Scientific Data

https://doi.org/10.1038/s41597-026-07713-2

Article in Press

Tephritid26: A standardized, multi-angle image dataset of quarantine-significant true fruit flies for deep learning-based identification

Received: 9 February 2026

Accepted: 22 June 2026

Published online: 27 June 2026

Cite this article as: Li, Z., Wang, X., Wu, Z. *et al.* Tephritid26: A standardized, multi-angle image dataset of quarantine-significant true fruit flies for deep learning-based identification. *Sci Data* (2026). https://doi.org/10.1038/s41597-026-07713-2

Zitao Li, Xingkai Wang, Zhuojie Wu, Mengyuan Yong, Marc De Meyer, Kostas Bourtzis, Bingbing Wei, Tianyu Zheng, Qian Zeng, Jin Mo, Ruosi Liu, Agus Susanto, Weiqi Liu, Wenchao Guo, Xinhua Ding, Xiaolei Huang, Ding Yang, Daifeng Cheng, Xin Yu, Fan Jiang & Xuankun Li

Tephritid26: A standardized, multi-angle image dataset of quarantine-significant true fruit flies for deep learning-based identification

Zitao Li[1, 2], Xingkai Wang[1, 2], Zhuojie Wu[3], Mengyuan Yong[2], Marc De Meyer[4], Kostas Bourtzis[5], Bingbing Wei[6], Tianyu Zheng[1, 2], Qian Zeng[7], Jin Mo[8], Ruosi Liu[9], Agus Susanto[10], Weiqi Liu[11], Wenchao Guo[12], Xinhua Ding[12], Xiaolei Huang[13], Ding Yang[1, 2], Daifeng Cheng[14], Xin Yu[3], Fan Jiang[15]*, Xuankun Li[1, 2]*

[1.] *State Key Laboratory of Agricultural and Forestry Biosecurity, College of Plant Protection, China Agricultural University, Beijing 100193, China*

[2] *Department of Entomology, College of Plant Protection, China Agricultural University, Beijing 100193, China*

[3] *School of Electrical Engineering and Computer Science, The University of Queensland St Lucia, QLD 4072, Australia*

[4] *Royal Museum for Central Africa, Invertebrates Section and JEMU, Tervuren B3080, Belgium*

[5] *Insect Pest Control Section, Joint FAO/IAEA Centre of Nuclear Techniques in Food and Agriculture, Department of Nuclear Sciences and Applications, International Atomic Energy Agency, P.O. Box 100, 1400 Vienna, Austria*

[6] *College of Plant Science & Technology, Huazhong Agricultural University, Wuhan 430070, China*

[7] *Technology Center of Mengla Customs, Mengla 666300, China*

[8] *Technology Center of Changsha Customs District, Changsha 410004, China*

[9] *Science and Technology Research Center of China Customs, Beijing 100026, China*

[10] *Pest Laboratory, Department of Plant Pest and Disease, Faculty of Agriculture, Universitas Padjadjaran, Ir. Soekarno Street Km. 21, Jatinangor, Sumedang Regency 45363, Indonesia*

[11] *Technology Center of Manzhouli Customs District, Manzhouli 021400, China*

[12] *Key Laboratory of Integrated Pest Management on Crops in Northwestern Oasis, Ministry of Agriculture and Rural Affairs, National Plant Protection Scientific Observation and Experiment Station of Korla, Xinjiang Key Laboratory of Agricultural Biosafety, Institute of Plant Protection, Xinjiang Uygur Autonomous Region Academy of Agricultural Sciences, Urumqi 830091, China*

[13] *State Key Laboratory of Agricultural and Forestry Biosecurity, College of Plant Protection, Fujian Agriculture and Forestry University, Fuzhou 350002, China*

[14] *Department of Entomology, South China Agricultural University, Guangzhou 510640, China*

[15] *Institute of Plant Inspection and Quarantine, Chinese Academy of Quality and Inspection & Testing, Beijing 100176, China*

Corresponding author:

Xuankun Li: xuankun.li@cau.edu.cn

Fan Jiang: f-jiang@outlook.com

Abstract

Accurate and rapid identification of quarantine-significant tephritids is critical to global agricultural biosecurity, but the application of deep learning is limited by the lack of large public image datasets. We present Tephritid26, a multi-angle image dataset of 26 tephritid species to address this gap. The dataset includes 38,081 images from 1,473 specimens across seven genera and two subfamilies, assembled through a global collaborative effort to source these regulated species. Specimens were mounted using a novel protocol combining varied thoracic attachment points and pin angles, and a rotational imaging setup then systematically captured each specimen from multiple perspectives to mimic real inspection conditions. The dataset is formatted for machine learning workflows. To demonstrate its utility, we trained deep learning models for species identification. ResNet-50, ConvNeXt-B, Vit-Small and Swin-Tiny all attained high species-level accuracy (Macro-Averaged F1-score > 96.75). Gradient-weighted Class Activation Mapping confirmed that the models focused on taxonomically informative morphological regions. This dataset serves as a benchmark for developing automated identification tools in phytosanitary applications.

Keywords: dataset; deep learning; multi-angle image; quarantine; true fruit fly; insect identification

Background & Summary

Tephritidae (true fruit flies) include numerous polyphagous species of major quarantine significance[1,2], whose accidental introduction can cause severe agricultural losses and trigger costly trade restrictions[3-5]. Rapid and accurate species identification at points of entry is therefore a cornerstone of global agricultural biosecurity[6]. While molecular methods provide high specificity[7,8], morphological identification remains indispensable due to its low cost, rapid turnaround when adult specimens are intercepted, and minimal equipment requirements, which is especially critical in resource-limited settings[9,10]. However, this approach relies on specialized expertise[11,12]. Accurate diagnosis demands a deep understanding of taxonomic terminology, the ability to discern subtle morphological characters, and practical skill in using identification keys[13,14].

Deep learning offers a promising avenue to augment morphological identification[15-17], yet progress is hampered by the scarcity of large-scale, public image datasets suitable for model training[18]. Existing resources are often limited to single, standardized views per species, which do not reflect the variable orientations and conditions encountered during real inspections[19-21]. This gap constrains the development of robust, automated tools that can support non-specialist personnel.

To address this, we present Tephritid26, one of the largest multi-angle image datasets of insect specimens created to date (Tables 1). Compared to other tephritid species dataset for deep learning research, this standardized, multi-angle image dataset includes 38,081 images of 26 quarantine-significant tephritid species across seven genera (Tables 2 and 3). Assembling such a dataset is inherently challenging, as these quarantine-significant species cannot be collected from a single geographic source (Table 3). Its creation therefore required a coordinated, global effort to obtain and share specimens. To maximize the coverage of observable angles from each specimen, we employed a novel mounting protocol that combines varied thoracic attachment points with a range of pin angles. Each mounted specimen was then systematically imaged through a full 360° rotation. This process generated 38,081 high-resolution images derived from 1,473 specimens (Table 3), comprehensively mimicking the diverse viewing-angles of a manual inspection.

By providing a comprehensive visual resource that captures the full range of specimen orientations, Tephritid26 is designed to accelerate the development of deep learning models for automated tephritid identification. We anticipate this dataset will serve as a foundational benchmark for researchers and practitioners working towards strengthening phytosanitary diagnostics through artificial intelligence.

Table 1. Comparison of existing insects benchmark datasets. (*) Multi Angle: each specimen is photographed from more than 5 different angles. N, not included; Y, included.

Dataset	Year	No. of Species	Domain	Type	Multi Angle*	Images / Text
Samanta et al.[22]	2012	8	Tea Insect Pests	Text	N	602
Venugoban et al.[23]	2014	20	Peddy Field Pests	Ecological Images	Y	200
Xie et al.[24]	2015	24	Field Crop Pests	Ecological Images	Y	1,440
Liu et al.[25]	2016	12	Peddy Field Pests	Ecological Images	Y	5,136
Deng et al.[26]	2018	10	Tea Insect Pests	Ecological Images	Y	563

PestNet[27]	2019	16	General Agricultural Pests	Trap Images	Y	88,670
IP102[28]	2019	102	General Agricultural Pests	Ecological Images	Y	75,222
AgriPest[29]	2021	14	General Agricultural Pests	Ecological Images	Y	49,707
Popkov et al.[19]	2022	26	Miridae Species	Specimen Images	N	3,792
Zhao et al.[30]	2022	17	Culicidae Species	Specimen Images	Y	9,900
Simović et al.[31]	2024	90	EPT Species	Specimen Images	N	16,650
HIVE[32]	2025	13,434	General Pests	Ecological Images	Y	2,292,900
Wang et al.[33]	2026	92	Reduviidae Species	Ecological and Specimen Images	Y	11,915
Li & Li[34]	2026	11	Pest Moth Species	Specimen Images	Y	15,349
Tephritid26 (present study)	**2026**	**26**	**Qurantine Tephritidae Species**	**Specimen Images**	**Y**	**38,081**

Table 2. Comparison of existing tephritid species classification research with deep learning approaches. (*) No species level.

Research	Year	Genus	Species	Images
Leonardo et al.[35]	2018	1	3	301
Shen et al.[36]	2021	4	*	5,287
Wang et al.[37]	2021	8	83	2,083
Salifu et al.[38]	2022	1	7	1,091
Tannous et al.[39]	2023	2	2	1,826
Bustamante Flores et al.[40]	2025	2	2	472
Ward et al.[41]	2025	12	34	1,380
Tephritid26 (present study)	**2026**	**7**	**26**	**38,081**

Methods

Specimen sourcing and identification. Specimens of 26 quarantine-significant tephritid species were sourced from three channels (Table 3): (1) field surveillance using fruit-baited traps, (2) interceptions at international ports of entry, and (3) insectary-reared colonies provided by the Insect Pest Control Laboratory of the International Atomic Energy Agency (IAEA). The complete species list with corresponding sources is provided in Table 3. All species identifications were verified through morphological examination by the corresponding author, F. Jiang[42,43], with voucher specimens deposited in the Entomological Museum of China Agricultural University (CAU), Beijing, China.

Specimen mounting protocol. To systematically generate a comprehensive range of viewing angles for imaging, we developed a multi-angle mounting protocol that departs from the standard practice (lateral adhesion with the body axis perpendicular to the pin). Each specimen was prepared in a set of 15 distinct posture-orientation combinations. This was achieved by adhering specimens to insect pins at three different thoracic positions: dorsal (mesonotum), lateral (right mesopleuron), and ventral (sternum). For each adhesion point, the pin was fixed at one of five angles relative to the specimen's body axis: (1) perpendicular (conventional), (2) tilted upward at 45°, (3) tilted downward at 45°, (4) parallel with the head oriented upward, and (5) parallel with the head oriented downward (Figure 1).

Image acquisition system. Pinned specimens were centered on a rotary stage with the camera aligned parallel to the rotary stage's plane. Photos were taken using two digital single-lens reflex (DSLR) camera systems under controlled lighting: a Canon EOS 5D Mark III camera with a Canon MP-E 65mm f/2.8 1-5X lens, and a Nikon Z 8 camera with a NIKKOR Z Micro 105mm f/2.8 VR S lens. All images were acquired under controlled diffuse omnidirectional lighting with a clean, uniform background. The distance between lens and the specimen was adjusted to keep most of the specimen in focus. No focus stacking was applied. Two camera systems were used interchangeably across the entire dataset to increase variability. To achieve sufficient depth of field, we used an aperture of f/16 for all species except *Bactrocera thailandica*, *Carpomya vesuviana*, *Zeugodacus diaphorus*, and a subset of *Bactrocera oleae* and *Ceratitis cosyra*, which were imaged at f/7.1. Under stable lighting conditions, the ISO and shutter speed were adjusted accordingly to ensure proper exposure.

Image preprocessing and dataset assembly. The raw image collection was processed to create the final deep learning-ready dataset. First, the primary subject (the tephritid) in each image was localized and cropped to minimize background. This was performed using a YOLO11 object detection model (https://github.com/ultralytics/ultralytics) initially trained on a manually annotated subset of 50 images per species, which were randomly sampled from Tephritid26. The automated cropping was reviewed and refined using the annotation tool X-AnyLabeling (https://github.com/CVHub520/X-AnyLabeling). After automated cropping, approximately 1,000 images were manually corrected by adjusting the bounding boxes to fully enclose the specimen pixels. The final Tephritid26 dataset was constructed by partitioning the images at the specimen level to prevent data leakage: all images derived from a single physical specimen were assigned exclusively to one of the three subsets—training (80%), validation (10%), or test (10%). For *Bactrocera latifron*, which has only one available specimen, the images were split into three subsets randomly.

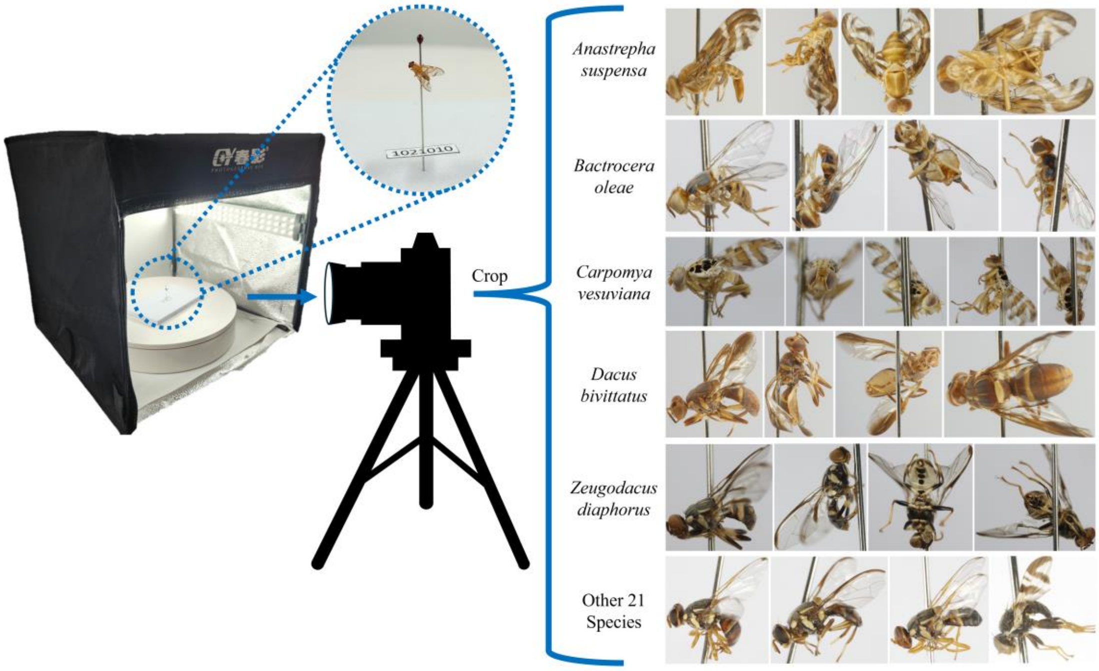

Figure 1 Outline of images collection of quarantine-significant tephritid species.

Data Record

Tephritid26 dataset could be accessed via Dryad (https://doi.org/10.5061/dryad.gf1vhhn36[44]). Original images and bounding box labels for Yolo11 and X-AnyLabeling were provided. The image files were named according to the following convention: "02+SpeciesID + IndividualID+1111+ SequentialID.JPG".

Table 3. Composition of Tephritid26. Species list with corresponding specimen and image totals. Serial numbers are employed for the renaming of image files.

Species	**SpecieID**	**Source**	**Image/Specimen**	**Train**	**Val**	**Test**
Anastrepha fraterculus	1019	IAEA: Argentina	2122/66	1657/52	228/7	237/7
Anastrepha ludens	1020	IAEA: Guatemala; Mexico	2259/98	1773/78	223/10	263/10
Anastrepha obliqua	1021	Reared By IAEA	2337/98	1866/78	242/10	229/10

Anastrepha suspensa	1031	IAEA: Mexico	621/17	467/13	77/2	77/2
Bactrocera correcta	1001	China: Yunnan	1313/57	1021/45	142/6	150/6
Bactrocera dorsalis	1003	China: Guangdong	2380/93	1973/74	193/9	214/10
Bactrocera latifrons	1005	China: Yunnan	4604/6	19/1	2/0	3/0
Bactrocera minax	1013	China: Hubei	2613/117	2083/93	261/12	269/12
Bactrocera oleae	1022	IAEA: Greece	766/25	608/20	66/2	92/3
Bactrocera thailandica	1029	China: Yunnan	864/28	678/22	90/3	96/3
Bactrocera tryoni	1006	IAEA: Australia	996/46	790/36	104/5	102/5
Bactrocera tuberculata	1007	China: Yunnan	784/25	629/20	72/2	83/3
Bactrocera umbrosa	1027	Indonesia: West Java	1391/42	1121/33	113/4	157/5
Bactrocera zonata	1002	Reared By IAEA	2112/98	1700/78	211/10	201/10
Carpomya vesuviana	1024	China: Xinjiang	1097/34	868/27	97/3	132/4
Ceratitis capitata	1018	Reared By IAEA	2443/102	1925/81	267/10	251/11
Ceratitis cosyra	1030	Mozambique : Cuamba; Mphingwe	709/25	563/20	59/2	87/3
Dacus bivittatus	1032	Mozambique : Mphingwe	950/22	755/17	70/2	125/3
Dacus punctatifrons	1034	Mozambique : Mphingwe	1145/19	921/15	126/2	98/2
Dacus trimacula	1025	China: Beijing	1005/36	781/28	109/4	115/4
Rhagoletis cerasi	1026	China: Inner Mongolia	811/34	646/27	71/3	94/4

Zeugodacus cilifer	1009	China: Yunnan	1320/67	1043/53	134/7	143/7
Zeugodacus cucurbitae	1010	China: Guangdong	1954/87	1538/69	209/9	207/9
Zeugodacus diaphorus	1028	China: Sichuan	1150/37	900/29	121/4	129/4
Zeugodacus scutellatus	1012	China: Sichuan	2595/90	2068/72	259/9	268/9
Zeugodacus tau	1011	China: Sichuan	2320/109	1865/87	228/11	227/11
Total			**38081/1473**	**30258/1168**	**3774/148**	**4049/157**

Technical Validation

Taxonomic and procedural controls. All specimen identifications were verified by expert taxonomists prior to imaging, and voucher specimens are permanently deposited in a curated collection (CAU). The image acquisition and preprocessing pipeline was designed to ensure consistency: lighting and camera settings were fixed, and the automated cropping step was uniformly applied to all images to minimize non-biological variance.

Benchmark classification performance. To validate that the Tephritid26 dataset supports accurate species-level identification, we established a standard classification benchmark using two widely adopted convolutional neural network (ResNet-50[45] and ConvNeXt-B[46]) and two Transformer network (Swin-Tiny[47] and Vit-Small[48]). This benchmark is intended to demonstrate the dataset's baseline utility, not to optimize model performance. Models were initialized with ImageNet-1K pretrained weights. During training, validation, and test, images underwent the same deterministic geometry: they were isotropically resized so that the shorter side equaled 224 pixels (preserving aspect ratio) and then center-cropped to 224×224 without anisotropic stretching. For non-square originals, only peripheral regions may fall outside the crop. Training also applied random horizontal flipping, in-plane rotation (±15°), and color jitter (brightness, contrast, and saturation: 0.2; hue: 0.05 when targeted augmentation was enabled). With targeted augmentation, training further included Gaussian blur (kernel size 5, σ uniformly sampled from 0.15–1.2, applied with probability 0.3), random autocontrast (probability 0.15), random sharpness adjustment (sharpness factor 1.8, probability 0.15), additive Gaussian noise on the [0, 1] tensor (standard deviation 0.02), and random erasing (probability 0.1, erasing area fraction 2–12% of the image, aspect ratio 0.3–3.3). All tensors were finally normalized with ImageNet mean [0.485, 0.456, 0.406] and standard deviation [0.229, 0.224, 0.225][49]. Validation and test used only resize, center crop, conversion to tensor, and the same normalization. All four models were trained for 20 epochs with a learning rate of 0.0005 and a batch size of 32.

Quantitative evaluation was based on standard multi-class classification metrics: accuracy, macro-averaged precision, macro-averaged recall, and macro-averaged F1-score. All four models achieved high performance on the held-out test set (Tables 4 and 5), with ConvNeXt-B attaining top-1 accuracy of 98.22%. The high accuracy confirms that the dataset contains sufficient and discriminative visual information for deep learning models to learn species-specific features. Favorable results from the confusion matrix (Figure 2), along with a smoother decline in the loss curve (Figure 3), indicate that all four distinct models performed well during training on this dataset, attesting to its usability and high quality.

Table 4. Benchmark classification performance on the Tephritid26 test set. Accuracy (A), Macro-Averaged Precision (P), Macro-Averaged Recall (R), and Macro-Averaged F1-score (F) are expressed as percentages (%) among four models on the Tephritid26 dataset.

Model	**A**	**P**	**R**	**F**
ConvNeXt-B[38]	98.22%	98.64%	98.61%	98.61%
ResNet-50[37]	97.21%	97.60%	97.39%	97.46%
Swin-Tiny[39]	97.78%	98.11%	98.02%	98.04%
Vit-Small[40]	97.18%	96.30%	97.67%	96.75%

Table 5. Comparison of per-class F1-scores, precision and recall on the Tephritid26 test set

(ConvNeXt-B). Macro-Averaged F1-score (F), Macro-Averaged Precision (P), and Macro-Averaged Recall (R) are expressed as percentages (%)

Species	F	P	R
Anastrepha fraterculus	93.17%	91.46%	94.94%
Anastrepha ludens	97.19%	95.93%	98.48%
Anastrepha obliqua	90.50%	93.90%	87.34%
Anastrepha suspensa	100.00%	100.00%	100.00%
Bactrocera correcta	95.83%	100.00%	92.00%
Bactrocera dorsalis	97.49%	95.11%	100.00%
Bactrocera latifrons	100.00%	100.00%	100.00%
Bactrocera minax	99.08%	98.18%	100.00%
Bactrocera oleae	100.00%	100.00%	100.00%
Bactrocera thailandica	98.97%	97.96%	100.00%
Bactrocera tryoni	99.51%	100.00%	99.02%
Bactrocera tuberculata	99.40%	98.81%	100.00%
Bactrocera umbrosa	100.00%	100.00%	100.00%
Bactrocera zonata	100.00%	100.00%	100.00%
Carpomya vesuviana	98.86%	99.24%	98.48%
Ceratitis capitata	100.00%	100.00%	100.00%
Ceratitis cosyra	98.29%	97.73%	98.85%
Dacus bivittatus	99.60%	100.00%	99.20%
Dacus punctatifrons	99.49%	98.99%	100.00%
Dacus trimacula	100.00%	100.00%	100.00%
Rhagoletis cerasi	100.00%	100.00%	100.00%
Zeugodacus cilifer	98.96%	97.95%	100.00%
Zeugodacus cucurbitae	99.76%	99.52%	100.00%
Zeugodacus diaphorus	99.61%	100.00%	99.22%
Zeugodacus scutellatus	99.25%	100.00%	98.51%
Zeugodacus tau	98.89%	100.00%	97.80%

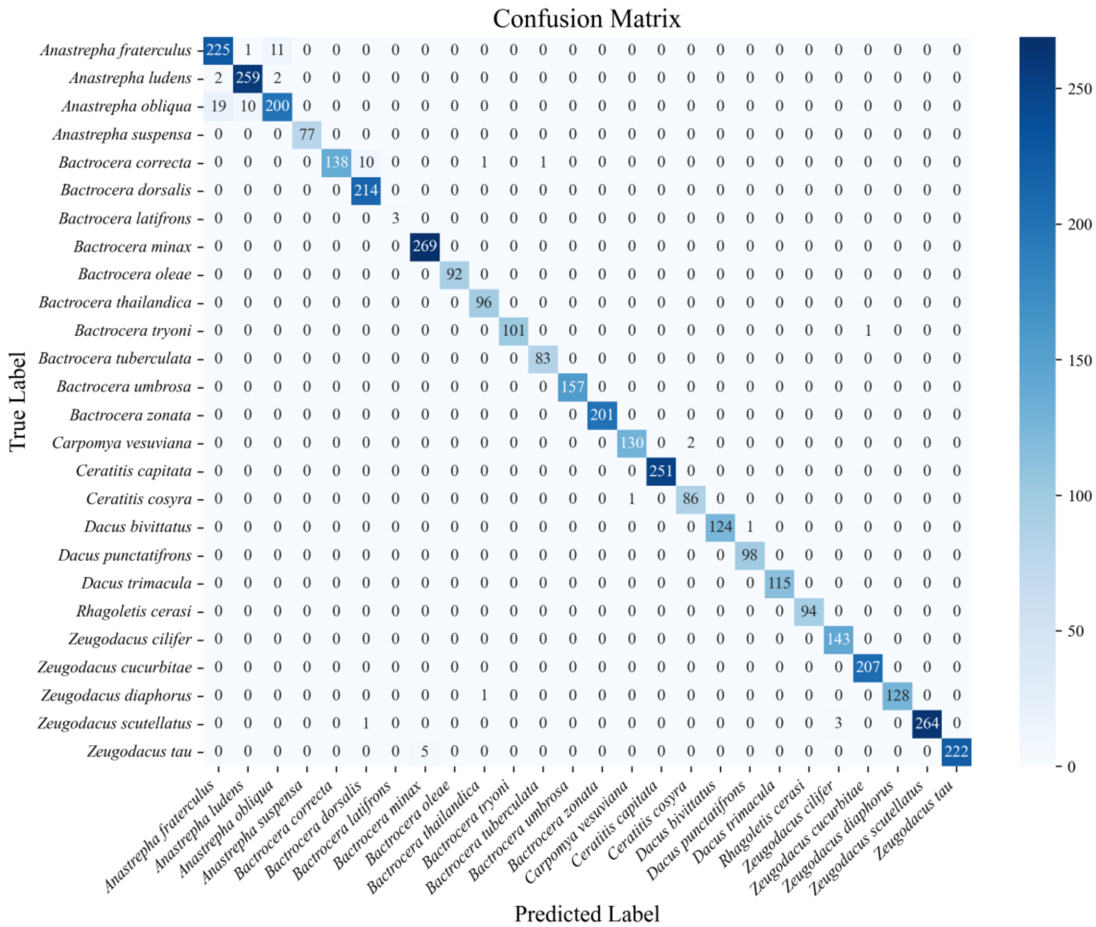

Figure 2. Confusion matrix of trained ConvNeXt-B model test-prediction.

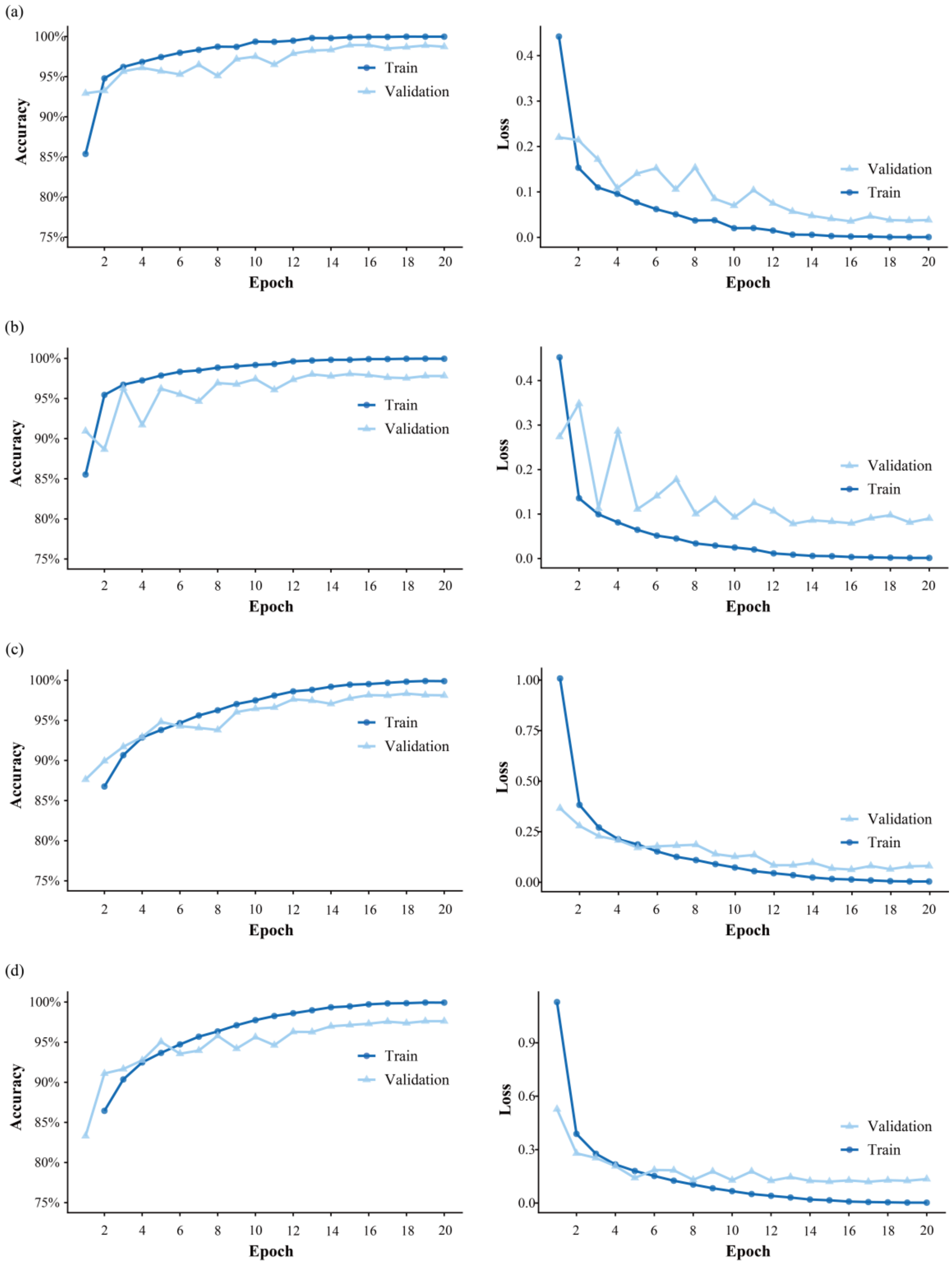

Figure 3. Model accuracy and loss curves of four trained backbone network models. (a) ConvNeXt-B; (b) ResNet-50; (c) Swin-Tiny; (d) Vit-Small.

Visual explanation of learned features. To ensure that the high classification accuracy stems from biologically meaningful features rather than spurious correlations, we employed Gradient-weighted Class Activation Mapping (Grad-CAM)[50], targeting the final depthwise convolutional layer in the last block of the ConvNeXt architecture. This technique visualizes the image regions most influential for the model's predictions. As shown in Figure 4, the models consistently attended to taxonomically diagnostic morphological structures, such as wing patterns, thorax markings, and abdominal features, across different viewing angles generated by our mounting protocol[42,43]. This confirms that the multi-angle imaging strategy successfully captures consistent, discriminative morphological information, and that models trained on Tephritid26 learn biologically plausible representations.

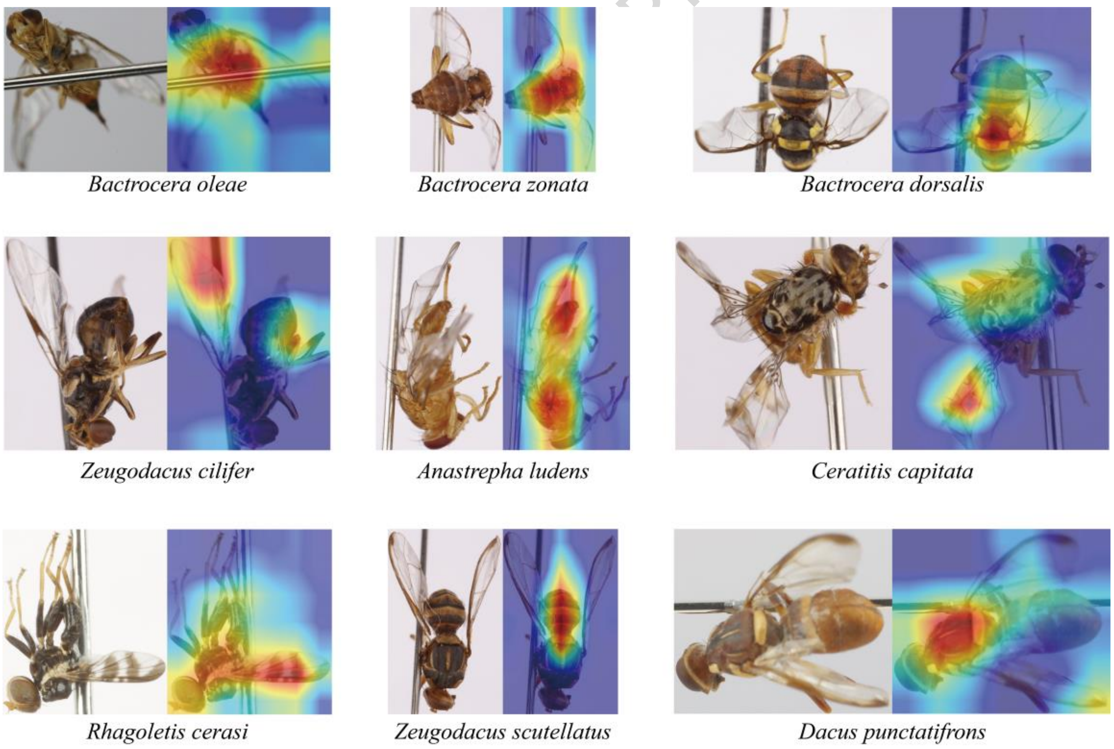

Figure 4. Grad-CAM analysis of trained model using the backbone network ConvNeXt-B. Warmer colors (red) indicate image regions that most strongly drive the classification decision, whereas cooler colors (blue) represent areas with minimal influence on the prediction.

The combined results from high benchmark accuracy and biologically grounded attention maps provide strong technical validation for the Tephritid26 dataset. They confirm its quality, completeness, and suitability for developing and benchmarking deep learning models aimed at the automated, image-based identification of quarantine-significant tephritids.

Data availability

Tephritid26 is deposited at Dryad (https://doi.org/10.5061/dryad.gf1vhhn36[44]).

Usage Notes

The dataset exhibits variability in sample counts across species (ranging from one specimen and 24 images to 117 specimens and 2,613 images per species). This imbalance may bias model performance toward dominant classes; mitigation strategies such as class-weighted loss functions, oversampling, or synthetic data augmentation (e.g., Generative Adversarial Networks) are recommended during training. For *Bactrocera latifron*, which represented by a single specimen, test images originate from the same individual as training images. Therefore, performance metrics reflect within-specimen repeatability cross-specimen generalization, and results for such rare species should be interpreted with this in mind. All images were captured under highly standardized laboratory conditions (controlled lighting, uniform background, fixed camera setup). This matches the controlled laboratory environment typical of phytosanitary specimen examination. However, models trained on these specimen-based images along may not directly generalize to uncontrolled field photographs; users targeting field applications should consider domain adaptation or supplementary field validation.

Code availability

The code of the study can be found at: https://github.com/lizitao2005/TrueFruitFly_Classification. The repository includes python files for models training, predicting and Grad-CAM analysis. For detailed instruction please read the documentation provided in the Github repository.

References

1 Badii, K. *et al.* Species composition and host range of fruit-infesting flies (Diptera: Tephritidae) in northern Ghana. *Int. J. Trop. Insect Sci.* **35**, 137-151 (2015).

2 Liquido, N. J. *et al.* Host plants of Mediterranean fruit fly (Diptera: Tephritidae) on the Island of Hawaii (1949-1985 survey). *J. Econ. Entomol.* **83**, 1863-1878 (1990).

3 Li, Z. *Prevention and Control of Biological Invasions: Potential Loss of Economically Important Fruit Flies* 1st edn (China Agricultural University Press, 2022).

4 Qin, Y. *et al.* Prediction of potential economic impact of *Bactrocera zonata* (Diptera: Tephritidae) in China: Peaches as the example hosts. *J. Asia-Pac. Entomol.* **24**, 1101-1106 (2021).

5 Vayssières, J. *et al.* The mango tree in central and northern Benin: damage caused by fruit flies (Diptera Tephritidae) and computation of economic injury level. *Fruits* **64**, 207-220 (2009).

6 Armstrong, K. *et al.* Fruit fly (Diptera: Tephritidae) species identification: a rapid molecular diagnostic technique for quarantine application. *Bull. Entomol. Res.* **87**, 111-118 (1997).

7 Hodgetts, J. *et al.* DNA barcoding for biosecurity: case studies from the UK plant protection program. *Genome* **59**, 1033-1048 (2016).

8 Jinbo, U. *et al.* Current progress in DNA barcoding and future implications for entomology. *Entomol. Sci.* **14**, 107-124 (2011).

9 Balmès, V. & Mouttet, R. Development and validation of a simplified morphological identification key for larvae of tephritid species most commonly intercepted at import in Europe. *EPPO Bull.* **47**, 91-99 (2017).

10 Zhang, G.-N. *et al.* Morphological characterization and distribution of sensilla on maxillary palpi of six *Bactrocera* fruit flies (Diptera: Tephritidae). *Florida Entomologist* **94**, 379-388 (2011).

11 Engel, M. S. *et al. The taxonomic impediment: a shortage of taxonomists, not the lack of technical approaches* Vol. 193 (Oxford University Press, 2021)

12 Hockland, S. Dilemmas for the quarantine diagnostician and taxonomist in the 21st century. in *Proceedings of the Fourth International Congress of Nematology, 8-13 June 2002, Tenerife, Spain.* 115-123 (2004).

13 Australia, P. H. *The Australian Handbook for the Identification of Fruit Flies*. Version 3.1 edn, (Plant Health Australia, 2018).

14 Foote, R. H. *et al. Handbook of the fruit flies (Diptera: Tephritidae) of America north of Mexico*. (Cornell University Press, 2020).

15 Ahmad, I. *et al.* Deep learning based detector YOLOv5 for identifying insect pests. *Applied Sciences* **12**, 10167 (2022).

16 Chakrabarty, S. *et al.* Application of artificial intelligence in insect pest identification-A review. *Artif. Intell. Agric.* **16**, 44-61 (2026).

17 Ejaz, M. H. *et al.* Crop pest classification using deep learning techniques: a review. Preprint at https://doi.org/10.48550/arXiv.2507.01494 (2025).

18 Teixeira, A. C. *et al.* A systematic review on automatic insect detection using deep learning. *Agriculture* **13**, 713 (2023).

19 Popkov, A. *et al.* Machine learning for expert-level image-based identification of very similar species in the hyperdiverse plant bug family Miridae (Hemiptera: Heteroptera). *Syst. Entomol.* **47**, 487-503 (2022).

20 David, K. J. *et al.* Taxonomic notes on the genus *Campiglossa* Rondani (Diptera, Tephritidae, Tephritinae, Tephritini) in India, with description of three new species. *ZooKeys* **977**, 75 (2020).

21 Singh, S. K. *et al.* An updated checklist for fruit flies (Diptera: Tephritidae) of India with new species and new distribution records. *Zootaxa* **5607**, 1-71 (2025).

22 Samanta, R. & Ghosh, I. Tea insect pests classification based on artificial neural networks. *Int. J. Comput. Eng. Res.* **2**, 1-13 (2012).

23 Venugoban, K. & Ramanan, A. Image classification of paddy field insect pests using

gradient-based features. *Int. J. Mach. Learn. Comput.* **4**, 1 (2014).

24 Xie, C. *et al.* Automatic classification for field crop insects via multiple-task sparse representation and multiple-kernel learning. *Comput. Electron. Agric.* **119**, 123-132 (2015).

25 Liu, Z. *et al.* Localization and classification of paddy field pests using a saliency map and deep convolutional neural network. *Sci. Rep.* **6**, 20410 (2016).

26 Deng, L. *et al.* Research on insect pest image detection and recognition based on bio-inspired methods. *Biosyst. Eng.* **169**, 139-148 (2018).

27 Liu, L. *et al.* PestNet: An end-to-end deep learning approach for large-scale multi-class pest detection and classification. *IEEE Access* **7**, 45301-45312 (2019).

28 Wu, X. *et al.* Ip102: A large-scale benchmark dataset for insect pest recognition. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 8787-8796 (2019).

29 Wang, R. *et al.* Agripest: A large-scale domain-specific benchmark dataset for practical agricultural pest detection in the wild. *Sensors* **21**, 1601 (2021).

30 Zhao, D. *et al*. A Swin Transformer-based model for mosquito species identification. *Sci. Rep.* **12**, 18664 (2022).

31 Simović, P. *et al.* Automated identification of aquatic insects: A case study using deep learning and computer vision techniques. *Sci. Total Environ.* **935**, 172877 (2024).

32 Kim, M. *et al.* HIVE: Hierarchical Identification in Visual Entomology for Pest Recognition. Preprint at https://dx.doi.org/10.2139/ssrn.5764792 (2025).

33 Wang, X. *et al*. Multi-angle, cross-domain fusion strategy enhances automated insect identification and hierarchical categorization: a case study on assassin bugs (Hemiptera: Reduviidae). *Cladistics*. https://doi.org/10.1111/cla.70029 (2026).

34 Li, Z. & Li, X. Deep Learning-Based Image Classification of Pupae from 11 Lepidoptera Pest Species. *Insects* **17**, 327 (2026).

35 Leonardo, M. M. *et al*. Deep feature-based classifiers for fruit fly identification (Diptera:

Tephritidae). In *2018 31st SIBGRAPI conference on graphics, patterns and images*. 41-47. (2018).

36. Shen, Y. *et al*. Systematics of Tephritid Fruit Flies: A Machine Learning Based Pest Identification System. *Proceedings* **68**, 0. (2021).

37. Wang, J. *et al*. An intelligent identification system combining image and DNA sequence methods for fruit flies with economic importance (Diptera: Tephritidae). *Pest Manag. Sci.* **77**, 3382-3395. (2021).

38. Salifu, D. *et al*. Leveraging machine learning tools and algorithms for analysis of fruit fly morphometrics. *Sci. Rep.* **12**, 7208. (2022).

39. Tannous, M. *et al*. A deep-learning-based detection approach for the identification of insect species of economic importance. *Insects* **14**, 148. (2023).

40. Bustamante Flores, E. A. *et al*. Fruit Fly Classification (Diptera: Tephritidae) in Images, Applying Transfer Learning. In *Mexican Congress on Artificial Intelligence* 85-99. (2025).

41. Ward, D. *et al*. A deep learning classification pipeline for identifying economically important tephritid fruit flies. Preprint at https://doi.org/10.21203/rs.3.rs-7802967/v1 (2025).

42 Drew, R. & Romig, M. *Tropical Fruit Flies of South-East Asia: Indomalaya to North-West Australia*. (CABI, 2013).

43 White, I. M. & Elson-Harris, M. M. *Fruit flies of economic significance: their identification and bionomics*. (CABI, 1992).

44 Li, Z. *et al.* Tephritid26. *Dryad* https://doi.org/10.5061/dryad.gf1vhhn36 (2026).

45 He, K. *et al.* Deep residual learning for image recognition. in *Proceedings of the IEEE conference on computer vision and pattern recognition.* 770-778 (2016).

46 Liu, Z. *et al.* A convnet for the 2020s. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 11976-11986 (2022).

47 Liu, Z. *et al.* Swin transformer: Hierarchical vision transformer using shifted windows. in *Proceedings of the IEEE/CVF international conference on computer vision.* 10012-10022

(2021).

48 Dosovitskiy, A. *et al*. An image is worth 16x16 words: Transformers for image recognition at scale. Preprint at https://doi.org/10.48550/arXiv.2010.11929 (2020).

49 Imambi, S., Prakash, K. B., & Kanagachidambaresan, G. R. PyTorch. In *Programming with TensorFlow: solution for edge computing applications*. 87-104 (2021).

50 Selvaraju, R. R. *et al.* Grad-cam: Visual explanations from deep networks via gradient-based localization. in *Proceedings of the IEEE international conference on computer vision.* 618-626 (2017).

Author contributions

Zitao Li completed the data collecting, model training, result analyzing and article writing; Xuankun Li, Xin Yu, Ding Yang and Zitao Li designed this study; Xingkai Wang and Zhuojie Wu guided the completion of the code; Fan Jiang, Daifeng Cheng, Qian Zeng, Jin Mo, Ruosi Liu, Agus Susanto, Weiqi Liu, Wenchao Guo and Xinhua Ding provided and identified the specimens of fruit flies; Mengyuan You and Tianyu Zheng participated in photographing of fruit flies; Fan Jiang, Kostas Bourtzis, Marc De Meyer, Xiaolei Huang and Bingbing Wei provided helpful comments on the manuscript. All authors have read the manuscript.

Competing Interests

All authors declare no competing interests.

Acknowledgements

We sincerely thank Ruiqing Dong, Haoyue Zhou and Zhiyu Zhang (China Agricultural University) and Yihang Li (University of Memphis) for specimen mounting and imaging. We thank Massimiliano Virgilio (Royal Museum for Central Africa), Wenzhao Yang (Agricultural and Rural Bureau of Xishan District, Kunming), Changying Niu (Huazhong Agricultural University), Luxi Liu (Technology Center of Chengdu Customs District), Zhihan Jiang (China Third Metallurgical Group Corporation LTD) and Jiaxin Liu for providing tephritid specimens. The authors acknowledge the China Agricultural University for providing high-performance computing platforms and support that contributed to the research results reported in this study.

Funding

This work was supported by the Joint Research Program of State Key Laboratory of Agricultural and Forestry Biosecurity (No. SKLJRP2509). Xuankun Li and Ding Yang was supported by the 2115 Talent Development Program of China Agricultural University. Fan Jiang was supported by the National Natural Science Foundation of China (No. 32500393).