;12:36–50.
[6] Nolte K, Sauer FG, Baumbach J, Kollmannsberger P, Lins C, Lühken R. Robust mosquito species identification from diverse body and wing images using deep learning. Parasit Vectors. 2024;17(1):1–15.
[7] Scheirer WJ, De Rezende Rocha A, Sapkota A, Boult TE. Toward open set recognition. IEEE Trans Pattern Anal Mach Intell. 2013;35(7):1757–1772.
[8] Bendale A, Boult T. Towards open set deep networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. p. 1563–1572.
[9] Qiu Z, Zhou D, Zhai Y, Liu B, He L, Cao J. VAEMax: Open-set intrusion detection based on OpenMax and variational autoencoder. In: 2024 5th Information Communication Technologies Conference (ICTC). IEEE; 2024. p. 98–105.
[10] Wang Z, Xu Q, Yang Z, He Y, Cao X, Huang Q. OpenAUC: Towards AUC-oriented open-set recognition. Advances in Neural Information Processing Systems. 2022;35:25033–25045.
[11] Palechor A, Bhoumik A, Günther M. Large-scale open-set classification protocols for ImageNet. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2023. p. 42–51.
[12] Mundt M, Pliushch I, Majumder S, Hong Y, Ramesh V. Unified probabilistic deep continual learning through generative replay and open set recognition. Journal of Imaging. 2022;8(4):93.

[13] Tien-Wen Sung, Zheng-Jiang Xiao, Qingjun Fan, You-Te Lu, Thi-Minh-Phuong Ha. Super-resolution of UAV images based on improved Generative Adversarial Networks. Journal of Network Intelligence. 2025;10(4): 2003–2019,
[14] Hevner AR, March ST, Park J, Ram S. Design science in information systems research. MIS Q. 2004;28(1):75–105.
[15] Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S. A design science research methodology for information systems research. J Manag Inf Syst. 2007;24(3):45–77.
[16] Zolbanin H, Aubert B. A process model for design-oriented machine learning research in information systems. J Strateg Inf Syst. 2025;34(1):101868.
[17] Bartumeus F, Garriga J, Falk M, Mosquito Alert Expert Community. AI Mosquito Alert Challenge Dataset 2023 (Version v1) [Internet]. Zenodo; 2025 [cited 2026 Jan 16]. Available from: https://doi.org/10.5281/zenodo.15063886
[18] Salma AB, Gannavarapu GM, Jameer S, Jempani VYV, Chappidi J, Ruqsar Zaitoon, et al. Deep learning-driven mosquito species identification using YOLOv5 for disease monitoring and control. J Theor Appl Inf Technol. 2025;103(4):1347–1348.
[19] Park J, Kim DI, Choi B, Kang W, Kwon HW. Classification and morphological analysis of vector mosquitoes using deep convolutional neural networks. Sci Rep. 2020;10(1):1012.
[20] Gregor S, Hevner AR. Positioning and presenting design science research for maximum impact. MIS Q. 2013;37(2):337–355.