Zoological Studies **64:**24 (2025)

A Systematic Review of Species Classification Using Deep Learning Algorithms and Gender Identification of *Tribolium castaneum* Using Convolutional Neural Networks

Anurupa Mistry[1], Chetas Hedaoo[2], Archana Sharbidre[3,*], Jayashri Bagade[4,*], and Sangeeta V. Pandit[5]

[1]*Department of Zoology, Savitribai Phule Pune University. Pune, Maharashtra, India. E-mail: anurupamistry@gmail.com (Mistry)*

[2]*Department of Electronics and telecommunication, Vishwakarma Institute of Information Technology, Pune, Maharashtra, India. E-mail: chetas.22111109@viit.ac.in (Hedaoo)*

[3]*Department of Zoology, Savitribai Phule Pune University. Pune, Maharashtra, India. *Correspondence: E-mail: aasharbidre@unipune.ac.in (Sharbidre)*

[4]*Department of Information Technology, Vishwakarma Institute of Technology, Pune, Maharashtra, India. *Correspondence: E-mail: jayashrihedaoo@rediffmail.com (Bagade)*

[5]*Department of Zoology, Savitribai Phule Pune University. Pune, Maharashtra, India. E-mail: drpanditsv@unipune.ac.in (Pandit)*

(Received 1 September 2023 / Accepted 16 April 2025 / Published -- 2025)
Communicated by Sheng-Feng Shen

ORCiD:
Anurupa Mistry: https://orcid.org/0000-0002-4677-826X
Chetas Hedaoo: https://orcid.org/0009-0005-3871-0379
Archana Sharbidre: https://orcid.org/0000-0003-1238-2846
Jayashri Bagade: https://orcid.org/0000-0002-9709-5530

Machine learning (ML) constitutes a division of artificial intelligence (AI) that aims to train computers how to perform specific tasks without explicit programming. Traditional ML tools are widely used for classification and identification of animals. However, these methods have some drawbacks because of the extensive manual reliance and the delay in data interpretation. To overcome this, Applied Deep Learning algorithms are used with Artificial Neural Networks (ANN) and Convolution Neural Network (CNN) models introduced to address species classification, characteristics detection, and pattern recognition tasks helping in accurate identification and classification of animals.

In this paper, we have tried to compile and deliver a recent comprehensive information on latest available investigations in the field of life sciences particularly used for animal identification. We have also accentuated the diverse applications of machine learning models including other parameters like, features, accuracy gained, database used and their limitations.

The red flour beetle, *Tribolium castaneum* (Coleoptera; Tenebrionidae) is a prevailing and detrimental secondary insect pest of stored grains along with derived products causing 7% to 35% annual loss. Despite of that, nowdays it is also extensively considered as a model organism for genetic disease investigation. While using it in scientific research, exact sex identification of these insects becomes a crucial preliminary step. Generally, pupal stage is used to sort these insects according to their sex and needs expert humans. It is crucial to employ image processing and ML algorithms to quickly identify gender of this insect which is not done yet.

We have used a CNN-based smart technique to recognize and categorize gender differences in *T. castaneum* using microscopic images in order to build an intelligent system for applied research. For this study, a dataset is created by taking 116 microscopic images of both the dorsal and ventral sides of pupae of two different sexes. In this algorithm, a 2D matrix of feature map is selected sequentially and the maximum value in the matrix is selected to generate a pooled feature map. The Rectified Linear Unit (ReLU) activation function is used for the CNN. The classification model has an accuracy between 97 and 98% with an F-score of 0.67. These results demonstrate the robustness of the classification model, which does not rely heavily on manual intervention compared to traditional machine learning (ML) tools and automates the processes of feature extraction and gender classification regardless of the position of the pupae in the images.

Keywords: Species Classification, Deep Learning, CNN, *Tribolium. Castaneum,* Gender Identification

Citation: Mistry A, Hedaoo C, Sharbidre A, Bagade J, Pandit SV. 2025. A systematic review of species classification using deep learning algorithms and gender identification of *Tribolium castaneum* using convolutional neural networks. Zool Stud **64:**24.

BACKGROUND

Tribolium castaneum, (Herbst 1797) commonly known as the red flour beetle and belonging to the Coleoptera order within the Tenebrionidae family, is a prevalent and destructive secondary insect pest that primarily targets stored grains and their derived products. This species exhibits sexual dimorphism (Sokoloff 1974; Rees 2004; Mahroof and Hagstrum 2012). Under optimal conditions, the developmental time for *T. castaneum* is approximately five days for eggs, twenty days for larvae, and seven days for pupae (Sokoloff 1974; Dawson 1964). Both larvae and adult beetles infest stored food and grains, posing a significant threat to agricultural commodities. Hana

2013 reported that the damage inflicted by these beetles accounts for a considerable percentage, ranging from 7% to 35%, of total agricultural production on an annual basis.

Nowdays, *T. castaneum* have gained recognition as a valuable model organism for investigating the underlying causes of genetic diseases. Notably, *T. castaneum* was the first species within the Coleoptera order to have its genome sequenced (Richards et al. 2008; Herndon et al. 2020). This type of insect species is progressively employed across a diverse array of biomedical investigations, covering fields like neurodegenerative ailments (such as Parkinson's disease), the signaling pathway for diuretics (which involves the function of vasopressin-like peptide and its receptor), interactions between hosts and pathogens (encompassing antagonistic interactions and co-evolution), as well as the domains of pharmacology and toxicology (specifically in the analysis of the impacts of psychoactive substances) (Denell 2008; Grunwald et al. 2013; Bingsohn et al. 2016; Adamski et al. 2019). Although there exist certain dissimilarities in cellular characteristics and overall structure between humans and the red flour beetles, a multitude of genetic, physiological, and immunological traits remain consistent. This conservation renders them a valuable model for investigating diverse facets of human biology. *T. castaneum*'s utilization as a model for practical research extends beyond postharvest management, encompassing inquiries into aging, the environment, and pest control (Thomson et al. 2014; Wijayaratne et al. 2018). Several studies have been conducted on the evolutionary and the pre- peri and post-mating sexual selection behavior of *T. castaneum* (Michalczyk et al. 2010).

The red flour beetle, known for its reddish-brown coloration and three-segmented clubbed antennae (Bousquet 1990), exhibits distinct sexual dimorphism. In *T. castaneum*, differentiation between males and females can be established by the presence of genital papillae during the pupal stage and sex patches in the bodies of adult insects. This differentiation is applicable to both pupal and adult stages. Notably, sex determination is most straightforward during the pupal stage (Kramarz et al. 2016). The morphological characteristics of the insect, influenced by both genotype and phenotype, indirectly impact the process of gender classification. Female pupae display pointed genital papillae, whereas male pupae possess stubby and barely noticeable papillae.

Accurate identification of the sex of individuals is a crucial initial step in characterizing the population. As the sizes of the beetles are very small (3–4 mm), it is difficult to be perceived with human eyes. However there is need identify it using microscopic images. Perseverance of large number of microscopic images create fatigue to human subjects which will endup in to incorrect outcome. This repeated task will be very well handled by machine. Thus machine learning plays a very vital role in classification of beetles. There is an urgent need to integrate some fast processing techniques to speed up the experiments going on them.

Machine learning has emerged as a viable alternative to traditional technical methodologies in various domains of science and technology. By leveraging data-driven approaches, machine learning techniques offer the potential to expedite the design process, minimize complexity, and enhance cost-effectiveness (Simeone 2018).

Machine learning is a subset of artificial intelligence dedicated to instructing computers in the execution of particular tasks, all without necessitating direct, explicit programming. Computers are fed structured data and 'learn' to become better at evaluating and acting on that data over time. A computational framework inspired by the structure of biological neural networks, which forms the foundation of the human brain, is commonly denoted as an artificial neural network. These networks are capable of effectively processing vast quantities of data. Artificial neural networks (ANNs) have been shown to be effective tools for various kinds of tasks, but they also have a number of disadvantages. As ANNs to generalize effectively, a lot of labelled training data is often needed. Overfitting or low accuracy might result from a lack of data or data with poor quality. Neural networks frequently perform the function of "black boxes," which means they can make accurate predictions but cannot be interpreted. The "curse of dimensionality" refers to the exponential increase in data required to generalize the machine learning model accurately as the number of dimensions or characteristics rises. Additionally, ANN requires features that are handmade for model training and testing.

Very few characteristics enables male and female *T. castaneum* pupae to be distinguished from one another. Because the pupae body is white in colour and has less texture, it is difficult for conventional machine learning approaches to correctly recognize it. Female and male pupae 117 could be distinguished by the size and shape of genital papillae, located anterior to the 118 urogomphi. Females have larger and finger like papillae whereas males has smaller papillae. To overcome this problems we need deep learning network to classify the beetles. Convolution Neural Network (CNN) model is used to classify beetls automatically. A Convolutional Neural Network (CNN) can comprise tens or even hundreds of layers, with each layer designed to learn and recognize different features within an image. During training, every image undergoes filtering at multiple resolutions, and the outcome of each convolution is employed as input for the subsequent layer. Starting with fundamental attributes such as brightness and edges, these filters can progress towards more intricate aspects, ultimately culminating in features that distinctly identify the object. Deep learning algorithms, notably convolutional neural networks (CNN), have garnered substantial attention due to their capabilities in pattern recognition tasks related to image analysis. Their popularity is notably prominent in the field of biological sciences.

In our proposed study, we aim to devise an intelligent approach utilizing deep learning techniques to discern and categorize disparities observed in both species and gender based on microscopic images captured from ventral and dorsal views. This endeavor seeks to contribute to the advancement of intelligent systems within the realm of applied research.

Literature survey:

The table 1 (presented below at the end of manuscript) provides a comprehensive overview of recent investigations in the field of life sciences, highlighting the diverse applications of machine learning models. Predominantly, supervised learning models have been employed to address species classification, characteristics detection, and pattern recognition tasks. An array of machine learning algorithms, including support vector machines (SVM), logistic regression (LG), random forests (RF), gradient boosting (GB), k-nearest neighbors (kNN), decision trees (DT), and deep learning (DL), have been employed to achieve these goals. The utilized datasets encompass both publicly available resources and researcher-generated collections through sample acquisition. The fundamental strategy in constructing classification models entails dividing the dataset into a training set for model development and a test set for the purpose of validating and evaluating the model's dependability. Accuracy measurements reveal that deep learning variants, including artificial neural networks (ANN) and convolutional neural networks (CNN), demonstrate superior robustness compared to other classification models. Nevertheless, it is important to acknowledge that the accuracy of these models is influenced by certain limitations, such as the quantity and quality of the datasets employed, as well as class imbalance, which arises when multiple species classes are involved in the classification and identification studies. It is widely acknowledged that there has been minimal research conducted so far regarding gender classification using computational methods in T. castaneum. In this proposed study, we aim to utilize a CNN-based intelligent approach to detect and differentiate gender disparities from microscopic images. This endeavor is aimed at contributing to the advancement of intelligent systems within the domain of applied research.

Table 1. Summery of features, accuracy, model and limitations

S. No.	Ref No.	Features	Model Used	Accuracy	Database	Limitations
1	Themozhi et al. 2019	Automatic feature extraction for image classification	Deep CNN model for classification	95.97–97.47%	National Bureau of Agricultural Insect Resources (NBAIR) dataset, Xie1 and Xie2	Lower accuracy for higher mini batch size.
2	Lee et al. 2021	machine learning model developed to diagnose malaria by leveraging patient-related information	support vector machine, random forest (RF), multi-layered perceptron, AdaBoost, gradient boosting (GB), and CatBoost	56.9–85.6%	Dataset containing information on parasitic diseases as well as a broader dataset encompassing various other diseases gathered from patient information found within PubMed abstracts spanning the years from 1956 to 2019.	Datasets utilized have small sizes, and they incorporate a constrained set of features without undergoing the feature selection process.
3	Safavi et al. 2022	ExtraTreesClassifier algorithm for selecting important predictive features in forecasting the disease occurrence and artificial neural network (ANN) algorithm in predicting the occurrence of LSDV infection in unseen data	ExtraTreesClassifier and ANN	97%	Data from multiple sources, including the Global Animal Disease Information System of FAO (Food and Agriculture Organization), meteorological data, the Gridded Livestock of the World (GLW 3) database, the GLC-SHARE Beta-Release v1.0 dataset, and the Natural Earth database.	Study incorporates data obtained from inactive accounts within veterinary facilities across different countries. The dataset comprises a limited amount of information and is characterized by a small number of predictor variables utilized for analysis.
4	González-Pérez et al. 2022	Automated categorization of mosquitoes based on their genus and gender by using five distinctive features extracted from wingbeat recordings.	logistic regression, (LR), gradient boosting (GB), random forests (RF), support vector machines (SVM) and a fully connected deep neural network (DNN)	Genus classification: 94.2%, Sex classification of Aedes: 99.4%, sex classification of Culex: 100%	Flight recordings of 4335 mosquito through a novel optical sensor	Slight overfitting: more training samples for genus classification model
5	Kirkeby et al. 2021	3 methods for automatic classification of insect groups	Fourier transformation of wingbeat frequency, Random Forest Classifier and 3-layer Neural network	80%	10,000 records of airborne insects discovered in oilseed rape (*Brassica napus)* fields, captured via an optical remote sensor	only 4 species of pests considered

6	Pataki et al. 2021	Deep learning model to find tiger mosquitoes (Aedes albopictus) from 7686 citizen-made mosquito photos between 2014 and 2019	deep learning model, ResNet5026	96%	Mosquito Alert's curated database	Data size used for testing is not optimum.
7	Kittichai et al. 2021	One stage and two stage learning methods for classifying species and gender of different mosquito species	deep learning model, YOLO	97–98.9%	10564 captured images of mosquitos	small number of images for two species in the training set. Some species are not identified through the learning methods
8	Cannet et al. 2022	Automatized identification of species of tsetse flies using deep learning architecture and Wing Interference Patterns	CNN	33–100%	A collection of 1,766 images depicting 23 distinct species of *Glossina* (tsetse flies).	A notably limited quantity of images belonging to a specific species is present within the test dataset for WIPs
9	Lei et al. 2019	3 convolutional layer model for identifying handwritten digits	Dilated CNN and HDC models	60–100%	MNIST data set	
10	Antipov et al. 2016	CNN model for gender predication from face image	CNN Ensemble model	96.8–97.3%	CASIA Web Face and Labelled Faces in the Wild (LFW)	Number of images in CASIA Web Face database is excessive with respect to the number of subjects
11	Chola et al. 2022	A classification model for determining the gender of Drosophila melanogaster by utilizing a combination of color, shape, and texture features.	support vector machines (SVM), Naive Bayes (NB), and K-nearest neighbour (KNN)	90%	Photographs of Drosophila specimens obtained from the National Drosophila Stock Centre, Department of Studies in Zoology, University of Mysore, India.	The dataset comprises a relatively small quantity of images, specifically 100 images categorized into two distinct classes.
12	Ozdemir et al. 2022	21 base criteria (keys) for classification of insect order using deep learning models	SSD MobileNET, YoloV4, and Faster R-CNN InceptionV3	67–81%	1500 insect images	Image issues can deteriorate the performance of the model.
13	Rabinovich et al. 2021	Experimental design using machine learning by taking temperature,	support vector machine, GBMs, Bayesian generalized linear	80%	Dataset of 228 insects through a combination of 4 features	Limited size of dataset

		exposure times, stage and sex of adult kissing bugs as features	models, linear models fitted with ordinary least squares			
14	Motta et al. 2019	Autonomous classification of adult mosquitoes by extracting features from the mosquito images using CNN model	CNN (LeNet, AlexNet, GoogleNet)	57.5–83.9%	4056 mosquito images extracted from ImageNet platform	Limited number of images lead to risk of overfitting
15	Bjerge et al. 2022	Insect Classification and Tracking algorithm (ICT) for real-time classification and tracking of insect species using intelligent camera system and deep learning model	deep learning model, YOLO	89%	2121 background images without insects and 5757 images with insects	Precision of the tracking algorithm is highly dependent on the duration of visibility of insects on the camera
16	Zare et al. 2022	Boosting method to train classifier and feature extraction from microscopic images for detecting leishmaniasis	Viola-Jones algorithm	83%	A dataset encompassing 300 images extracted from 50 laboratory slides obtained from lesions under suspicion of leishmaniasis.	size of data set is quite low and accuracy depends on the resolution of images
17	Bellin et al. 2021	Classifier models for identifying two species of mosquitos using geometric morphometrics data and pairwise comparison for feature extraction	support vector machine (SVM), random forest (RF), artificial neural network (ANN) and an ensemble model (EN).	73-81%	664 mosquito specimens of Maculipennis complex	The proportion of two species in the test set is highly skewed (training set: 1:1 and test set: 24:222)
18	Markovic et al. 2021	Classifier models to predict the appearance of insects during a season on a daily basis using 21 parameters	K-Nearest Neighbours, Support Vector Machines, Decision Tree, Random Forest, Multi-layer Perceptron classifier, Ada Boost, Gaussian Naive Bayes and Quadratic	75–86.3%	*Helicoverpa armigera* insects from 17 locations in the northern part of Serbia during 2019 and 2020.	Temperature and humidity are the only environmental parameters considered

			Discriminant, Analysis			
19	Shen et al. 2018	Classifier model for detection of stored grain insects by applying an inception structure for convolution neural network	R-CNN	88%	12508 images of six different species of insect	Accuracy is highly dependent on the resolution of insect images
20	Wittek et al. 2022	Supervised machine learning predictive classifiers for pigeon behaviours using multivariate time series data for 10,424,241 frames as input	Decision Trees, Random Forest	87%	Eight naïve adult homing pigeons each receiving 10-20 sessions	Absence of filters during the implementation of machine-learning based tracking software DeepLabCut that resulted in tracking glitches and instances of anatomically implausible movements observed in pigeons during the tracking process..
21	Borba et al. 2021	Distinguishing taxonomic species by analyzing morphological, morphometric, and ecological data from capilliards.	J48, Random Tree, REPTree, LMT, Majority Voting	82–97%	Samples procured from two helminth collections associated with institutions, containing a total of 28 distinct species and 8 different genera.	Utiilized dataset provides only a limited representation of the actual biological diversity observed within capillariids.
22	Abdelaziz et al. 2022	Automated classification of vertebrate species from the 3D images of their remains using 8 extracted physical features	Support Vector Machines (SVM), K-Nearest neighbours (KNN) and decision tree (DT) classifiers	83.4–93.7%	2052 3D images for three main classes	Models show very high training accuracy implying that there might be issues of overfitting. K-Fold cross validation is not performed.
23	Patel et al. 2020	Deep learning classification of Galápagos Snake Species using 6 external parameters extracted from images	R-CNN	75%	247 images of snakes making up 9 species	Dataset size is limited, not enough images to train the model properly and model classification is highly

						dependent on the resolution of images.
24	Acevedo et al. 2009	Automated classification of calls of nine frogs and three bird species using four standard call variables or eleven variables that included three standard call variables and a coarse representation of call structure	Support Vector Machines, Decision Trees and Discriminant Analysis	71.45–94.95%	10,061 isolated calls	The model accuracy is dependent on the species type and call frequency
25	Xie et al. 2016	Acoustic classification model of frogs using 14 features extracted from frog call recordings	linear discriminant analysis, K-nearest neighbour, support vector machines, random forest, and artificial neural network	94–99%	Recordings of 24 frog with the duration ranging from eight to fifty-five seconds	The accuracy diminishes with higher background noise
26	Petrescu et al. 2021	Fear classification model using 40 types of features from the physiological data	Decision Trees, k-Nearest Neighbours, Support Vector Machine and artificial networks	91.7–93.5%	Peripheral signals from DAEP dataset	Imbalanced classes and single self-assessment of the emotional status for the video extract
27	Shia et al. 2021	Physical features extracted from images for unsupervised classification of malignant tumours in breast	combination of locally weighted learning (LWL) and sequential minimal optimisation	84.70%	677 US images	Smaller dataset, clinical limitation of the application and biases associated with the observers
28	Lapp et al. 2021	Automated call recognition method for frog calls and choruses using 9 parameters	RIBBIT (repeat interval-based bioacoustic identification tool) classifier	90%	70 Audio files	Smaller dataset and accuracy dependent on heavily overlapping choruses, background noise and other species with similar vocalizations
29	Bisgin et al. 2018	Species identification of beetles based on 3 sets of image features	SVM, ANN	80–85%	set of 6900 images of 15 species of beetles	Limiting number of specimen images per species, quality of

						images and difficult pairs due to high entomological similarities
30	Zhu et al. 2021	Neural network for identification of 3 specific types of promoters in DNA sequence of species including Homo sapiens, Mus musculus, Drosophila melanogaster and Arabidopsis thaliana	Capsule neural network	80–98%	promoter sequences for four different species from the EPDNew database	Validity of model is compared against independent dataset that are subjected to continuous updating, the other tools selected for comparison only focusses on single species and most approaches do not identify all the promoters properly
31	Bisgin et al. 2022	Automated recognition through elytral pattern of food-contaminating beetles	CNN	90%	27 species of beetles collected from U.S. Department of Agriculture's (USDA) Animal and Plant Health Inspection Service (APHIS) laboratory	Accuracy dependent on better resolution and higher number of images for training set
32	Tannous et al. 2023	Automated identification of insect species using feature pyramid extracted from images with 3 different resolutions and scales	CNN	93%	1826 images of 2 species of insects	Small sized and morphologically similar insects are difficult to identify
33	Loti et al. 2021	Machine learning based automated identification of chili pest and disease in leaves using features extracted through 6 deep learning tools	support vector machine (SVM), a random forest (RF), and artificial neural network (ANN)	49-–92%	974 images of leaves	Small number of testing samples of 2 classes, loss of certain features during feature extraction process and patterns of discoloration
34	Veiner et al. 2022	Analyzing transcriptomic patterns and identifying genes linked to the honeybee waggle dance by	Support Vector Machine (SVM), Random Forest (RF), Generalized Linear Model (GLMNET)	66.7–100%	15314 gene counts of whole honeybee genome across 32 bees	Small size of training and test set

		employing the top 20 gene features for characterization.				
35	Gupta et al. 2023	Data augmentation with deep learning methodology for automatic identification of castor pest insects	convolutional neural networks VGG16, VGG19, and ResNet50	71.2–82.18%	Dataset of 372 images organized into six different insect pest classes	Data unbalancing (disproportionate number of species in dataset for classification)
36	Aladhadh et al. 2022	Pest detection using a deep learning framework using cross stage partial network (CSP) for feature extraction from insect images	CNN, Faster RCNN, YOLO-5	57.3–98%	Ants class: 392 images, grasshopper class: 315 images, palm weevil class: 48 images, shield bug class: 392 images, and wasps' class: 318 images	Disproportionate number of insect classes in the training and test datasets
37	Liu et al. 2022	Automated recognition of tomato pests using deep learning model implementing Triplet Attention Module (TAM) for feature extraction	Deep learning model YOLOv4-TAM	95%	2,893 images of induced plate pests collected from Shouguang tomato greenhouse	Image quality affects the accuracy and detection of anchor boxes that correspond to the pest dataset
38	Alsanea et al. 2022	Autodetection model for red palm weevil using region-based CNN to extract the features to enclose image with the bounding boxes	convolutional neural network (R-CNN)	99%	6000 images generated from available 300 images using data augmentation techniques	Non-availability of the dataset for the proposed model, dataset too small for the robust model creation
39	Dai et al. 2022	Autodetection of citrus psyllids using deep learning method by implementing High-resolution network (HRNet) for feature extraction from the images	Cascade region-based convolution neural networks (R-CNN)	89%	Dataset comprising of 500 high-definition sample images of plants sourced from the Citrus HLB Test Base at South China Agricultural University.	Small target detection range of citrus psyllid in the image
40	Spiesman et al. 2021	classification and identification of various bumble bee species using images by applying deep learning model	Deep learning models ResNet, Wide ResNet, InceptionV3, and MnasNet	85.8–91.7%	120,000 images belonging to 42 species of bumble bees	Significant fluctuations in error rates for species with limited sample sizes, primarily due to the extent of differences within the

						same species and their distinct characteristics from other species.
41	Zhao et al. 2022	automatic recognition of mosquito species by implementing an identification model based on the Swin Transformer architecture	Convolutional neural network (CNN) models	80–100%	9,900 mosquito images covering 7 genera and 17 species	Lack of images of particular sex of some mosquito species lead to unbalanced dataset

The following table offers a comprehensive overview of recent investigations conducted in the field of life sciences (Table 1), highlighting the various applications of machine learning models. Primarily, supervised learning models have been utilized to address tasks such as species classification, characteristics detection, and pattern recognition. To achieve these objectives, a variety of machine learning algorithms, including support vector machines (SVM), logistic regression (LG), random forests (RF), gradient boosting (GB), k-nearest neighbors (kNN), decision trees (DT), and deep learning (DL), have been implemented. The datasets used encompass both publicly available resources and collections generated by researchers through sample acquisition. The fundamental approach to developing classification models involves partitioning the dataset to create a training set for model training and a test set for the purpose of validating and assessing the model's credibility. Accuracy measurements indicate that deep learning variants, such as artificial neural networks (ANN) and convolutional neural networks (CNN), exhibit superior robustness compared to other classification models.

Researchers have employed various machine learning and deep learning classification models to study insects and pests. For example, Kirkeby et al. 2021, utilized three different methods to classify insect groups and achieved an impressive accuracy rate of 80%. Chola et al. 2022, employed classical machine learning methods to classify the gender of *Drosophila melanogaster*, achieving a maximum accuracy of 90%. Rabinovich et al. 2021, used machine learning models for experimental design, testing the thermal limits of kissing bugs and achieving an accuracy of 80%. Veiner et al. 2022, employed Support Vector Machine (SVM), Random Forest (RF), and Generalized Linear Model (GLMNET) to characterize and identify genes associated with honeybee waggle dance, achieving accuracy levels ranging from 67% to 100%. Perez et al. 2022, utilized a deep neural network to classify mosquitoes by genus and sex, achieving an accuracy exceeding 94%. Similar studies were conducted by Pataki et al. 2021; Kittichai et al. 2021; Motta et al. 2019 Bellin et al. 2021; and Zhao et al. 2022, who used deep learning models such as ResNet5016 and YOLO, as well as neural network models like ANN and CNN, to classify mosquito species based on insect images. These studies achieved accuracies ranging from 54% to 100%. Themozhi et al. 2019 and Aladhadh et al. 2022, applied CNN models to classify and detect crop pests using images, achieving accuracy levels higher than 95%. Liu et al. 2022, used a deep learning model to classify tomato pests from a collection of images, achieving an accuracy of 95%. Additionally, researchers have employed faster deep learning models like R-CNN for classifying various insects. Ozdemir et al. in 2022 used R-CNN to identify key indicators for insect classification, achieving an accuracy of over 80%. Shen et al. in 2018 and Alsanea et al. in 2022 applied the R-CNN classification model to images of stored grain insects and for the auto-detection of red palm weevil, respectively, resulting in accuracy levels of 88% and 99%. Bisgin et al. 2022 and Tannous et al. 2023, utilized CNN models for automated identification of food-contaminating beetles and insect species, achieving accuracy levels of 90% and 93%, respectively. Dai et al. 2022, employed cascade region-based convolutional neural networks (R-CNN) for the autodetection of citrus psyllids from images with an accuracy level of 89%. Aladhadh et al. in 2022 applied CNN and faster R-CNN models for the autodetection of pests from insect images with accuracy levels between 57% and 98%.

In addition to species classification studies, machine learning and deep learning models have also been applied in various other biological studies. Lee et al. 2020, used support vector machines, random forest (RF), multi-layered perceptron, AdaBoost, gradient boosting (GB), and CatBoost models for malaria diagnosis and achieved accuracy levels between 56.9% and 85.6%. Safavi et al. 2022, employed ExtraTreesClassifier and ANN models for forecasting and predicting the occurrence of LSDV infection, achieving a highest classification accuracy of 97%. Zare et al. 2022, applied a boosting method to train classifiers and extract features from microscopic images for detecting leishmaniasis with an accuracy level of 83%. Petrescu et al. 2021, used decision trees, k-nearest neighbors, support vector machines, and artificial neural network models for fear classification from physiological data, with accuracy levels between 91.7% and 93.5%. Shia et al. 2021, employed a combination of locally weighted learning and sequential minimal optimization for unsupervised classification of malignant breast tumors, achieving an accuracy level of 84.7%. Deep learning methods have been applied by Antipov et al. 2016 and Patel et al. 2020, who used CNN ensemble and R-CNN models for gender prediction from face images and classification of Galápagos Snake Species, with accuracies greater than 96% and 75%, respectively. Zhu et al. 2020, applied neural network models for the identification of three specific types of promoters in the DNA sequences of species including *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, and *Arabidopsis thaliana*, with classification accuracy ranging from 80% to 98%.

However, it is important to acknowledge that the accuracy of these models is influenced by certain limitations, such as the quantity and quality of the datasets used, as well as class imbalance that occurs when multiple species classes are involved in classification and identification studies. It is well established that the realm of gender classification through computational means in *T. castaneum* remains significantly underexplored. Given the minute sizes of these beetles, human classification becomes challenging, necessitating the fusion of image processing and machine learning techniques to facilitate species and gender identification, which could speed up the ongoing experiments on them. Here, in this planned work, a machine learning-driven intelligent methodology is developed by utilizing microscopic images (from ventral and dorsal viewpoints) to discern and categorize gender disparities. The ultimate objective is to foster the development of intelligent systems within the applied research domain

MATERIALS AND METHODS

Insect rearing and image Acquisition

The primary *T. casteneum* culture was procured from ROSS Lifescience Pvt. Ltd., Pune. It was thereafter cultured in wheat adding 5% yeast at the Zoology department of Savitribai Phule Pune University in an ideal environment at 33 ± °C and 70% relative humidity in a BOD incubator (Halliday et al. 2014). Images of *T. castaneum* at the pupal stage were acquired using digital stereo microscope (Nikon SMZ1270) and MIchrome 6 (6MP) color microscopic camera. The microscopic picture dataset comprises a total of 116 photographs of pupae of two distinct sexes, male and female, with each class including 58 images. These photos were taken from both the dorsal and ventral sides using constant angle and magnification (40X) with a dark background. Figure 1a–d shows representative photos of male and female *T. castaneum* pupa. In this study, machine learning model is train and tested with image size 128 × 128 × 3.

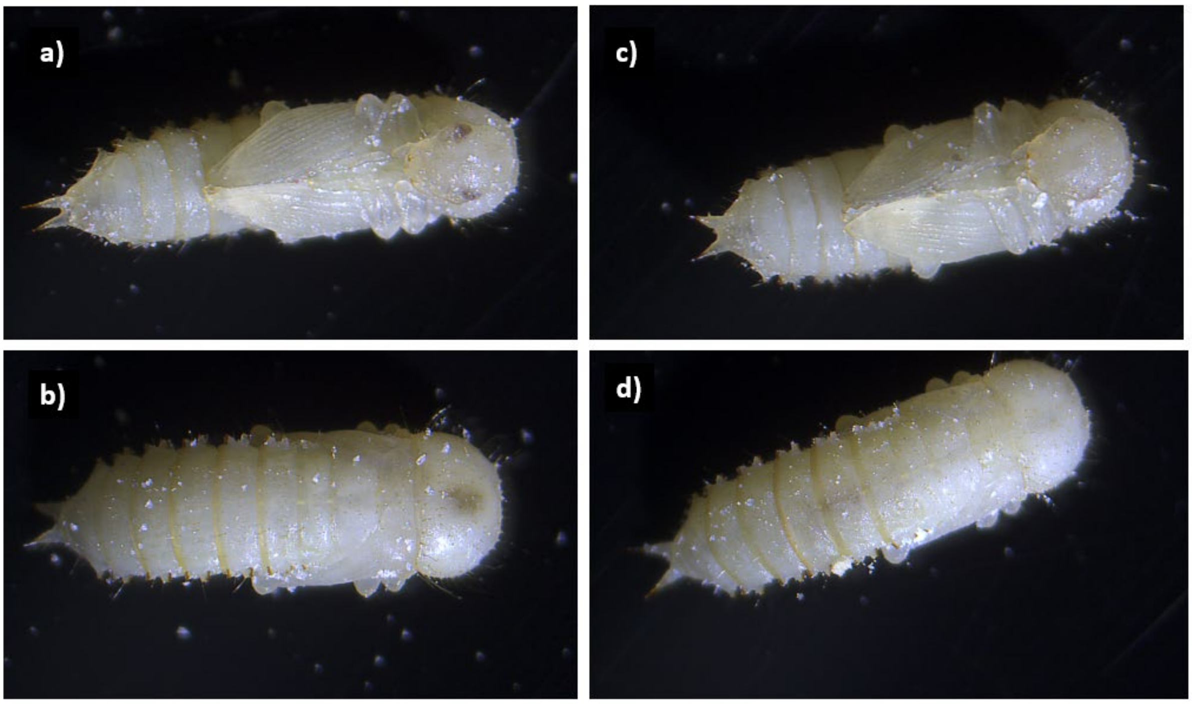

Fig. 1. Specimen photographs of *T. castaneum* pupae under the microscope. a) and b) ventral and dorsal view of female pupa, c) and d) ventral and dorsal view of male pupa.

Methodology

The methodology for the proposed study is depicted in figure 2. The prepared images, as discussed in previous sections, are given as input to a deep neural network. A total of 116 original images were used in this study. To improve model generalization, data augmentation was applied. Each training image was augmented using techniques such as rotation, horizontal and vertical flipping, zooming, and brightness adjustments, resulting in 10 augmented versions per image. This expanded the dataset to a total of 1,160 images. The dataset was then split into training and testing subsets, with 70% of the images used for training the network and 30% used for evaluating its performance. The trained network classifies the images as male and female pupae. Since the algorithm deals with the classification of the pupae, a Convolutional Neural Network (CNN) is applied, which relieves us from the need to locate the pupae in the given image and focuses on recognizing the gender of the pupae. The important feature of CNN is obtaining the abstract features as the image propagates through the deeper layers of the network. Each image is expressed as 128 × 128 matrix.

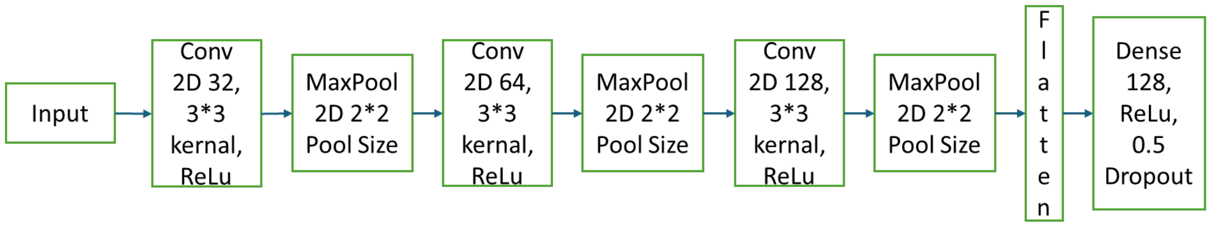

Fig. 2. Architecture of the proposed model.

The first step of the CNN is defining the convolutional layer. Each image matrix is defined with a kernel, which is a set of random numbers also called kernel weights. These weights are adjusted during each training cycle, allowing the kernel to extract significant features. In the applied algorithm, the kernel is of size 2-D (*e.g.*, 2 × 2). The kernel slides over the complete given image matrix, and the corresponding values are multiplied and summed to create a single scalar value. The next layer is the pooling layer, which reduces the feature map generated by the convolutional layer. The commonly used algorithms for the pooling layer are max pooling, min pooling, and Global Average Pooling (GAP). The applied algorithm works on the max 2-D pooling algorithm. In the max pooling layer, a small window (typically 2 × 2) slides across the feature map, and for each window, the maximum value is selected to create a downsampled (pooled) feature map. Following this, the Rectified Linear Unit (ReLU) activation function is applied to introduce non-linearity by replacing all negative values with zero. This function returns zero if the input is negative or the value itself.

Following this, a fully connected layer is designed as the final layer of the CNN network. In this layer, each neuron is connected to all the neurons of the previous layer. The input to the fully connected layer comes from the last pooling or convolutional layer, and its output represents the final output of the CNN.

We chose to develop this model to overcome the problem of overfitting for this specific test case scenario. During the early stages of the development of the solution to the problem we tried to implement VGG16 architecture but faced a serious problem of overfitting, as the image was microscopic and had fine details which were not necessarily needed for the dectection problem as seen in Figure 1. Hence a simpler CNN model proved to be higher performing than more denser models. The architecture of the CNN model includes three convolutional layers followed by max-pooling layers, with the first layer having 32 filters, the second layer having 64 filters, and the third layer having 128 filters, all using a 3 × 3 kernel size and ReLU activation. Each convolutional layer is followed by a 2 × 2 max-pooling layer to reduce the spatial dimensions. After the convolutional and pooling layers, the output is flattened and passed through a fully connected dense layer with 128 units and a ReLU activation function, followed by a dropout layer with a dropout rate of 0.5 to prevent overfitting. We use binary cross entropy as loss function with batch size of 32 which provided us with good number of gradient updates. Figure 2 depicts the detailed architecture of the model. The final output layer uses a softmax activation function to classify the images into the respective categories. This architecture allows the model to learn complex patterns in the images while mitigating overfitting by incorporating dropout and appropriate pooling operations. Reffer to figure 3 for the representation.

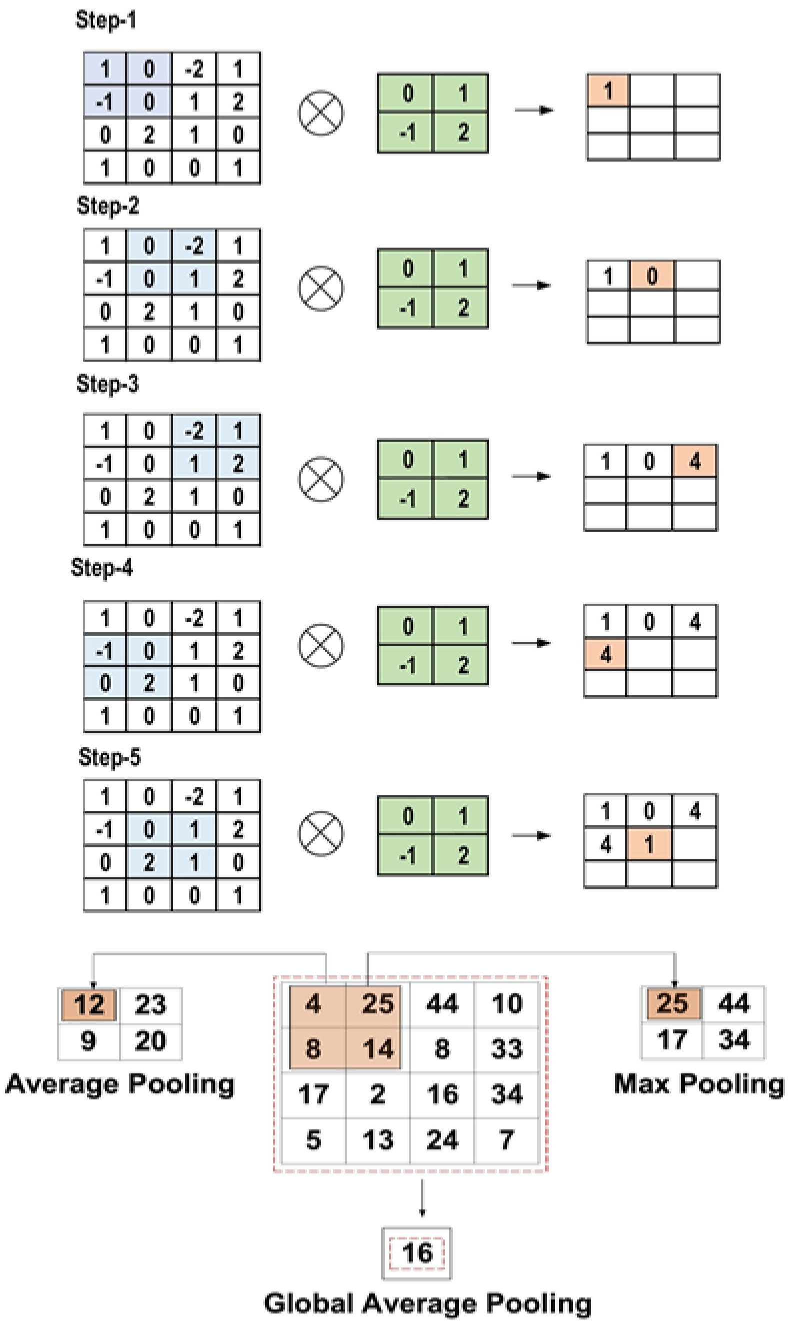

Fig. 3. Procedure and filters for CNN model.

RESULTS

As discussed in the methodology, the first step in identifying key factors for pupal gender classification was the design and implementation of a convolutional layer. The employed model was trained on 812 images in the training dataset and validated and tested on 348 images in the test dataset. To minimize the risk of overfitting with a limited dataset, all images were resized to a resolution of 128 × 128 × 3. This reduction in resolution decreased the input dimensionality, effectively lowering the model's complexity. Additionally, the resizing process enhanced key morphological features by creating a zooming effect, allowing the network to focus on more prominent and consistent traits rather than fine-grained noise, which can contribute to overfitting in small datasets. While this approach improved the model's generalization capability, future work will include the application of interpretability techniques such as Grad-CAM to further validate that the CNN is attending to biologically meaningful regions within the images.

Extracted features by the CNN were passed to a dense network to classify the sex of the pupae. Experimental results showed a high classification accuracy. Notably, it was observed that reducing the image resolution improved the model's ability to differentiate key features. The model was implemented using the TensorFlow library in Python, requiring only 8 MB of memory. It was trained and tested on an Nvidia RTX 3060 using Keras version 3.0.5, CUDA version 11.2, and cuDNN version 8.1, which facilitated efficient processing and inference.

This overall configuration led to a classification accuracy between 97% and 98%, as illustrated by the accuracy and loss plots in figure 4. The model also achieved an F-Score of approximately 0.96 (Fig. 5). These results demonstrate the effectiveness of the proposed model for real-time, automated classification of *T. castaneum* pupae. A comparison of the model's performance with other state-of-the-art techniques is provided in table 2.

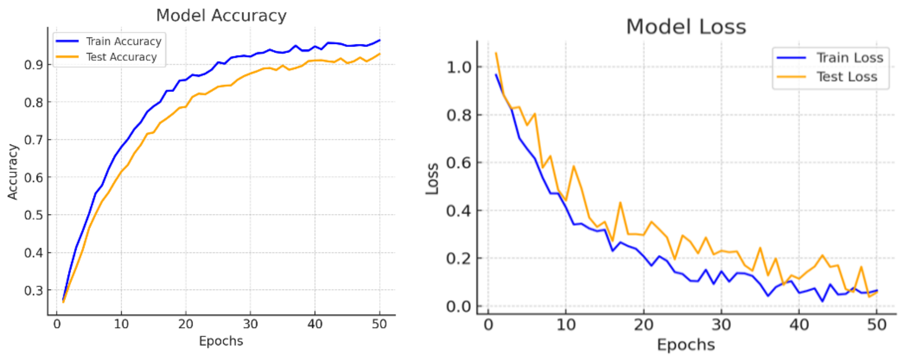

Fig. 4. Plots of accuracy and loss of employed model.

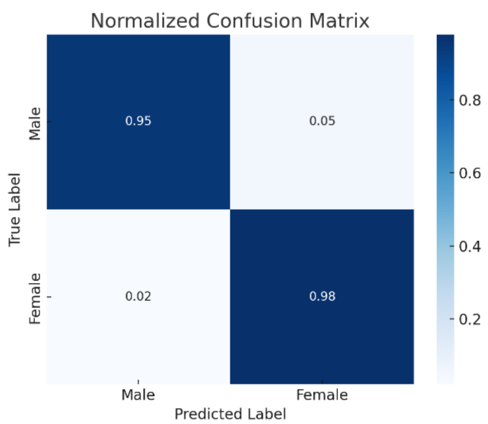

Fig. 5. Confusion matrix for the deployed model.

Table 2. Comparison of results between proposed and other state of the art methods

S. No.	Ref No.	Model Used	Accuracy	Database
1	Themozhi et al. 2019	Deep CNN model for classification	95.97–97.47%	40 pest types from field crops
2	Cannet et al. 2022	Convolutional neural network (CNN) models	33–100%	*Glossina* spp. (tsetse flies)
3	Tannous et al. 2023	Convolutional neural network (CNN) models	93%	The Mediterranean fruit fly *Ceratitis capitata*, and the olive fruit fly *Bactrocera oleae*
4	Alsanea et al. 2022	convolutional neural network (R-CNN)	99%	*Rhynchophorus ferruginous*
5	Dai et al. 2022	Cascade region-based convolution neural networks (R-CNN)	89%	Citrus psyllid
6	Zhao et al. 2022	Convolutional neural network (CNN) models	80–100%	*Ae. vexans*, *Coquillettidia ochracea*, *Mansonia uniformis*, *An. vagus* and *Toxorhynchites splendens*
7	Proposed methodology	Convolutional neural network (CNN) models	97–98%	*Tribolium castaneum*

DISCUSSION

We have already comprehensively discussed the use of deep learning algorithms for species identification in the literature survey section, so we are excluding that information in this section to avoid repetition and considering only CNN-based methods that are similar to our study. Themozhi et al. (2019) classified and detected 40 different crop pests using photos with greater than 95% accuracy. Tannous et al. (2023) and Cannet et al. (2022) used CNN models to automatically identify Glossina spp. (tsetse flies) and food-contaminating beetles, respectively, with accuracy levels of 93% and 3–100%. Alsanea et al. (2022) achieved accuracy values of 99% by applying the R-CNN classification model to photos of stored grain insects and for the auto-detection of red palm weevils, respectively. Cascade region-based convolutional neural networks (R-CNN) were used by Dai et al. (2022) to recognize citrus psyllids from photos with an accuracy of 89%. Zhao et al. (2022) classified mosquito species based on insect pictures using deep learning models like ResNet5016 and YOLO as well as neural network models like CNN. Compared with these results, our experimental set up network performed well with accuracy 97–98% and F1 score 0.67. Nevertheless, identification of gender of pupae in given image challenges minute level feature extraction. The employed network result indicate that cutting-edge systems for tracking the dynamics of pests populations can be created using machine learning methods. This investigation of the population dynamics of important pupae may encourage the creation of novel methods based on a field-based distributed network of automatic monitoring system. This would improve the effectiveness and sustainability of control efforts. With little data pre-processing and a similar design, this approach can be applied to other species of interest.

CONCLUSIONS

This finding can be used in ecological and entomological study as well as a variety of developmental research projects in which determining the gender of *T. castaneum* will be needed. Automation of current process helps to minimise the efforts and time in sex-specific categorisation of *T. castaneum* beetles with 97–98% accuracy. To compensate for the lost characteristics in the photographs and attain more accuracy, we intend to further improve our algorithm by including Image Enhancement techniques in future.

List of abbreviations

ML, Machine learning.
AI, Artificial Inteligence.
ANN, Artificial neural network.
DL, Deep learning.
CNN, Convolution neural network.
ReLU, Rectified Linear Unit.
GAP, Global Average Pooling.
NBAIR, National Bureau of Agricultural Insect Resources.
RF, Random forest.
GB, Gradient boosting (GB).
FAO, Food and Agriculture Organization.
LR, Logistic regression.
SVM, support vector machines (SVM).
DNN, Deep neural network.
NB, Naive Bayes.
KNN, K-nearest neighbour.

RIBBIT, repeat interval-based bioacoustic identification tool.
GLMNET, Generalized Linear Model.

Ackowledgment: Authors are thankful to Dr. Kedar Deobhankar, CEO, Dr. Deepak Phal, MD and Mr. Kishor Raut, In-charge of Entomology laboratory of Ross Lifescience Pvt. Ltd., Pune for providing authentic pure culture of *Tribolium castaneum*, Dr. Milind Sardesai, Prefessor of Department of Botany, SPPU for providing photography instruments and Dr. Atul Kulkarni, Dean-Industry relation of VIIT for providing the help in machine learning tools.

Authors' contributions: AM – Insect rearing, data collection and manuscript writing; CH – Data analysis and interpretation; AS- Study conception and manuscript writing; JB - Data interpretation and manuscript writing; SVP – Manuscript writing

Competing interests: The authors declare that they have no conflicts of interest.

Consent for publication: Not applicable.

Availability of data and materials: All data are provided within the manuscript.

Ethics approval consent to participate: Not applicable.

REFERENCES

Abdelaziz H A, Sallah M, Elgarayhi A, Al-Tahhan FE. 2022. Accurate automatic classification system for 3D CT images of some vertebrate remains from Egypt. JTUMED **16:**632–645. doi:10.1080/16583655.2022.2096391.

Acevedo MA, Corrada-Bravo CJ, Corrada-Bravo H, Villanueva-Rivera LJ, Aide TM. 2009. Automated classification of bird and amphibian calls using machine learning:A comparison of methods. Ecol Inform **4:**206–214. doi:10.1016/j.ecoinf.2009.06.005.

Adamski Z, Bufo SA, Chowański S, Falabella P, Lubawy J et al. 2019. Beetles as model organisms in physiological, biomedical and environmental studies – a review. Front Physiol **10:**319. doi:10.3389/fphys.2019.00319.

Loti ANN, Mohd Noor MR, Chang SW. 2021. Integrated analysis of machine learning and deep learning in chili pest and disease identification. J Sci Food Agric **101:**3582–3594. doi:10.1002/jsfa.10987.

Aladhadh S, Habib S, Islam M, Aloraini M, Aladhadh M et al. 2022. An Efficient Pest Detection Framework with a Medium-Scale Benchmark to Increase the Agricultural Productivity. Sensors **22:**9749. doi:10.3390/s22249749.

Alsanea M, Habib S, Khan, NF, Alsharekh MF, Islam M, Khan S. 2022. A deep-learning model for real-time Red Palm Weevil detection and localization. J Imaging **8:**170. doi:10.3390/jimaging8060170.

Antipov G, Berrani SA, Dugelay JL. 2016. Minimalistic CNN-based ensemble model for gender prediction from face images. Pattern Recognit Lett **70:**59–65. doi:10.1016/j.patrec.2015.11.011.

Bellin N, Calzolari M, Callegari E, Bonilauri P, Grisendi A et al. 2021. Geometric morphometrics and machine learning as tools for the identification of sibling mosquito species of the Maculipennis complex (*Anopheles*). Infect Genet Evol **95:**105034. doi:10.1016/j.meegid.2021.105034.

Bingsohn L, Knorr E, Vilcinskas A. 2016. The model beetle *Tribolium castaneum* can be used as an early warning system for transgenerational epigenetic side effects caused by pharmaceuticals. Comp Biochem Physiol C Toxicol Pharmacol **185:**57–64. doi:10.1016/j.cbpc.2016.03.002.

Bisgin H, Bera T, Ding H, Semey HG, Wu L et al. 2018. Comparing SVM and ANN based machine learning methods for species identification of food contaminating beetles. Sci Rep **8:**6532. doi:10.1038/s41598-018-24926-7.

Bisgin H, Bera T, Wu L, Ding H, Bisgin N et al. 2022. Accurate species identification of food-contaminating beetles with quality-improved elytral images and deep learning. Front Artif Intell **5:**952424. doi:10.3389/frai.2022.952424.

Bjerge K, Mann HM, Høye TT. 2022. Real-time insect tracking and monitoring with computer vision and deep learning. Remote Sens Ecol **8:**315–327. doi:10.1002/rse2.245.

Borba VH, Martin C, Machado-Silva JR, Xavier SC, de Mello FL et al. 2021. Machine learning approach to support taxonomic species discrimination based on helminth collections data. Parasit Vectors **14:**230. doi:10.1186/s13071-021-04721-6.

Bousquet Y. 1990. Beetles associated with stored products in Canada: An Identification Guide. Canadian Government Publishing Centre, Ottawa, Canada.

Cannet A, Simon-Chane C, Akhoundi M, Histace A, Romain O et al. 2022. Wing Interferential Patterns (WIPs) and machine learning, a step toward automatized tsetse (*Glossina* spp.) identification. Sci Rep **12:**20086. doi:10.1038/s41598–022–24522–w.

Chola C, Benifa JV, Guru DS, Muaad AY, Hanumanthappa J et al. 2022. Gender identification and classification of drosophila melanogaster flies using machine learning techniques. Comput Math Methods Med. doi:10.1155/2022/4593330.

Chola C, Muaad AY, Bin Heyat MB, Benifa JB, Naji WR et al. 2022. BCNet: a deep learning computer-aided diagnosis framework for human peripheral blood cell identification. Diagnostics **12:**2815. doi:10.3390/diagnostics12112815.

Dai F, Wang F, Yang D, Lin S, Chen X et al. 2022. Detection method of citrus psyllids with field high-definition camera based on improved cascade region-based convolution neural networks. Front Plant Sci **12:**3136. doi:10.3389/fpls.2021.816272.

Dawson PS. 1964. Age at sexual maturity in female flour beetles, *Tribolium castaneum* and *T. confusum*. Ann Entomol Soc **57:**1–3. doi:10.1093/aesa/57.1.1.

Denell R. 2008. Establishment of *Tribolium* as a genetic model system and its early contributions to evo-devo. Genetics **180:**1779–1786. doi:10.1534/genetics.104.98673.

González-Pérez MI, Faulhaber B, Williams M, Brosa J, Aranda C et al. 2022. A novel optical sensor system for the automatic classification of mosquitoes by genus and sex with high levels of accuracy. Parasit Vectors **15:**1–11. doi:10.1186/s13071-022-05324-5.

Grünwald S, Stellzig J, Adam IV, Weber K, Binger S et al. 2013. Longevity in the red flour beetle *Tribolium castaneum* is enhanced by broccoli and depends on nrf-2, jnk-1 and foxo-1 homologous genes. Genes & Nutrition **8:**439–448. doi:10.1007/s12263-012-0330-6.

Gupta SB, Yadav R, Tyagi PK. 2023. Developing precision agriculture using data augmentation framework for automatic identification of castor insect pests. Front Plant Sci **14:**310. doi:10.3389/fpls.2023.1101943.

Hana U. 2013. Competitive advantage achievement through innovation and knowledge. J Compet **5:**82–96. doi:10.7441/joc.2013.01.06.

Halliday W D, Blouin-Demers G. 2014. Red flour beetles balance thermoregulation and food acquisition via density-dependent habitat selection. J Zool **294:**198–205. doi:10.1111/jzo.12168.

Herndon N, Shelton J, Gerischer L. et al. 2020. Enhanced genome assembly and a new official gene set for *Tribolium castaneum*. BMC Genom **21:**47. doi:10.1186/s12864-019-6394-6.

Kirkeby C, Rydhmer K, Cook SM, Strand A, Torrance MT et al. 2021. Advances in automatic identification of flying insects using optical sensors and machine learning. Sci Rep **11:**1555. doi:10.1038/s41598-021-81005-0.

Kittichai V, Pengsakul T, Chumchuen K, Samung Y, Sriwichai P et al. 2021. Deep learning approaches for challenging species and gender identification of mosquito vectors. Sci Rep **11:**4838. doi:10.1038/s41598-021-84219-4.

Kramarz P, Małek D, Naumiec K, Zając K, Drobniak SM. 2016. Response of development and body mass to daily temperature fluctuations: a study on *Tribolium castaneum*. Evol Biol **43:**356–367. doi:10.1007/s11692-016-9375-6.

Lapp S, Wu T, Richards-Zawacki C, Voyles J, Rodriguez K M, Shamon H, Kitzes J. 2021. Automated detection of frog calls and choruses by pulse repetition rate. Biol Conserv **35:**1659–1668. doi:10.1111/cobi.13718.

Lee YW, Choi JW, Shin EH. 2021. Machine learning model for predicting malaria using clinical information. Comput Biol Med **129:**104151. doi:10.1016/j.compbiomed.2020.104151.

Lei X, Pan H, Huang X. 2019. A dilated CNN model for image classification. IEEE Access **7:**124087–124095. doi:10.1109/ACCESS.2019.2927169.

Liu J, Wang X, Liu G. 2022. Tomato pests recognition algorithm based on improved YOLOv4. Front Plant Sci **13:**814681. doi:10.3389/fpls.2022.814681.

Mahroof RM, Hagstrum DW. 2012. Biology, Behavior, and Ecology of Insects in Processed Commodities. *In*: Hagstrum DW, Phillips TW, Cuperus G (eds) Part I – Ecology of Storage Systems. Stored product protection, Kansas State University Agricultural Experiment Station and Cooperative Extension Service, USA.

Marković D, Vujičić D, Tanasković S, Đorđević B, Ranđić S et al. 2021. Prediction of pest insect appearance using sensors and machine learning. Sensors **21:**4846. doi:10.3390/s21144846.

Michalczyk Ł, Martin OY, Millard AL, Emerson BC, Gage MJ. 2010. Inbreeding depresses sperm competitiveness, but not fertilization or mating success in male *Tribolium castaneum*. Proc Biol Sci **277:**3483–3491. doi:10.1098/rspb.2010.0514.

Motta D, Santos AÁB, Winkler I, Machado BAS, Pereira DADI et al. 2019. Application of convolutional neural networks for classification of adult mosquitoes in the field. PLoS ONE **14:**e0210829. doi:10.1371/journal.pone.0210829.

Ozdemir D, Kunduraci MS. 2022. Comparison of deep learning techniques for classification of the insects in order level with Mobile Software Application. IEEE Access **10:**35675–35684. doi:10.1109/ACCESS.2022.3163380.

Pataki BA, Garriga J, Eritja R, Palmer JR, Bartumeus F et al. 2021. Deep learning identification for citizen science surveillance of tiger mosquitoes. Sci Rep **11**, 4718.doi:10.1038/s41598-021-83657-4.

Patel A, Cheung L, Khatod N, Matijosaitiene I, Arteaga A et al. 2020. Revealing the unknown: real-time recognition of Galápagos snake species using deep learning. Animals **10:**806. doi:10.3390/ani10050806.

Petrescu L, Petrescu C, Oprea A, Mitruț O, Moise G et al. 2021. Machine learning methods for fear classification based on physiological features. Sensors **21:**4519. doi:10.3390/s21134519.

Rabinovich JE, Costa A, Muñoz IJ, Schilman PE, Fountain-Jones NM. 2021. Machine-learning model led design to experimentally test species thermal limits: the case of kissing bugs (Triatominae). PLoS Negl Trop Dis **15:**e0008822. doi:10.1371/journal.pntd.0008822.

Rabinovich JE. 2021. Morphology, life cycle, environmental factors and fitness – a machine learning analysis in kissing bugs (Hemiptera, Reduviidae, Triatominae). Front Ecol Evol **9:**651683. doi:10.3389/fevo.2021.651683.

Rees DP. 2004. Insects of Stored Products. Collingwood, Australia: CSIRO Publishing.

Richards S, Gibbs RA, Weinstock GM, Brown S J, Denell R et al. 2008. The genome of the model beetle and pest *Tribolium castaneum*. Nature **452:**949–955. doi:10.1038/nature06784.

Safavi AE. 2022. Assessing machine learning techniques in forecasting lumpy skin disease occurrence based on meteorological and geospatial features. Trop Anim Health Prod **54:**55. doi:10.1007/s11250-022-03073-2.

Shen Y, Zhou H, Li J, Jian F, Jayas DS. 2018. Detection of stored–grain insects using deep learning. Comput Electron Agric **145:**319–325. doi:10.1016/j.compag.2017.11.039.

Shia WC, Lin LS, Chen DR 2021. Classification of malignant tumours in breast ultrasound using unsupervised machine learning approaches. Sci Rep **11:**1–11. doi:10.1038/s41598-021-81008-x.

Simeone O. 2018. A very brief introduction to machine learning with applications to communication systems. IEEE Trans Cogn Commun Netw **4:**648–664. doi:10.1109/TCCN.2018.2881442.

Sokoloff A. 1974. The biology of *Tribolium* with special emphasis on genetic aspects. Oxford University Press, London, Great Britain.

Spiesman BJ, Gratton C, Hatfield RG, Hsu WH, Jepsen S et al. 2021. Assessing the potential for deep learning and computer vision to identify bumble bee species from images. Sci Rep **11:**1–10. doi:10.1038/s41598-021-87210-1.

Tannous M, Stefanini C, Romano D. 2023. A Deep-Learning-Based Detection Approach for the Identification of Insect Species of Economic Importance. Insects **14:**148. doi:10.3390/insects14020148.

Thenmozhi K, Reddy US. 2019. Crop pest classification based on deep convolutional neural network and transfer learning. Comput Electron Agric **164:**104906. doi:10.1016/j.compag.2019.104906.

Thomso MS. 2014. A selfish gene chastened: *Tribolium castaneum* Medea M 4 is silenced by a complementary gene. Genetica **142:**161–167. doi:10.1007/s10709-014-9763-8.

Veiner M, Morimoto J, Leadbeater E, Manfredini F. 2022. Machine learning models identify gene predictors of waggle dance behaviour in honeybees. Mol Ecol Resour **22:**2248–2261. doi:10.1111/1755-0998.13611.

Wijayaratne LKW, Arthur FH, Whyard S. 2018. Methoprene and control of stored–product insects. J Stored Prod Res **76:**161–169. doi:10.1016/j.jspr.2016.09.001.

Wittek N, Wittek K, Keibel C, Güntürkün O. 2022. Supervised machine learning aided behavior classification in pigeons. Behav Res Methods **55:**1624–1640. doi:10.3758/s13428-022-01881-w.

Xie J, Towsey M, Zhang J, Roe P. 2016. Acoustic classification of Australian frogs based on enhanced features and machine learning algorithms. Appl Acoust **113:**193–201. doi:10.1016/j.apacoust.2016.06.029.

Zare M, Akbarialiabad H, Parsaei H, Asgari Q, Alinejad A et al. 2022. A machine learning–based system for detecting leishmaniasis in microscopic images. BMC Infect Dis **22:**48. doi:10.1186/s12879-022-07029-7.

Zhao DZ, Wang XK, Zhao T, Li H, Xing D et al. 2022. A Swin Transformer-based model for mosquito species identification. Scientific Reports **12(1)**:18664. doi:10.1038/s41598-022-21017-6.

Zhu Y, Li F, Xiang D, Akutsu T, Song J et al. 2021. Computational identification of eukaryotic promoters based on cascaded deep capsule neural networks. Brief Bioinform **22:**bbaa299. doi:10.1093/bib/bbaa299.