Systematics of Tephritid Fruit Flies: A Machine Learning Based Pest Identification System †

Yefeng Shen [1]*, Md Zakir Hossain [1,2]*, Shafin Rahman [3] and Khandaker Asif Ahmed [2]*

1   The Australian National University (ANU), ACT 2601, Australia
2   The Commonwealth Scientific and Industrial Research Organisation (CSIRO), ACT 2601, Australia
3   North South University, Dhaka, Bangladesh
*   Correspondence: zakir.hossain@csiro.au (M.Z.H.); khandakerasif.ahmed@csiro.au (K.A.A.)
†   Presented at the 1st International Electronic Conference on Entomology (IECE 2021), 1–15 July 2021; Available online: https://iece.sciforum.net/.

Abstract: Tephritid fruit flies (Diptera:Tephritidae) are the major economically important agricultural pests around the world. Numerous control measures are undergoing to reduce their abundance. An efficient pest identification system is a prerequisite for such tasks. Typically, the classification/identification of different insect species is done based on either external body features or DNA barcoding. However, those approaches are time-consuming by nature, requiring expert knowledge in relevant fields. Several machine learning (ML) models have been successfully deployed in the field of systematics, but there is a lack of ML models for fruit fly species. This study aims to curate and validate a comprehensive tephritid image database and build ML models to automatically identify tephritids from mixed of tephritids and non-Tephritid dipteran flies, and classify four major genera of notorious tephritid flies, namely, Anastrepha, Ceratitis, Rhagoletis, and Bactrocera. The images of our experiment were collected from the iNaturalist database. The dataset is cleaned by removing uninformative images using a deep learning model (Inception-V3) and unsupervised k-mean clustering. Several state-of-the-art ML models are tested on the dataset, results in highest accuracy of 95.44% with EfficientNet-B0 model to identify tephritid flies. Moreover, the EfficientNet-B2 model achieved 89.65% accuracy for classifying representatives of the major tephritid genus and showed the potential to enhance the identification accuracy. Overall, this work of the systematics of harmful fruit flies can be transformed into a practical and effective detection tool and can be implemented easily with existing agricultural pest control systems.

Keywords: Tephritid fruit fly; Machine Learning; Artificial Intelligence; Insect systematics

1. Introduction

Fruit flies (Family:Tephritidae) are one of the most destructive agricultural pests around the world. Until 2018, over 4,000 species have been identified under this family and 350 among them has been considered as economically harmful[1]. Australia is one of the largest agricultural crop producers around the world and due to strong invasiveness of fruitflies, Australia is under threats of numerous invasive fruit flies. Till date, more than 300 fruitfly species have been recorded in Australia, within these several species of *Bactrocera* and *Ceratitis* genera are the most prominent, causing millions of dollar losses yearly [2,3]. For a sustainable development of agriculture, it is important to reduce the prevalence of fruitfly species, and several environmental friendly, benign biological control measures is being implemented, namely Sterile insect techniques (SIT)[4], Male Annilation Technique (MAT)[5], Integrated Pest Management (IPM)[6] etc. Systematics of fruitlfies is the steeping stone of such measures, which contribute to estimate pest species prevalance within target area and suggest effective control measure to adopt[7].

Traditionally, fruitfly systematics is done based on morphological and genotypic features. Morphological identification need careful collection and storage procedures, fol-

lowed by manually curation by expert taxonomists[1]. This procedure is time-consuming, error-prone and often leads to miss-identification of closely related species[8,9]. Genetic features showed potential to be used as a rapid and accurate identification tool for fruitfly systematics [10]. This process utilise a short section of DNA from a specific gene, especially cytochrome c oxidase I (COI or COX1), 16S rRNA or 18S rRNA genes, which serve as an unique "barcode" for the species[11]. Compared to morphological key based systematics, DNA barcoding is automatic, can resolve species complexities and can be exercised with low-amount of DNA. But DNA barcoding is often costly, need sophisticated instruments and molecular biology expertise for sample preparation. Using cost-efficient Machine learning (ML) models can overcome these disadvantages easily[12].

Several image-based ML models had been adopted in the field of insect systematics[13, 14]. Support vector machine (SVM) get better accuracy then traditional neural networks (NNs) for order level classification and substantial improvements in other feature extraction models further enhance the accuracy nearly 20% for a dataset comprising 24 insect classes [15,16]. However, NN and SVM along with other traditional ML algorithms (e.g. KNN and Naive Bayes) heavily rely on the quality of images and extracted features. Real-life images with complex backgrounds often create inevitable noises in the dataset, which resulted in poor efficacy in classifying insects [17]. Moreover, the dataset need heavily manual curation to standardise images. Principal component analysis (PCA) and scale-invariant feature transform (SIFT) performed well to some extent [15,18]. In contrast, advanced neural networks such as convolutional neural networks (CNN) has better tolerance for different systematics tasks[19,20]. It focus on the whole images rather than only on pre-defined low-level features and able to train with more deeper and fine-tune body features. While comparing with other ML models, multi-layerd CNN model outperformed others to classify insect classes [21]. Numerous convolution models had been proposed to classify lepidopteran insects, but most of them utilised simple, unified images to build the database [22–24]. Recently, computer-intensive deep convolutional neural network (DCNN) models showed outstanding performance for complexed background insect images and showed high potential to be used in insect pest systematics [20,25,26]. Even though numerous ML models have been developed in insect systematics, there is a scarcity of ML model for fruitflies and possible implementation in biological control measures. There is also huge gap for proper and large sized dataset to train the models effectively.

The aim of this study is to curate a comprehensive fruitfly image database from publicly available data-sources and develop a ML model to be used in the field of fruitfly systematics. The model will distinguish the major economically important agricultural pests, tephritids, from mixed of tephritids and non-tephritid dipteran flies and will classify among four genera of fruitflies. Based on the risk analysis of 180 economically important *Tephritids* in six countries (China, USA, South Africa, Argentina, Italy, and Australia)[27], we selected four most destructive genera namely, *Anastrepha*, *Ceratitis*, *Rhagoletis*, and *Bactrocera*. Images were collected from iNaturalist [1] and pre-processed with an Inception-V3 model and unsupervised k-mean clustering to remove uninformative and irrelevant images. We validated our dataset with three traditional ML classifiers, KNN, Naive Bayes and SVM, along with some state-of-the-art DCNN models including ResNets, EfficientNets, and ResNeSts. Our model can distinguish fruitflies from non-fruitfly dipteran fly, and further classify fruitflies into one of the four major genera. The implication of current study can be used on insect pest monitoring and servillance, as well as can be adopted for species-level classification works.

2. Materials and Methods:

The experimental design of the current study is to collect image datasets from publicly available data-repositories, build an automated method to extract general features of insects

---

[1] https://www.inaturalist.org

Table 1. Details of insect images: (a) FF2 and (b) TF4 databases

(a) FF2 dataset

Insect class	Pre-cleaning	Post-cleaning
Non-Tephritid flies	83,941	45,103
Tephritid fruitflies	5,649	5,287

(b) TF4 dataset

Insect class	Pre-cleaning	Post-cleaning
Anastrepha	674	671
Ceratitis	1,463	747
Rhagoletis	1,030	948
Bactrocera	1,061	1,043

with complex backgrounds, and finally apply ML including DCNN models to classify tephritids.

Data collection: For insect images collection, we used the iNaturalist website, which is frequently used by citizen scientists and biologists to share observations of biodiversity across the world and contain 92.30% to 97.30% proper classification of biological entities [28]. Our first dataset, namely "**FF2**", includes all images of dipteran fruitflies available at the website and further divided into two major groups - tephritids and non-tephritid dipteran flies. We used "diptera" search term and downloaded 89,247 images using in-house python script. The search term "tephritidae" was used to download images of "tephritids", whereas non-tephritid dataset was build by subtracting tephritid images from the diptera dataset. In total, we downloaded 89,247 images for FF2 dataset, containing 5,306 and 83,941 *Tephritidae* and non-Tephritid dipteran flies, respectively. Our second dataset, namely "**TF4**", contains images of four major genera of tephritid fruitflies, namely *Anastrepha*, *Ceratitis*, *Rhagoletis*, and *Bactrocera*. We used respected genus names as search terms, and found 674, 1463, 1030 and 1061 images for *Anastrepha*, *Ceratitis*, *Rhagoletis*, and *Bactrocera*, respectively. After downloading the images, our **TF4** dataset contains 4228 images in total and divided under respected genera names.

Data pre-processing: Due to some poor-quality, uninformative images and large scale of the dataset, it is essential to develop an automatic method to detect and remove irrelevant images. The automatic pre-processing method follows several steps: **1)** Pre-trained Inception-V3 model was used to get a 2,048 dimensional feature vector and a 1000 dimensional score vector corresponding to ImageNet classes, with reduced size of 299×299. **2)** Principal component analysis (PCA) and a non-linear t-Distributed stochastic neighbor embedding (t-SNE)[29] were applied to reduce feature dimensions from 2,048. **3)** k-means clustering was considered to group similar images in a single cluster. Elbow curve (see Figure 1(a)) was used to determine the value of k, which is 20 as the Within-Cluster-Sum-of-Squared Errors (WCSS) became very steady after the value. The 2d visualization of the cluster is shown in Figure 1(b). **4)** It was found that flies are identified in 14 classes (range 303 to 320 excluding 306, 311, 312, 315) among the 1,000 ImageNet classes, we call the range as *'flies range'*. Samples probability scores are shown in Figures 1(c) and 1(d). If highest peak was not found in the *'flies ranges'* and sample score in the *'flies ranges'* was <0.15 in any clusters, they were excluded. Images were also removed from other clusters if sample score in the *'flies ranges'* was <0.15. Other images on the other clusters were used for further analysis.

Since we used a fixed Inception-V3 model pre-trained on ImageNet to quickly scan our databases, the extracted features do not describe the images very accurately. Inevitably, we lose a certain amount of images using this automated process. To quantify our loss, we randomly selected 100 samples from deleted photos and manually examined them to calculate the loss rate. The average loss rate ranges from 17% to 23%. We employ this automatic cleaning method mainly on the class non-tephritid dipteran flies from **FF2** and *Ceratitis* from **TF4**, which turned out to include many irrelevant images of other species. For other classes with few uninformative images, k-mean clustering and simple manual check on few clusters were conducted to filter images. Detailed information about the pre- and post-cleaned **FF2** and **TF4** databases are presented in Table 1.
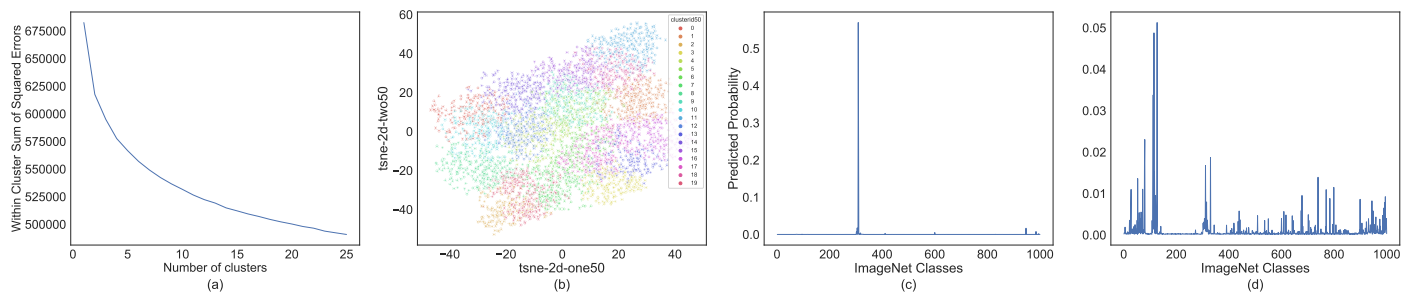
Figure 1. Automated Data Cleansing Methods, where (a) The Elbow Method Graph, (b) 2D Plot of T-SNE on PCA50, (c) Mean Probability Distribution of Kept Cluster, (d) Mean Probability Distribution of Removed Cluster

Classification models: We employed transfer learning (TL) with pre-trained models on our databases instead of training entire DCNN models from scratch (with random initialization). PyTorch provides different DCNN pre-trained models trained on sufficient data from ImageNet, including 1.2 million images from 1000 categories [30]. TL maximizes the performance of DCNN models on our image datasets and ensures fast convergence. Before training, all images need to be rescaled to a uniform size based on the input requirements of different DCNN models and then normalized by calculating the Z-score with the mean [0.485, 0.456, 0.406] and standard deviation [0.229, 0.224, 0.225] of three color channels computed on the ImageNet. Our experiments included the two major applications of TL. One is utilizing the pre-trained network as a fixed feature extractor. Without training to update weights, the base convolutional network can already learn meaningful features quickly. Using extracted feature maps, classification models like SVM, KNN, and NB were trained to classify the target fruit-fly classes. The second way to apply TL is by fine-tuning the pre-trained models. We started at the trained weights but modified the structure of a few top layers and unfroze some layers for retraining, which could train classifiers more relevant to our dataset conditions. Finally, we adapted ResNet, ResNeSt, and EfficientNet series to compare their performance.

3. Results

Experimental setup: We conduct stratified 10-fold cross-validation. For each fold, 80% of data are used for training, 10% for validation, and the rest for testing. We compute the loss and accuracy of the training and validation set after each epoch during training. The best weights are updated if a higher accuracy appeared on the validation set. Models are evaluated on the testing set after training each fold.

Evaluation: We return a confusion matrix after predicting the testing set and then compute accuracy (ACC), Precision (PPV), Recall (TPR), and F1-score as the measures. Since **FF2** has an imbalanced problem, we take the weights of class into account and employ weighted measures. The results are listed in Table 2 which reflect the average performance of cross-validation.

Implementation details: For traditional ML classifiers, we decided to extract features by the ResNet152 model. The average pooling layer of ResNet152 was selected as the end layer, returning a vector of 2,048 features for each image. After processing all images in the dataset, a large feature map with the dimension of 2,048 are prepared for running different ML classifiers. In nonlinear classification, Radial Basis Function (RBF) often performs the best among all kernel functions [31]. Therefore, we trained our SVM with the RBF kernel: $exp(-\gamma\|x - x'\|^2)$, where, γ determines the effect of a single training input. Another important parameter in defining RBF is $'c'$, which determines the smoothness of the decision surface. We used grid-search in LibSVM and cross-validation to find the optimal 'c' and 'gamma' values on the training set before SVM prediction. In KNN, the number of neighbor observations is set to be four based on computing euclidean distance. The likelihood of the features given a class was assigned to be Gaussian for NB classifiers.

Table 2. Performance comparison among the models on (a) FF2 and (b) TF4 datasets

(a) FF2 dataset				
Model	**ACC**	**PPV**	**TPR**	**F1-score**
ResNet+Complement NB	68	68	88	74
ResNet+SVM with RBF	76	76	89	80
ResNet+KNN4	90	90	86	85
ResNet50	89.97	90.70	**98.96**	94.64
ResNeSt50	93.46	92.29	90.85	91.52
EfficientNet-B0	**95.44**	**96.73**	98.22	**97.47**

(b) TF4 dataset				
Model	**ACC**	**PPV**	**TPR**	**F1-score**
ResNet+KNN4	42	44	42	43
ResNet+Gaussian NB	52	53	52	52
ResNet+SVM with RBF	62	63	62	62
ResNet50	75.60	75.84	75.60	75.57
ResNeSt50	85.45	85.68	85.45	85.45
EfficientNet-B0	87.57	87.69	87.57	87.56
EfficientNet-B2	**89.65**	**89.71**	**89.65**	**89.64**

For DCNN based image classification, we tuned the parameters according to the following rules: (1) We reset the size of the final fully connected layer to satisfy the actual number of target classes in our datasets that are 2 for **FF2** and 4 for **TF4**; (2) We unfreeze the weights of all layers to be retrained; (3) Adam optimizer is adapted with a batch size of 32; (4) The default loss function is cross-entropy; (5) We set the learning rate to start at 10^{-3} and decrease to a tenth after every seven epochs with a scheduler; (6) Models are trained 15 epochs in a fold on **FF2** and 50 epochs on **TF4** to guarantee convergence.

FF2 is an imbalanced dataset containing 5,287 Tephritids and nearly 8.5 times lower than 45,103 non-Tephritid dipteran flies. We improved some of the default methods to solve this problem: (1) SVM was implemented with assigned class weights [1, 8.5] to tell the model paying more attention to the category with fewer images. (2) we applied the complement Naive Bayes instead of Gaussian NB, which is particularly strong in dealing with imbalanced data[32].

Hardware setup: The experiments are conducted on a 2.6 GHz Intel Core i7-10750H CPU with 16 GB of RAM and an NVIDIA GeForce RTX 2070 GPU with 16 GB of memory. The work was implemented in Jupyter Notebook, Python, and other supporting libraries such as Sklearn and Pytorch.

Classification performance: As we see from Table 2, EfficientNet-B0 surpassed all other models with the highest ACC (95.44%), PPV (96.73%), and F1-score (97.47%) where Complement NB performed the worst on **FF2**. ResNet50 got a high TPR (98.96%). The confusion matrix in Figure 2(a) presents the prediction result of the optimal EfficientNet-B0 model on one testing set. 8 of the 4,511 non-tephritids were miss-classified as tephritids and 19 of the 528 tephritids were miss-classified as non-tephritids. For **TF4**, EfficientNet-B2 turned out to be the best with over 89.00% values in all measures as shown in Table 2(b). ResNet+KNN4 obtained lower than 50% accuracy in identifying the genus of fruit fly pests. Similarly, the prediction result of the optimal EfficientNet-B2 model is presented in Figure 2(b). 80.60% of *Anastrepha*, 88.46% of *Bactrocera*, 94.67% of *Ceratitis*, and 88.42% of *Rhagoletis* were correctly classified.
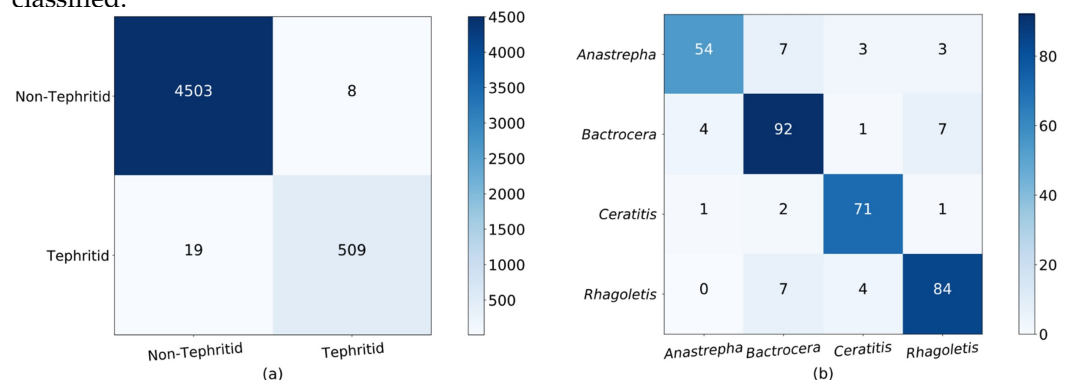
Figure 2. Confusion matrix for (a) EfficientNet-B0 on FF2, (b) EfficientNet-B2 on TF4 datasets.

4. Discussion

Our study aims to build a large dataset of fruit fly images with complex ecological backgrounds to routinely identify economically important tephritid fruit flies and its major genera. We collected images from online open source, built an automated approach to pre-process images and finally evaluated the dataset using traditional ML and DCNN based systems - to distinguish between tephritids and non-tephritid dipteran flies, and among four most destructive fruit flies. The step-by-step improvement in the prediction accuracy illustrated the high quality of our dataset, which can be widely employed by other researchers to evaluate the optimization of models in tephritids identification.

We evaluated our model, on both **FF2** and **TF4** datasets. For **FF2** dataset, compared to other models EfficientNet-B0 achieves the highest testing accuracy of 95.44%. ResNet with traditional classifiers (NB, SVM and KNN4) results in poor accuracy, since they treat all features independently and weights equally. Compared to ResNet50, the improved ResNeSt50 model with more attention on target area, results in 3.5% increase in accuracy. As the number of the tephritid images are 8.5 times lower than the non-tephritid flies, the identification of later provide accuracy with less miss-classified instances as shown in Figure 2. Since **TF4** dataset is much smaller than **FF2** dataset in size, we were able to scale up EfficientNets to EfficientNet-B2 and the later obtained the highest accuracy of 89.65% in distinguishing among *Anastrepha*, *Bactrocera*, *Ceratitis*, and *Rhagoletis* fruitflies. Like previous, traditional ML techniques do not perform well with ResNet extracted deep features. The optimal SVM only achieve nearly 60% accuracy where KNN and NB are even close to a random classifier. In contrast, DCNN models show outstanding improvements with at least 75% prediction accuracy based on a ResNet50 model. ResNeSt50 further improve the accuracy to 85.45%. Among the four genera *Ceratitis* and *Anastrepha* show highest and lowest accuracy of 94.67% and 80% (please see Figure 2(b)). The reason might be due to unique morphological characteristics of *Ceratitis* than other three genera or discrepancy in the number of images.

In comparison, Kasinathan et al.[21] utilised traditional ML models, with NB, SVM, KNN, and CNN classifiers to annotate few agricultural insect orders and families, and with optimal CNN classifier, their model achieved approximately 90% accuracy. Deng et al. [23] conducted a linear SVM on invariant features extracted by ROI and achieved accuracy of 85.5% to classify orders of tea plant insects. The classification between genus or species is challenging than distinguishing insects orders due to higher similarities in morphological appearances. CNN model of Hansen et al.[33] got 74.9% accuracy to predict ground beetles species. Motta et al.[34] implemented CNN models including LeNet, AlexNet and GoogleNet to classify three mosquito species and achieved highest of 76.2% accuracy. It is notable that, all of these studies considered small, heavily curated insect images and most importantly does not contain any representatives of dipteran fruitflies. We found few studies, related to ours. Martins et al.[19] tested several CNN models on field collected images and got nearly 92% in identifying *C. capitata* and *Grapholita Molesta*, which are visually distinguishable and belongs to different insect orders. Recently, CNN and DCNN models showed potential for species-level classifications of *Bactrocera* and *Anastrapha*, with accuracy ranging from 92.04% for complex background images to 95.68% for noise-free images [20,26]. The collected images of our study are shared by different people worldwide and the shooting environment, the size of the fruit fly target, and the lighting conditions varies widely which is much more complicated than collecting at a specific field and time for most researches. Under such heavy challenge, we still achieved 89.65% and 95.44% prediction accuracies on **TF4** and **FF2** datasets using EfficientNet-B2 and EfficientNet-B0 models respectively, and our model showed potential to enhance the accuracy further.

In summary, our model outperformed most of the studies to distinguish between tephritids and non-tephritids, and among the four tephritid genera. There is still much scope to enhance the accuracy by fine-tuning some parameters and enhancing computing resources or number of representative images. Models with larger scales in the EfficientNet

series and ResNeSt series are expected to provide better classification results in future study.

5. Conclusions

There are numerous biological control programs undergoing worldwide to control tephritid fruitflies [35,36] and our study can improvise such programs by identifying prevalent pest species of a particular area, before applying specific control measure. The outcome of current study is not only the model, but also the comprehensive dataset, which can be utilised in pest dynamics studies of a target area. Our model can be adopted easily for species-level classification of tephritid fruitflies as well as for other dipteran flies. It will be helpful to determine non-targeted insects, which cause benefits to crops and maintain ecological balance in the environment, but often overlooked in other control measures. In return, the possibility of exotic pest incursion can be decreased, the economically dangerous fruit flies can be quickly detected and better monitored. For field level implementation, low-cost application based tool can be built easily using the dataset and pre-trained model.

References

1. Health, P. *The Australian Handbook for the Identification of Fruit Flies Version 3.1*; planthealthaustralia, 2018.
2. Sutherst, R.W.; Collyer, B.S.; Yonow, T. The vulnerability of Australian horticulture to the Queensland fruit fly, *Bactrocera (Dacus) tryoni*, under climate change. *Australian Journal of Agricultural Research* **2000**, *51*, 467–480.
3. Sciarretta, A.; Tabilio, M.R.; Lampazzi, E.; Ceccaroli, C.; Colacci, M.; Trematerra, P. Analysis of the Mediterranean fruit fly [*Ceratitis capitata* (Wiedemann)] spatio-temporal distribution in relation to sex and female mating status for precision IPM. *PloS one* **2018**, *13*, e0195097.
4. Kubi, C.; Van Den Abbeele, J.; de Keken, R.; Marcotty, T.; Dorny, P.; van den Bossche, P.; van den Bossche, P.; Akoda, K.; Djagmah, B.; De Deken, R.; others. CRP: Improved and Harmonized Quality Control for Expanded Tsetse Production, Sterilization and Field Application. *Evolutionary Biology*, *22*, 1516–1525.
5. Christenson, L. The male annihilation technique in the control of fruit flies; ACS Publications, 1963.
6. Elliott, N.; Farrell, J.; Gutierrez, A.; van Lenteren, J.C.; Walton, M.; Wratten, S. *Integrated pest management*; Springer Science & Business Media, 1995.
7. Kapoor, V. Taxonomy and biology of economically important fruit flies of India. *Israel Journal of Entomology* **2005**, *35*, 459–475.
8. Freidberg, A. CABIKEY to the Indo-Australian Dacini fruit flies (CD-ROM). By IM White & DL Hancock (Wallingford, CAB International, 1997). *Bulletin of Entomological Research* **1999**, *89*, 109–109.
9. Virgilio, M.; White, I.; De Meyer, M. A set of multi-entry identification keys to African frugivorous flies (Diptera, Tephritidae). *ZooKeys* **2014**, p. 97.
10. Armstrong, K.; Ball, S. DNA barcodes for biosecurity: invasive species identification. *Philosophical Transactions of the Royal Society B: Biological Sciences* **2005**, *360*, 1813–1823.
11. Hebert, P.D.; Cywinska, A.; Ball, S.L.; Dewaard, J.R. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **2003**, *270*, 313–321.
12. Wang, J.; Chen, Y.; Hou, X.; Wang, Y.; Zhou, L.; Chen, X. An intelligent identification system combining image and DNA sequence methods for fruit flies with economic importance (Diptera: Tephritidae). *Pest Management Science* **2021**.
13. Yang, H.P.; Ma, C.S.; Wen, H.; Zhan, Q.B.; Wang, X.L. A tool for developing an automatic insect identification system based on wing outlines. *Scientific reports* **2015**, *5*, 1–11.
14. Favret, C.; Sieracki, J.M. Machine vision automated species identification scaled towards production levels. *Systematic Entomology* **2016**, *41*, 133–143.
15. Wang, J.; Lin, C.; Ji, L.; Liang, A. A new automatic identification system of insect images at the order level. *Knowledge-Based Systems* **2012**, *33*, 102–110.
16. Xie, C.; Zhang, J.; Li, R.; Li, J.; Hong, P.; Xia, J.; Chen, P. Automatic classification for field crop insects via multiple-task sparse representation and multiple-kernel learning. *Computers and Electronics in Agriculture* **2015**, *119*, 123–132.

17. Xia, D.; Chen, P.; Wang, B.; Zhang, J.; Xie, C. Insect detection and classification based on an improved convolutional neural network. *Sensors* **2018**, *18*, 4169.

18. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *International journal of computer vision* **2004**, *60*, 91–110.

19. Martins, V.A.; Freitas, L.C.; de Aguiar, M.S.; de Brisolara, L.B.; Ferreira, P.R. Deep Learning applied to the Identification of Fruit Fly in Intelligent Traps. 2019 IX Brazilian Symposium on Computing Systems Engineering (SBESC). IEEE, 2019, pp. 1–8.

20. Peng, Y.; Liao, M.; Deng, H.; Ao, L.; Song, Y.; Huang, W.; Hua, J. CNN–SVM: a classification method for fruit fly image with the complex background. *IET Cyber-Physical Systems: Theory & Applications* **2020**, *5*, 181–185.

21. Kasinathan, T.; Singaraju, D.; Uyyala, S.R. Insect classification and detection in field crops using modern machine learning techniques. *Information Processing in Agriculture* **2020**.

22. Ding, W.; Taylor, G. Automatic moth detection from trap images for pest management. *Computers and Electronics in Agriculture* **2016**, *123*, 17–28.

23. Deng, L.; Wang, Y.; Han, Z.; Yu, R. Research on insect pest image detection and recognition based on bio-inspired methods. *Biosystems Engineering* **2018**, *169*, 139–148.

24. Zhu, L.Q.; Ma, M.Y.; Zhang, Z.; Zhang, P.Y.; Wu, W.; Wang, D.D.; Zhang, D.X.; Wang, X.; Wang, H.Y. Hybrid deep learning for automated lepidopteran insect image classification. *Oriental Insects* **2017**, *51*, 79–91.

25. Cheng, X.; Zhang, Y.; Chen, Y.; Wu, Y.; Yue, Y. Pest identification via deep residual learning in complex background. *Computers and Electronics in Agriculture* **2017**, *141*, 351–356.

26. Leonardo, M.M.; Carvalho, T.J.; Rezende, E.; Zucchi, R.; Faria, F.A. Deep feature-based classifiers for fruit fly identification (diptera: Tephritidae). 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). IEEE, 2018, pp. 41–47.

27. Qin, Y.; Paini, D.R.; Wang, C.; Fang, Y.; Li, Z. Global establishment risk of economically important fruit fly species (Tephritidae). *PLoS One* **2015**, *10*, e0116424.

28. Unger, S.; Rollins, M.; Tietz, A.; Dumais, H. iNaturalist as an engaging tool for identifying organisms in outdoor activities. *Journal of Biological Education* **2020**, pp. 1–11.

29. Linderman, G.C.; Steinerberger, S. Clustering with t-SNE, provably. *SIAM Journal on Mathematics of Data Science* **2019**, *1*, 313–332.

30. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009, pp. 248–255.

31. Divya, B.; Santhi, M. SVM-based Pest Classification in Agriculture Field. *International Journal of Recent Technology and Engineering (IJRTE)* **2019**, *7*, 2277–3878.

32. Rennie, J.D.; Shih, L.; Teevan, J.; Karger, D.R. Tackling the poor assumptions of naive bayes text classifiers. Proceedings of the 20th international conference on machine learning (ICML-03), 2003, pp. 616–623.

33. Hansen, O.L.; Svenning, J.C.; Olsen, K.; Dupont, S.; Garner, B.H.; Iosifidis, A.; Price, B.W.; Høye, T.T. Species-level image classification with convolutional neural network enables insect identification from habitus images. *Ecology and evolution* **2020**, *10*, 737–747.

34. Motta, D.; Santos, A.Á.B.; Winkler, I.; Machado, B.A.S.; Pereira, D.A.D.I.; Cavalcanti, A.M.; Fonseca, E.O.L.; Kirchner, F.; Badaró, R. Application of convolutional neural networks for classification of adult mosquitoes in the field. *PLoS one* **2019**, *14*, e0210829.

35. Yosiaki, I.; Kakinohana, H.; Yamagishi, M.; Kohama, T. Eradication of the melon fly, *Bactrocera cucurbitae*, from Okinawa, Japan, by means of the sterile insect technique, with special emphasis on the role of basic studies. *Journal of Asia-Pacific Entomology* **2003**, *6*, 119–129.

36. Estes, A.; Nestel, D.; Belcari, A.; Jessup, A.; Rempoulakis, P.; Economopoulos, A. A basis for the renewal of sterile insect technique for the olive fly, *Bactrocera oleae (Rossi)*. *Journal of Applied Entomology* **2012**, *136*, 1–16.