Multimedia Tools and Applications (2022) 81:3223–3251
https://doi.org/10.1007/s11042-021-11693-3

1167: DATA SCIENCE ON MULTIMEDIA DATA: CHALLENGES AND APPLICATIONS

An ultra-specific image dataset for automated insect identification

D. L. Abeywardhana[1] · C. D. Dangalle[1] · Anupiya Nugaliyadde[2] · Yashas Mallawarachchi[3]

Received: 22 May 2020 / Revised: 14 June 2021 / Accepted: 19 August 2021 /
Published online: 9 November 2021

Abstract
Automated identification of insects is a tough task where many challenges like data limitation, imbalanced data count, and background noise needs to be overcome for better performance. This paper describes such an image dataset which consists of a limited, imbalanced number of images regarding six genera of subfamily Cicindelinae (tiger beetles) of order Coleoptera. The diversity of image collection is at a high level as the images were taken from different sources, angles and on different scales. Thus, the salient regions of the images have a large variation. Therefore, one of the main intentions in this process was to get an idea about the image dataset while comparing different unique patterns and features in images. The dataset was evaluated on different classification algorithms including deep learning models based on different approaches to provide a benchmark. The dynamic nature of the dataset poses a challenge to the image classification algorithms. However transfer learning models using softmax classifier performed well on the current dataset. The tiger beetle classification can be challenging even to a trained human eye, therefore, this dataset opens a new avenue for the classification algorithms to develop, to identify features which human eyes have not identified.

Keywords Automated insect identification · Limited data · Tiger beetles · Inter-class similarity

✉ D. L. Abeywardhana
lakminiab@stu.cmb.ac.lk

[1] University of Colombo, Colombo, Sri Lanka

[2] Murdoch University, Perth, Australia

[3] Sri Lanka Institute of Information Technology, Malabe, Sri Lanka

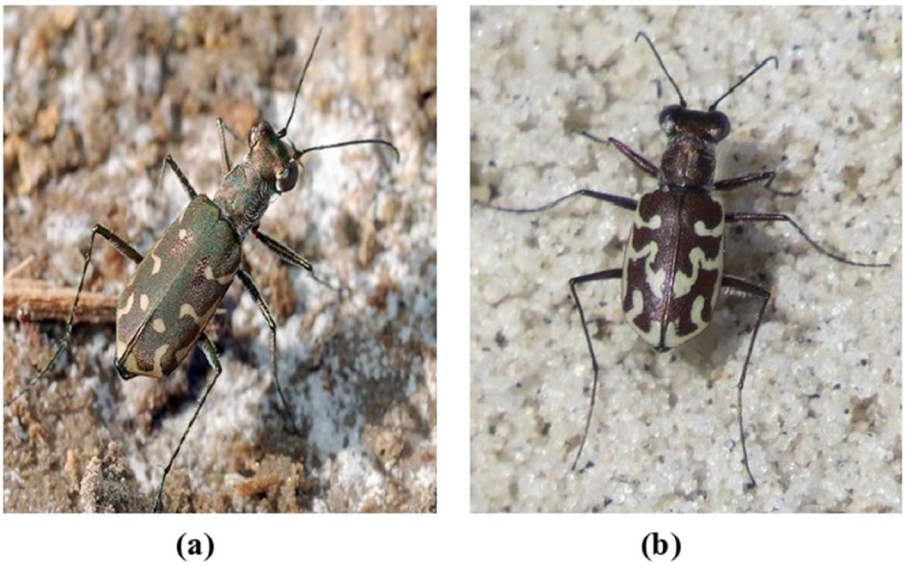


Fig. 1 Visually similar species from two different genera (a) *Myriochila distinguenda* (b) *Calomera angulata*

1 Introduction

There are millions of species on earth, of which 1.2 million have already been formally described [38]. Some species have been identified and described using molecular techniques while morphology and morphometrics have been used to identify others. However, the application of molecular techniques requires considerable expertise knowledge, high cost and time while morphology-based identification poses challenges in identifying cryptic and less abundant taxa [5, 21]. Difficulties in morphological identification are mainly due to the number of visually similar categories which are required to identify inter-class similarity. Therefore, traditional species identification requires expert knowledge of each species in detail and consumes a large amount of time. The lack of expert knowledge, cost of species identification and high time consumption has shown importance towards image classification using Artificial Intelligence [35].

The rapid improvement in machine learning models especially convolutional neural networks (CNN) image classification has shown dramatic performances. These models are capable of identifying features that are visible and non-visible to the naked eye to classify objects [7, 28]. However, these models require a large number of training data from each class to learn the features, which are organized to form a feature vector, an arbitrary length, vector which collects all the properties that are useful in describing the object under analysis [54] to enhance the model for better performance [29, 44, 50]. As a result, Ultra-specific classifications such as insect species classification which have inter-genus similarities in morphology, require an extremely large amount of image data to gain high validation accuracy [22, 41, 58]. Therefore, this is a challenging task for automated models to extract features to identify these insect species who are with enormous inter-class similarities (Fig. 1).

This paper introduces an ultra-specific data set on tiger beetles, which is one of the first datasets that focus on entomological specific image data. The tiger beetle images require expert knowledge to classify them as they were taken from various image sources and various environments. Each image has different noise levels, different size and a different zoom level on the beetle. Therefore, classifying these images using computational models are challenging as general computational models are not capable of handling noise and variations of the images to classify. Hence, the dataset would provide a good platform for new image analysis methods for real-world ultra-specific image classification, with highly different noise levels and variations. In this paper, benchmarks are also set to show the responses to each machine learning algorithm.

Almost all the dominant image classification datasets used in computer vision tend to have a uniform (balanced) and copious distribution of images across object categories. As an example, the well-known, highly diverse ImageNet dataset [12] consists of more than 14 million images belonging to more than 20,000 categories with a typical category consisting of several hundred images. However, the feature vectors identified for such datasets are not substantially suitable for highly specific scenarios such as identification of zoological image sets of species belonging to the same family. To become more specific for animal species identification by machine learning, the Caltech-UCSD Birds-200-2011 Dataset [13] which is comprised of 11,788 bird images and Stanford Dogs Dataset which contain images of 120 breeds of dogs from around the world, has been built using images and annotation from ImageNet for the task of fine-grained image categorization [30]. Further, iNaturalist species classification and detection dataset, consisting of 859,000 images from over 5000 different species of plants and animals [24] and IP102, a large-scale dataset specifically constructed for insect pest recognition which contains more than 75,000 images belonging to 102 categories [56] have been developed. All above-mentioned datasets have at least more than 10,000 images in their collection, but this is often impractical, or even impossible, as in many real-world scenarios some species are more abundant and easier to photograph, and some are rare and endemic and not visible in the common environment frequently.

In some scenarios, the image collections in the databases have been taken in proper light condition, in specific dimensions (angle) with less background noise. Marques et al. [37], describes ant genera identification methodology using an image dataset as an online database on ant biology, the Antweb (http://www.antweb.org). The images in the above dataset were taken using powerful tools like Automontage and Leica microscope so that all the images in the dataset have the same (high level) quality and also the dimensions of the images have been restricted to three views as frontal, lateral and dorsal. Larios et al. [32], proposes a methodology for the identification of stonefly larvae, an insect inhabiting water. To capture high-quality images of stonefly larvae a mechanical apparatus has been built where each specimen is manually inserted into acrylic and then pumped through a tube. During this process, an infrared detector positioned along the tube detects the passage of the specimen and a side fluid jet captures the specimen using a QImaging MicroPublisher 5.0 RTV 5 megapixel colour digital camera attached to a Leica MZ9.5 high-performance stereo microscope at 0.63x magnification. Illumination has been provided by gooseneck light guides powered by Volpi V-Lux 1000 cold light sources and diffusers have also been installed in this apparatus to reduce glare, specular reflections and hard shadows. Gutierrez et al. [20], describes a pest detection and identification method for tomato plants where the dataset has been generated using both manually and automatically obtained images collected in a specific environment (Mendelu's cultivation chamber). In the manual procedure, all the images have been captured by AP-3200 t-PGE and monochrome camera DataCam

2016R using a standard display system connected to a PC. Different types of lenses and lighting systems have been also used when necessary. In the automated procedure, two GigE UI-5240CP cameras have been set up and using Raspberry Pi 3 microcontroller both the camera and the movement structure are controlled. The microcontroller is programmed to take pictures both with and without artificial lighting in different directions and angles. However, the above set-ups can be impractical for some real-world problems when dealing with images taken in a normal environment deprived of proper light intensity with background noise. Furthermore, the rarity of the species prevents access to a large number of images for each class. Therefore, machine learning models struggle to achieve accurate classifications results. Hence it is proven that class imbalance, availability of limited images with inter-class similarity and availability of different quality images with huge background noise are properties of a real-world scenario where computer vision models require to deal with [24].

1.1 Research contribution

The paper introduces a real-world tiger beetle image dataset that has been classified to genera level. This dataset uses images from various sources ranging from high-definition cameras to images scraped from the internet. The data set is created by an expert entomologist to test various machine learning models. Base-line machine learning models are used for experiments and are described as a benchmark for the dataset. Furthermore, this shows a path for future work required for a highly-specified machine learning model for ultra-specific insect classification.

2 The Tiger beetle dataset[1]

The classification dataset of tiger beetles (Coleoptera, Carabidae, Cicindelinae) from Sri Lanka is comprised of a limited number of images. This dataset consists of images of tiger beetles belonging to six genera of tribes Cicindelini (ground-dwelling tiger beetles) and three genera of tribe Collyridini (arboreal tiger beetles) [1] (Fig. 2).

There are significant morphological differences between species of tribe Cicindelini (Fig. 3a) and tribe Collyridini (Fig. 3b) to distinguish each tribe using visuals (Table 1).

Sources: [10, 15, 43].

The different genera of tribe Cicindelini have distinctive elytral patterns that enable the identification of genera. Further, small variations of elytral patterns may also be seen in species within the same genera.

[1] https://github.com/lakminia/Tiger-beetle-image-dataset; https://fos.cmb.ac.lk/opendata/tigerbeetles/

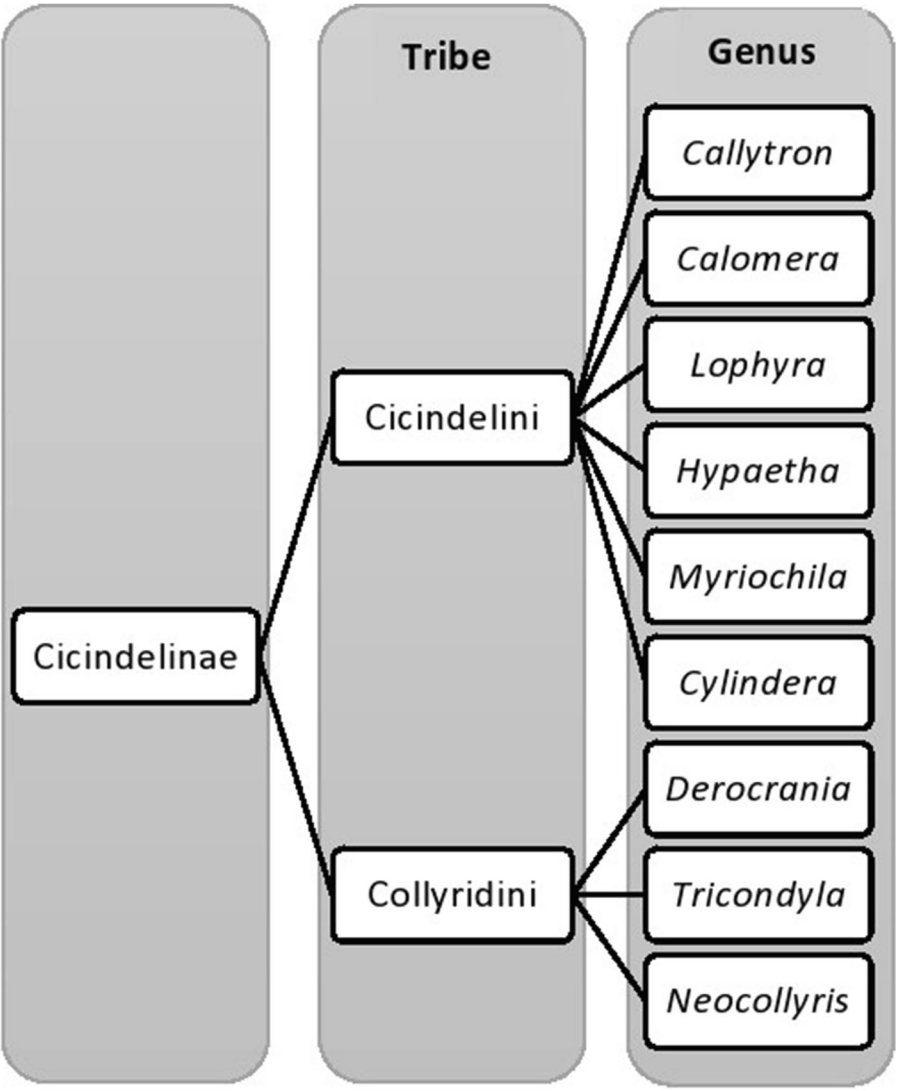


Fig. 2 Tiger beetle tribes and genera of the image dataset

2.1 Tiger beetle genera used in the dataset

2.1.1 Tribe Cicindelini

Genus *Callytron* The dataset of the present study used images of only *Callytron limosa* of genus *Callytron*. *Callytron limosa* has a unique elytral maculae pattern. Maculae have reduced to a narrow white continuous lateral band that extends from the humeral angle to the apical spine (Fig. 4).

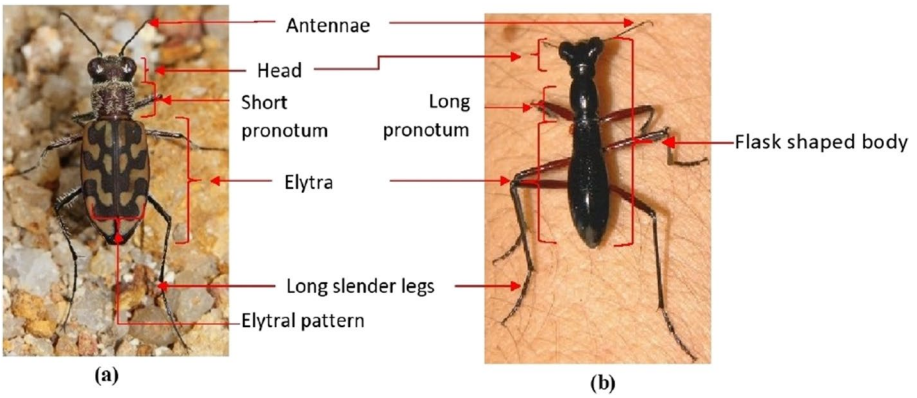


Fig. 3 (a) *Lophyra catena* of tribe Cicindelini, (b) *Tricondyla* of tribe Collyridini

Table 1 Morphological differences between tribe Cicindelini and tribe Collyridini

Ground dwelling tiger beetles (Cicindelini)	Arboreal tiger beetles (Collyridini)
The body is not conical/ flask-shaped.	The body is conical/ flask-shaped.
Short pronotum.	Long, slender pronotum.
Elytra consist of different elytral patterns.	No elytral patterns.
Episterna of the metasternum not very narrow and not strongly furrowed.	Episterna of the metasternum very narrow and strongly furrowed.
Tarsal pads are present only in pro-thoracic legs of males.	Tarsal pads present in all legs of both sexes (*Collyris, Tricondyla*).

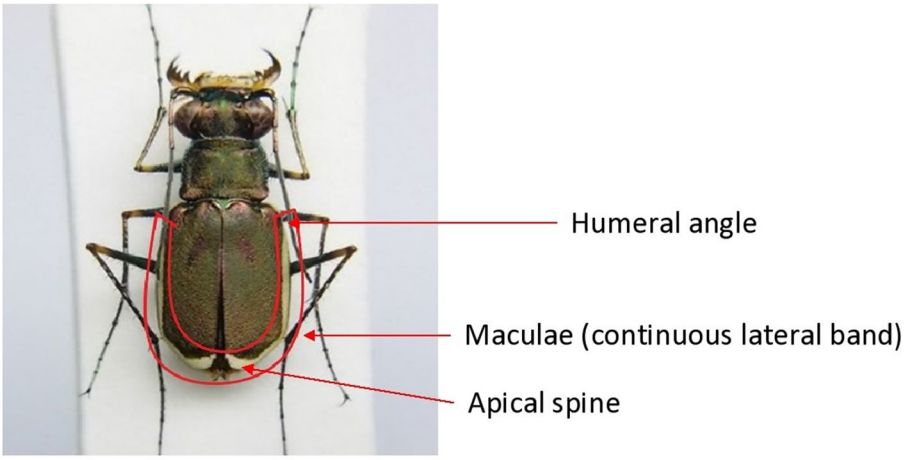


Fig. 4 *Callytron limosa*

Genus *Calomera* The dataset utilized images of the species *Calomera angulata* and *Calomera cardoni* of the genus *Calomera*. Both species have a complete humeral lunule on

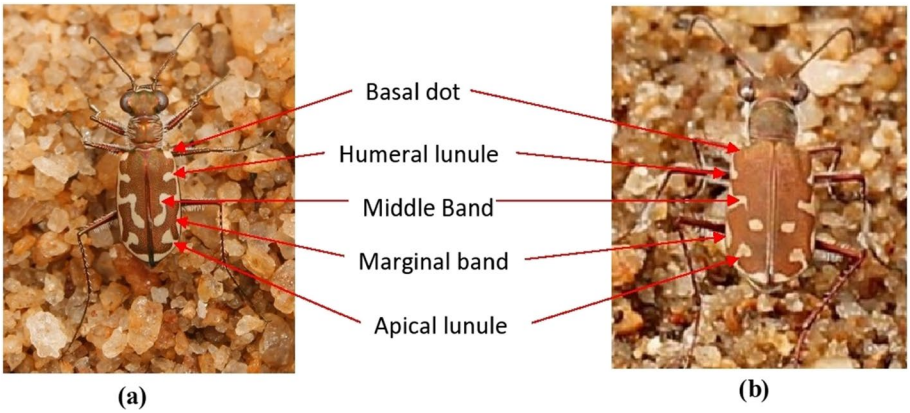


Fig. 5 (a) *Calomera angulata;* (b) *Calomera cardoni*

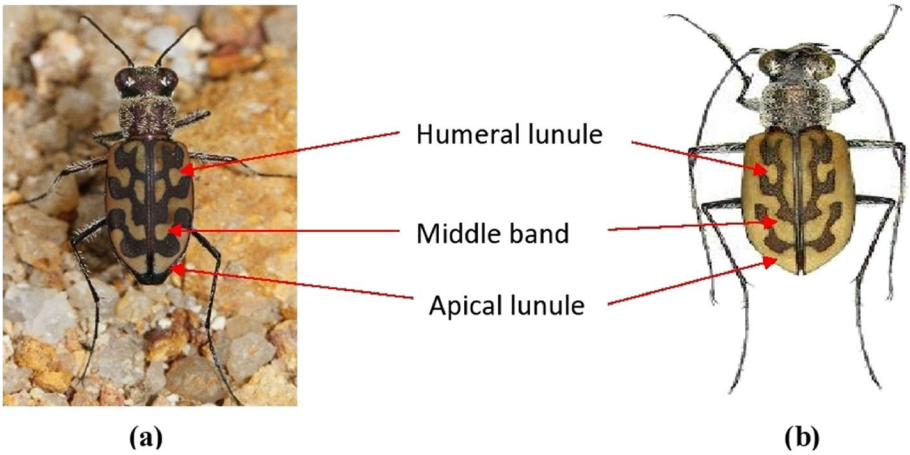


Fig. 6 (a) *Lophyra catena*; (b) *Lophyra cancellata*

elytra but can be differentiated using the middle band of elytra [3]. In *C. angulata* the middle band consists of a transverse portion that concaves anteriorly, while in *Cybister cardoni* the transverse portion of the middle band does not concave anteriorly. The terminal portion of the middle band of *C. angulata* is broadly connected to the transverse portion while it is separate or narrowly connected in *C. cardoni* [11]. Further, *C. angulata* has a dark bronze coloured head and pronotum, with greenish punctures and white markings. White colour in elytra extends from the shoulders to the apex, with an interruption before the apical lunate patch (Fig. 5a). The elytral surface of *C. cardoni* is dark brown or metallic brown with yellowish-white discontinuous lateral lunules. The middle band terminates medially to form a separate spot (Fig. 5b) [52].

Genus *Lophyra* *Lophyra catena* and *Lophyra cancellata* of genus *Lophyra* are morphologically very similar. However, in *Lophyra catena* the genae of the head are setose, while in *Lophyra cancellata* they are glabrous (smooth and without hairs) [22]. Humeral lunule,

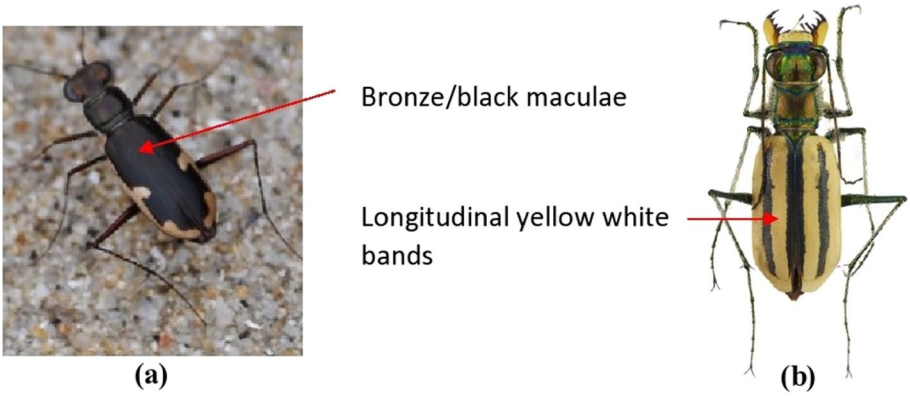


Fig. 7. (a) - *Hypaetha biramosa;* (b) - *Hypaetha quadrilineata.*

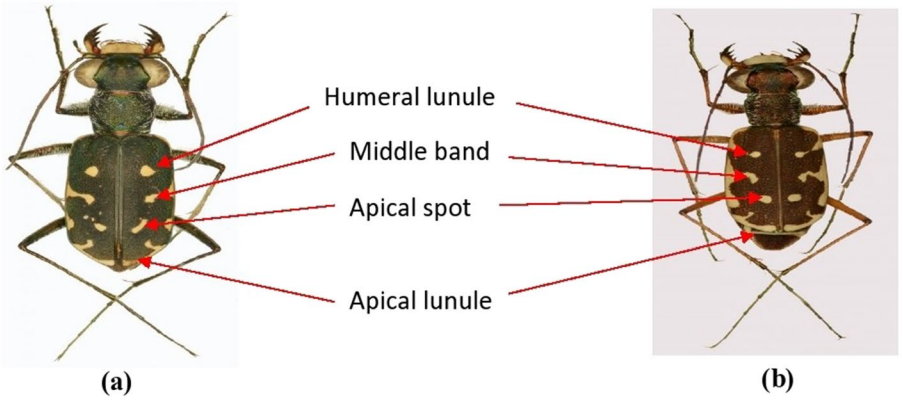


Fig. 8 (a) *Myriochila distinguenda;* (b) *Myriochila* (*Monelica*) *fastidiosa*

the middle band and apical lunule of both the species of this genera are clearly visible and similar to each other. They can be used to distinguish genus *Lophyra* from other genera of tribe Cicindelini (Fig. 6).

Genus *Hypaetha* Images of two species *Hypaetha biramosa, Hypaetha quadrilineata* were included in this genus. More than 90% of the elytral surface of *Hypaetha biramosa* is covered with dark brown/black maculae medially and a yellowish-white continuous band extends from humeral angle to the apex. This band invaginates around half of the length of the elytra to separate the dark brown/black maculae into two sections (Fig. 7a).

Hyla quadrilineata maculae exist as two longitudinal yellow-white bands on an elytron (Fig. 7b). The abdomen of *H. quadrilineata* is setose laterally, while it is glabrous in *H. biramosa* [22].

Genus *Myriochila* *Myriochila distinguenda* and *Myriochila* (*Monelica*) *fastidiosa* represent the genus *Myriochila* of Sri Lanka. Both species are characterized by the standard pattern of maculae on the elytra consisting of humeral lunule, middle band and apical

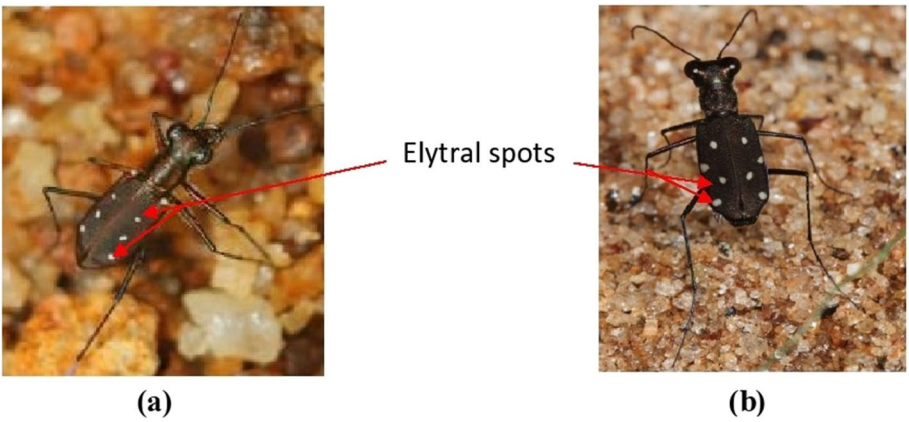


Fig. 9 (a) *Cylindera* (*Ifasina*) *waterhousei;* (b) *Cylindera* (*Ifasina*) *labioaenea*

lunule. In *Myriochila distinguenda* the basal portion of the elytral humeral lunule is separated while in *Myriochila* (*Monelica*) *fastidiosa* it is joined to the apical end of the humeral lunule (Fig. 8).

Genus *Cylindera* The species belonging to this genera are characterized by yellowish-white spots on elytra and do not have humeral, apical lunules or marginal/ middle bands. However, the position, number and shape of the elytral spots in each species of *Cylindera* may vary distinguishing the species from one another (Fig. 9) [3, 15].

2.1.2 Tribe Collyridini

The main differences between the genera of tribes Collyridini and Cicindelini are found related to the body size, body colour and characteristics of pronotum. In the present study, three genera of tribe Collyridini were considered, *Tricondyla* and *Derocrania* of sub-tribe Tricondylina, and *Neocollyris* of sub-tribe *Collyridina*.

Genus *Derocrania* Species of this genus are smaller and slender than species of genus *Tricondyla* of the same sub-tribe. Eyes are prominent and pronotum more elongate, slender and narrow at the top. Elytra is elongate and almost widened behind or sometimes in some species very strongly raised behind or almost flat, with very variable sculpture. Legs are long and slender (Fig. 10) [10, 15].

Genus *Tricondyla* Species of genus *Tricondyla* are characterized by a large head that is deeply excavated between the eyes. The neck behind the eyes is parallel-sided. Elytra narrowed in front and dilated and very convex behind. The pronotum is almost parallel sideded, broad, constricted in front and behind, sometimes a little convergent but without a collum in front (Fig. 11) [10, 15].

Genus *Neocollyris* Species of this genus are small and slender with bright blue elytra that are almost punctured. However, elytra vary in colour, size and sculpture and some species have strongly rugose sculptures in the middle. Many species of this genus are hard to

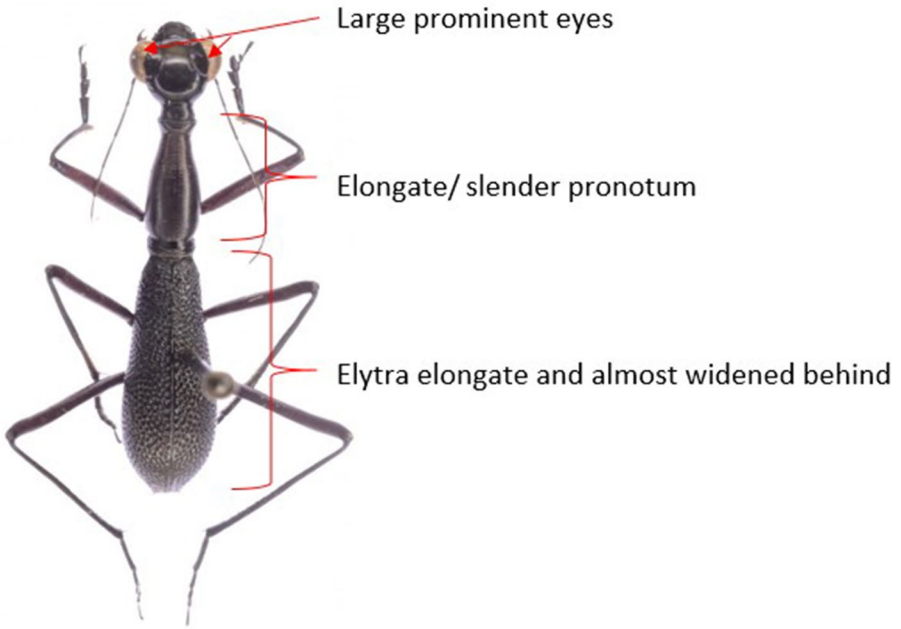
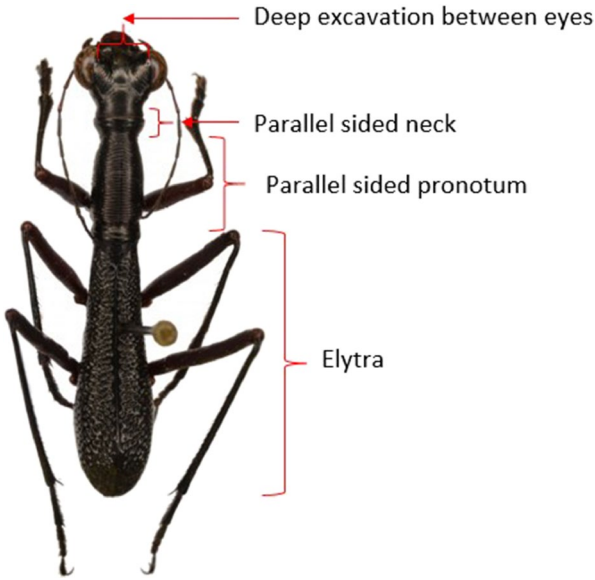


Fig. 10 *Derocrania schaumi*

Fig. 11 *Tricondyla granulifera*



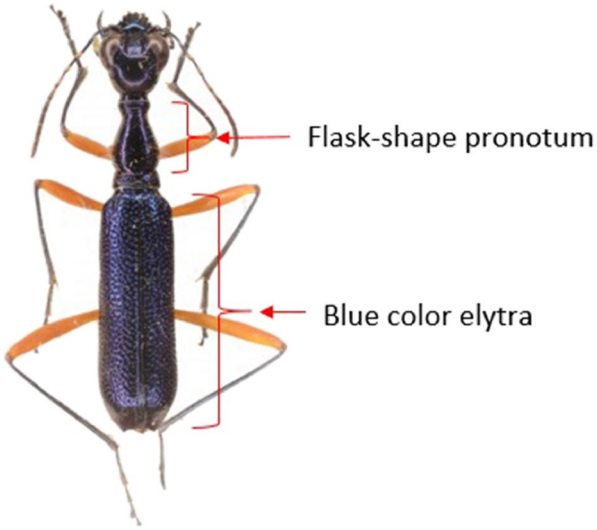


Fig. 12 *Neocollyris bonelli*

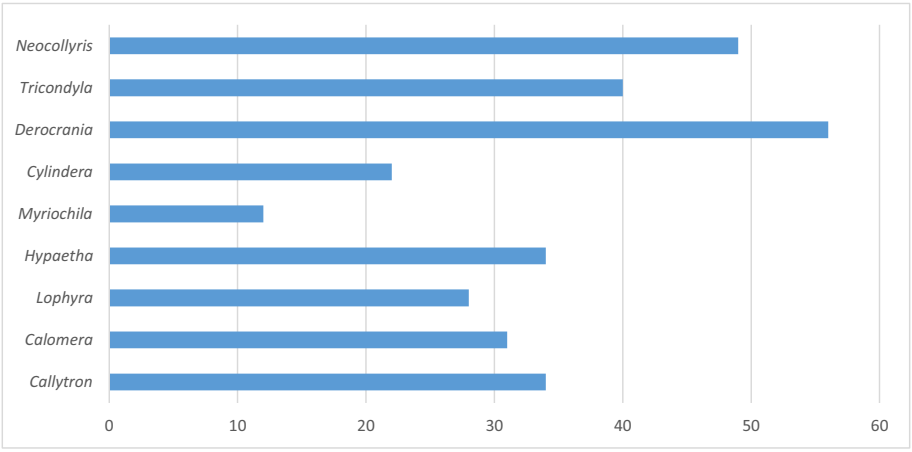


Fig. 13 Image samples collected for each genus

identify from one another. However, differences are found in the shape of the pronotum which is generally flask-shaped (Fig. 12) [15, 39].

2.2 Image collection

The dataset comprises of images taken during field investigations, from wildlife and nature photographers using different camera types with different image quality, tiger beetle publications and websites. However, due to the rarity of certain species and endemicity, only a limited number of images were collected from each genus (Fig. 13).

Fig. 14 Images obtained from different sources (a) DSLR (macro lens) camera capture; (b) smartphone captures; (c) image taken from a website (genus- *Derocrania)*

Fig. 15. Images of species taken in different angles (genus-Derocrania)

2.3 Features of the images

2.3.1 Image sources

Images were collected from different sources

- Mobile phone captures
- DSLR camera captures
- Images from websites/blogs (iNaturalist, Shnao, Project Noah, JungleDragon, Thailandwildlife, My Shot Gallery of Bengkulu)

Therefore, the images have different quality (Fig. 14), views and angles (Fig. 15).

2.3.2 Image background

Furthermore, salient parts in each image have extremely large variations in size (Fig. 16). Subsequently, most of these images are with noisy backgrounds and inconsistency as most of these images were taken in the actual habitat environment of the species (Fig. 17).

2.3.3 Image diversity based on colour saturation

In order to understand certain features of the images and to obtain an overall idea about the colour intensity distribution, combined colour histograms of images for each genus

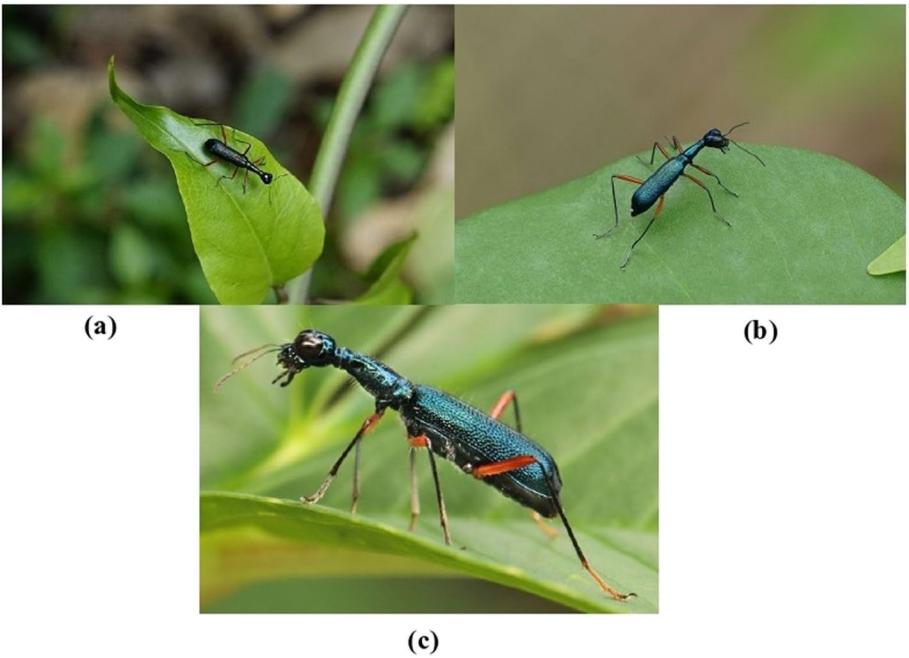


Fig. 16. (a) Species occupying very little space; (b) species occupying a part of it; (c) species occupying most of the image (genus-*Neocollyris*)

Fig. 17. Images with background noise

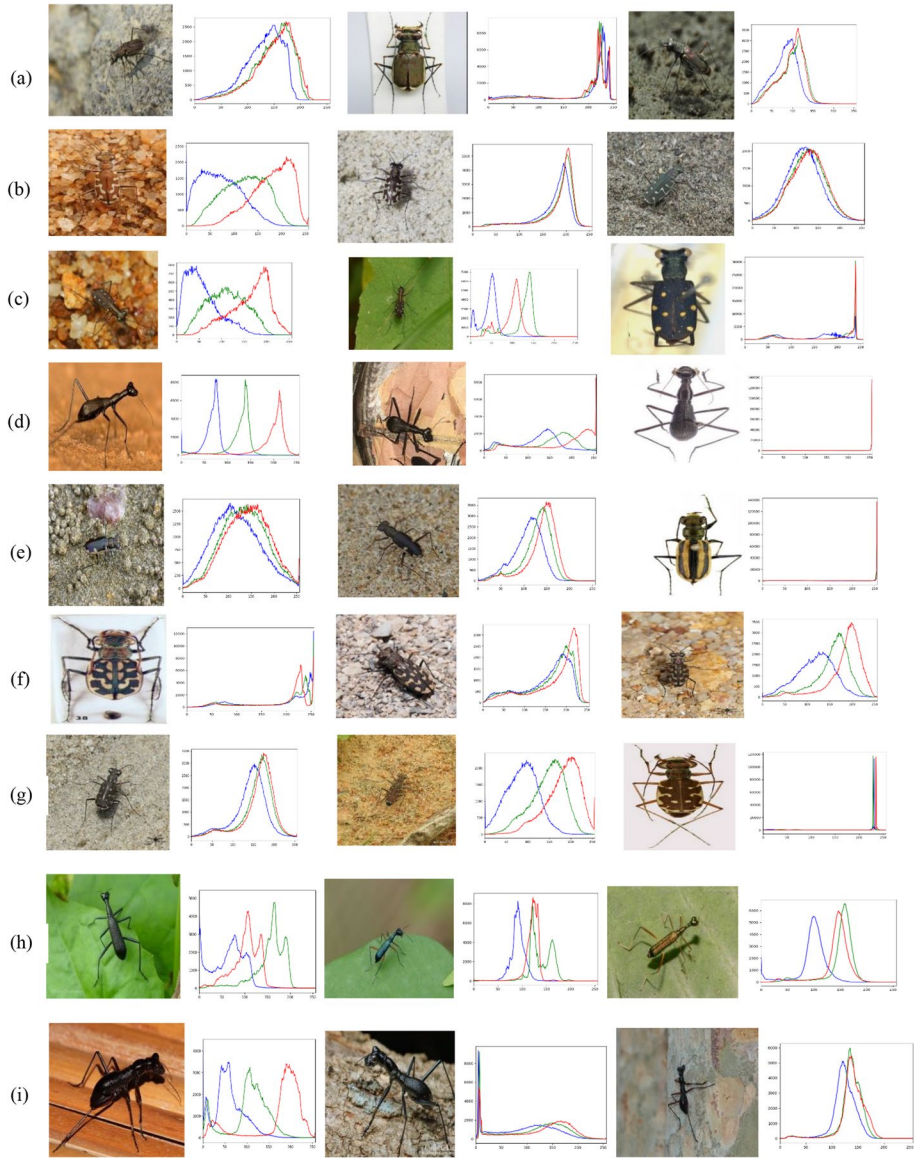


Fig. 18. Sample images of each genus and corresponding colored histogram related to the specific image ((a) *Callytron*; (b) *Calomera*; (c) *Cylindera*; (d) *Derocrania*; (e) *Hypaetha*; (f) *Lophyra*; (g) *Myriochila*; (h) *Neocollyris*; (i) *Tricondyla*)

of the dataset were plotted (Fig. 18). Each graph is plotted with pixel values (ranging from 0 to 255) on X-axis and the corresponding number of pixels in each image on Y-axis [17, 42]. The graphs depict variations in colour contrast, brightness, the intensity distribution of images due to background noise of images resulting from their environment.

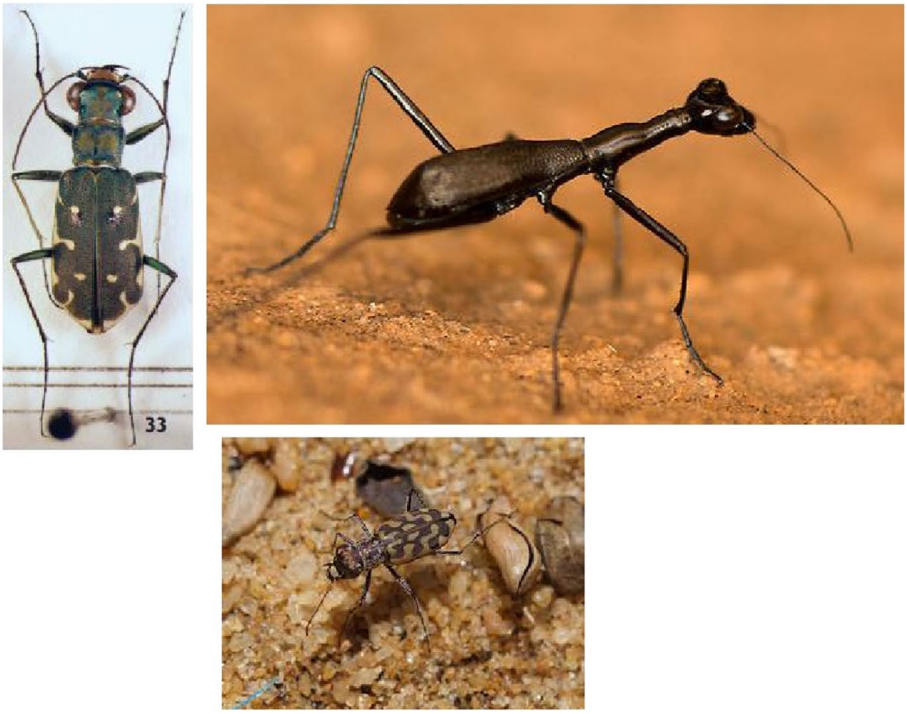


Fig. 19 Images with different sizes

2.3.4 Image diversity based on image size

Since the images were collected from different sources the resolution and dimensions of the image collection varied from one image to another. Thus, the sizes of the images in the dataset ranged from (136 × 319 pixels) to (5184 × 3459 pixels) (Fig. 19).

Limitations in the dataset

- Imbalanced image count - The dataset is highly imbalanced, with certain genera having more images than others.
- Background noise in images.
- Availability of a limited number of images due to endemicity and rareness of species.
- Different quality images.
- Images taken in different angles/ dimensions.

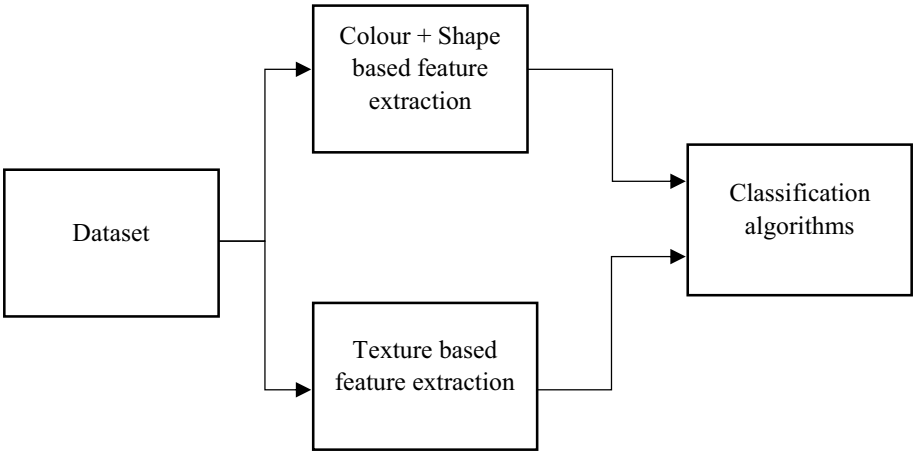


Fig. 20 Feature extraction and classification

3 Classification algorithms

3.1 Classical machine learning classifiers

As the current task is related to supervise learning category where it is required to identify the dependencies between the target prediction output and the input features, and as the image dataset comprises of nine genera of subfamily Cicindelinae, multi-class classification should be done. Before passing images to multi-class classifications, feature extraction needs to be done as preprocessing steps to perform the classification task more effectively. When deciding about the features that could quantify these nine genera of tiger beetles, variations in texture, colour patterns (Fig. 18) and shape were considered as global feature vectors, which describe the image as a whole to generalize the entire object. Texture defines the consistency of patterns and colours in an object/image. Since tiger beetles can be identified to genus level from combinations of patterns and colour variations of elytra, texture can be selected as a promising feature to extract unique features of each class (which helps to identify class separately) in the present dataset. However, when considering the shape as a single vector for feature extraction, it is less likely to produce good results since classes of the current dataset have many attributes in common due to inter-class similarities (Fig. 1). Therefore we combined both colour and shape as a single feature descriptor in order to describe the image more effectively. Texture based feature extraction can be done based on different concepts such as Ant Lion Optimizer (ALO) [2, 55], Grey Level Co-occurrence Matrix (GLCM) [16]. For the current process, the texture was quantified using haralick texture features [23] which are calculated from Grey Level Co-occurrence Matrix, (GLCM), a matrix that counts the co-occurrence of neighbouring grey levels in an image. Further, colour variations and shape were quantified using colour histograms [4] and Hu Moment [25] respectively. After the feature selection process data were passed through eight supervised, multi-class classification algorithms [33] (Fig. 20).

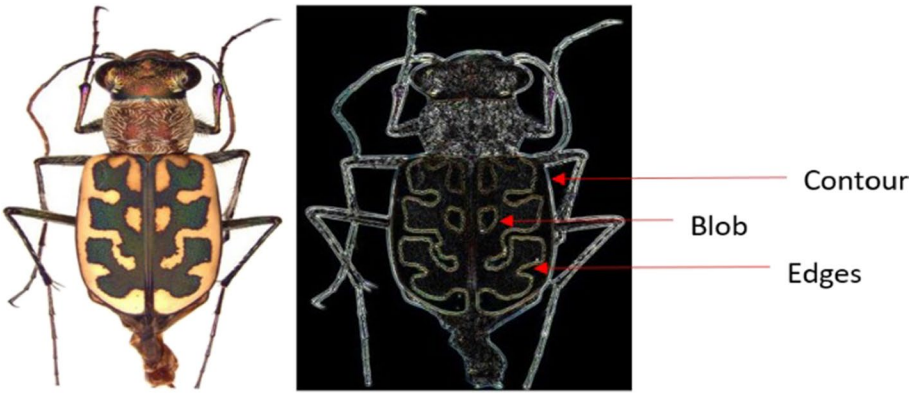


Fig. 21. Entomological view against computer vision based descriptors. Genus: Lophyra

I. Logistic Regression - As the current scenario is a multiclass classification problem a multinomial logistic regression approach was implemented by using Broyden–Fletcher–Goldfarb–Shanno (bfgs) algorithm as the optimization algorithm.
II. Linear Discriminant Analysis (LDA) - The objective of LDA is to project the dataset onto a lower-dimensional space with good class-separability to avoid overfitting and also reduce computational costs.
III. KNeighborsClassifier - It is important to find the best K (neighbour) value to obtain the optimum accuracy from KNN for the dataset. Therefore, to get the best possible fit for the images set $K = 1$ was selected.
IV. DecisionTree Classifier - Classification tree was selected for the current problem. This algorithm performs variable screening/feature selection implicitly which is an advantage for feature extraction. Tree-based learning algorithms are considered as the best and frequently used supervised learning methods as they assist predictive models with high accuracy, stability and affluence of interpretation. Unlike linear models, these algorithms are able to map non-linear relationships well.
V. RandomForest Classifier - RandomForest is a classification algorithm that evolves from decision trees. It consists of a collection of a large number of individual decision trees. For the current problem, 25 decision trees were used in the forest. To classify a new instance, each decision tree provides a classification for input data and this classification algorithm collects the classifications and chooses the most voted prediction as the result [36].
VI. GaussianNB - When Naïve Bayes is extended to real-valued attributes by assuming Gaussian distribution it is called Gaussian Naive Bayes.
VII. SVM - Support Vector Machine algorithms are based on the idea of decision planes (hyperplane) which outline decision boundaries that help to classify the data points. SVM algorithms use a set of mathematical functions that are defined as the kernel. The purpose of the kernel is to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions. RBF kernel function was used for this scenario [14].
VIII. ExtraTreesClassifier - An ensemble learning technique. This is very similar to random forest classifier and differs only from the manner of construction of the decision trees [51]. The number of decision trees used in the classifier was 10.

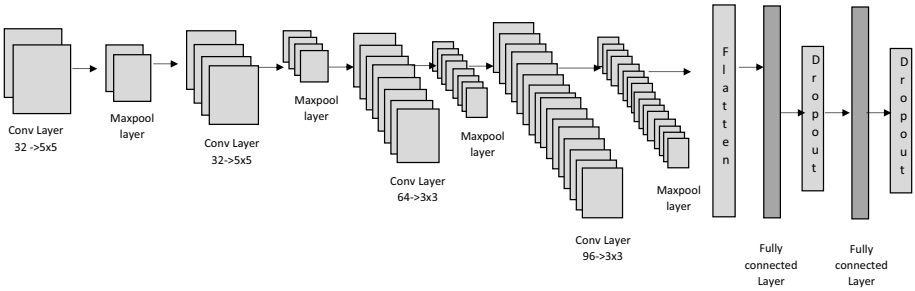


Fig. 22. Architecture of CNN model with 4 convolutional layers with 2 fully connected layers

Above mentioned traditional machine learning methodologies (classification algorithms) use trivial structures to handle limited data and computing units. When the target objects have rich meanings with high background noise and inconsistency, the performance and generalization ability of the above models are insufficient [14]. Therefore, it is important to move to deep learning methodologies which can attain higher accuracy. The ultimate goal of deep learning is to represent an image in a hierarchical manner by increasing complexity per each layer as seen in natural world objects where combination of many compositional units increase the diversity of the resulting structure (Fig. 21).

3.2 Deep learning techniques

The pre-processing required for deep learning models are much lower when compared with traditional classification algorithms. Because these approaches work by extracting features from images which eliminate the need of manual feature extraction. At the same time arrangement of these models play a major role in designing and creating new architectures for performance improvement. Therefore, model evaluations instigated from simple structures to complex form, since convolutional neural networks arguably considered a black box evaluation [6]. Based on strategy these models were evaluated in three stages.

- CNN models without using pre-trained weights.
- CNN models fine-tuned with transfer learning approach (using pre-trained weights).
- Extract features from pre-trained CNN models, and classify using SVM.

Further, as the target dataset mostly consists of images that were taken in normal habitat environments which consist of a fair amount of background noise and inconsistency, tiger beetles dataset was evaluated on Faster Region-based Convolutional Network method (Faster R-CNN) [8].

I. **CNN models without using pre-trained weights**

- **CNN model with 4 convolutional layers and 2 fully connected layers**

This convolutional neural network model comprises 6 hidden layers including 4 convolutional layers and 2 fully connected layers. The first and second convolutional layers of the model contain 32 kernels of size 5×5, while the third convolutional layer contain 64 kernels of size 3×3 and the final convolutional layer contain 96 kernels of 3×3 (Fig. 22).

- **AlexNet model**

This is a well-known CNN model which follows a standard neural network architecture of stacked and connected layers. It comprises eight layers that need to be trained, five convolutional layers followed by three fully connected layers, as well as a max-pooling layer [31, 49, 57].

- **SqueezeNet model**

This model features a great reduction in parameter space and computational complexity through channel projection bottleneck (squeeze layers). Further, similar to residual networks, the model uses indemnity-mapping shortcut connections which allow for stable training of deep network models. This model is comprised of "fire modules", where the input map is first fed through a bottlenecking channel-projection layer and then divided into two-channel sets. The first channel set is expanded through a 3×3 convolution and the second one through channel projection. The final convolution map is globally average-pooled into a 512-vector and then fed to a fully connected layer with 2048 units [19].

II. **CNN models that are fine-tuned with transfer learning strategy (using pre-trained weights) and classified using softmax classifier**

Transfer learning has become a conventional procedure (meaning that a classifier is already trained on a large-scale dataset like ImageNet dataset before the actual training begins). Here the classifier will only be fine-tuned to the specific classification problem by training a small number of high-level network layers proportional to the amount of available problem-specific training data [40]. Since the model has already learned certain features from a large dataset this method is more suitable for these kinds of datasets which are with a limited number of data. Transfer learning (TL), is commonly used in the computer vision area which allows building more precise models efficiently [26, 46, 48].

Therefore as the initial phase, it is required to get pre-trained models with weights loaded. Eight deep learning networks were selected as target models.

- AlexNet
- InceptionV3
- ResNet
- SqueezeNet
- VGG16
- VGG19
- DenseNet121
- Inception-ResNet v2

For pre-trained weights, we used weights gained by training the above models on top of ImageNet dataset to extract general classification-supporting features like curves and edges from several front layers. As the proposed architecture, we removed the final fully connected layer from each model and then used the remaining portion of the model as a feature extractor for the current dataset. These extracted features are called "Bottleneck features".

Then the fully connected layers were developed according to the target dataset below and the extracted bottleneck features were used to get the classes as outputs for the

problem. Finally, results were classified using softmax (multi-nomial logistic regression) classifier.

III. **Extract features from pre-trained CNN models, and classify using SVM**

Image features from different CNN architectures pre-trained on the ImageNet dataset were submitted to a linear support vector machine classifier, which is trained on the target problem. For this strategy, we selected the same Convolutional Neural Network models as target pre-trained deep learning models, which were used for earlier evaluation (AlexNet, InceptionV3, ResNet50, SqueezeNet, VGG16, VGG19, DenseNet121, Inception-ResNet v2). The main reason to select support vector machine algorithm as the classifier was that SVMs are more appropriate for small datasets where the number of dimensions are greater than the number of examples [53]. Further, SVMs are memory-efficient since they only use a subset of training points or support vectors and they generalize well to high dimensional spaces as well.

IV. **Faster-RCNN**

The advancement of CNN models based on regional paradigm have able to address the object detection and classification process [45]. The main objective of the regional paradigm was to improve the precision of object detection while improving overall model performance. Since Faster-RCNN was identified as the most state-of-the-art version of RCNN models, it was selected to evaluate the current tiger beetle image dataset. In Faster R-CNN each input image is passed to a ConvNet which returns feature maps for the image. Then the region proposed network (RPN) is applied on the above feature maps to get object proposals and the region of interest (ROI) pooling layer brings all the object proposals to a specific size. Finally, these proposals are passed through a fully connected layer to classify and predict the bounding boxes of each image [34]. The target dataset was evaluated on top of Faster-RCNN based on two base convolutional neural networks which are VGG16 and ResNet50 [8].

4 Results and discussion

4.1 Experimental setup

The original images set was divided into a training and test set where around 75% of the total images in each genus were put into a training set while the rest (25%) were placed into a test set. As deep artificial neural networks require a large corpus of training data to learn effectively and to avoid over-fitting, the image quantity was increased by expanded the dataset artificially by image augmentation [44, 47]. Image augmentation was conducted using a combination of multiple processing, such as random rotation, shifts, shear and flips, etc. For this, we used Augmentor [9], a Python package designed to aid the augmentation and artificial generation of image data for machine learning tasks (Fig. 23).

The augmentation procedures used to increase the amount of available training data are listed below.

- Horizontal flipping
- Vertical flipping

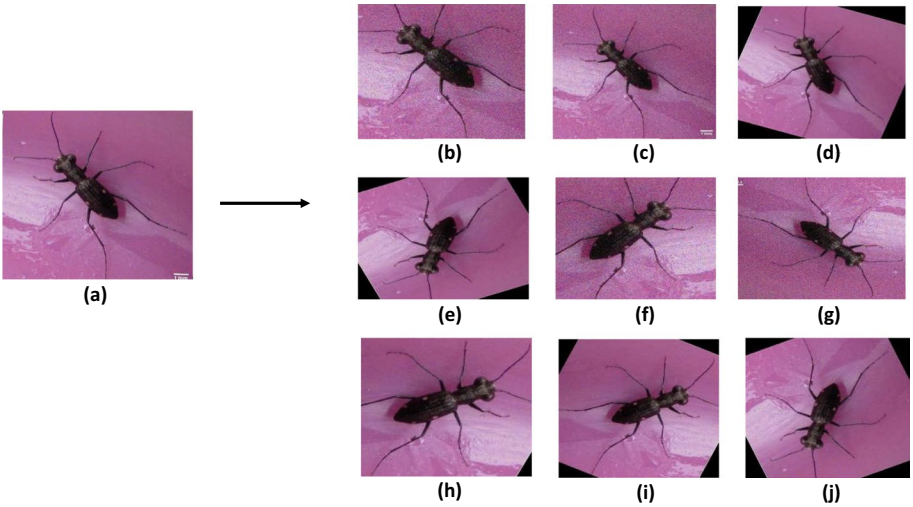


Fig. 23 Image Augmentation (a) Original image. (b)Zoomed-in image. (c) Zoomed-out image. (d) 45^0 rotated image to right (e) 90^0 rotated image to left (f) Vertically flipped and zoomed-in image (g) 180^0 rotated image to right (h) 90^0 rotated image to right (i) 90^0 rotated and zoomed-out image (j) 90^0 rotated image to left

- Zooming - Zoom in to an image at a random location within the image, while maintaining its size. The amount by which the image is zoomed is a randomly chosen value.
- Rotate - 90^0, 45^0,180^0

Table 2 Separation of images to train and test set and creation of 2 datasets as Approach I and Approach II

Tribe	Genus	Approach I		Approach II	
		Original training image quantity	Original testing image quantity	Training image quantity (after augmentation)	Testing image quantity after augmentation)
	Callytron	29	6	100	35
	Calomera	39	5	100	35
Cicindelini	*Lophyra*	33	4	100	35
	Hypaetha	37	4	100	35
	Myriochila	37	4	100	35
	Cylindera	20	7	100	35
	Derocrania	37	21	100	35
Collyridini	*Tricondyla*	35	7	100	35
	Neocollyris	44	11	100	35

After data augmentation process the training image quantity increased up to 100 images per genus and test image quantity to 35 for each genus, which is still a low image quantity for a deep convolutional neural network. However, increasing image quantity to thousands

using image augmentation is also unreasonable since it will again subject the model to over-fitting.

Finally, two approaches were initiated, one with original, imbalanced data (Approach I) and the other with augmented, balanced data (Approach II) (Table 2).

In order to perform a benchmark on the dataset, the previously mentioned classification algorithms were trained on the datasets of Approach I (original, imbalanced data) and Approach II (augmented, balanced dataset) while deep learning models were trained on only the augmented, balanced dataset. All the classification algorithms were executed on lenovo ideapad 300 machine having inter® Core™ i7 -6500U CPU @ 2.50GHz-2.59Ghz, 8 GB RAM with 64bit Operating System and deep learning models were executed on Google colaboratary using GPU on Jupiter notebook [18].

4.2 Test results and discussion

According to Table 3, classification algorithms using texture-based feature extraction is more suitable for the classification of the current dataset than classification algorithms using colour and shape. The highest test accuracy of 49.43% was gained from the linear discriminant analysis algorithm based on texture-based feature extraction. Further, higher validation accuracies were gained from all the classification algorithms which were trained on the original imbalanced images dataset than the augmented balanced images dataset. Therefore, it is possible to state that the orientation of the images supports the machine learning models for classifications and orientation is a key factor for image classification using a machine learning approach.

When considering deep learning models which were trained without using pre-trained weights on augmented balanced tiger beetle images set, SqueezeNet gave the highest validation accuracy of 63.81% (Table 4).

According to Table 5, the validation accuracies have increased when the complexity of the model decreases. Therefore, it can be assumed that model complexity is not always the solution for better accuracy. At the same time [27] it can be elaborated that CNN models with fewer parameters have several advantages such as require less communication across servers during distributed training and more feasible to deploy on FPGAs and other hardware with limited memory.

When results depicted in Table 5 compared with Tables 6 and 7, it clearly emphasizes that deep learning models which were trained using the transfer learning approach have able to provide more promising results than the deep learning models which were trained without using pre-trained weights on augmented balanced tiger beetle images set. Since the models which were trained using transfer learning approach have already trained on some features using a larger dataset (ImageNet), the baseline performances of the deep learning architectures have improved. Further computation time taken to train the above models (Table 6, Table 7) has also become low due to knowledge transfer. However, transfer learning models classified using softmax classifier have provided slightly better validation accuracies than transfer learning models which were classified using SVM classifier.

Results depicted in Table 8 corroborates that although the Faster-RCNN model is an object detection network, which is also composed of a feature extraction network (a pre-trained CNN), the complex architecture is unable to provide better accuracy for the present real-world tiger beetle dataset. The main reason for Faster-RCNN models to perform inferiorly on top of the current scenario is mainly due to the data limitation. Due

Table 3 Test accuracies of classification algorithms

Feature vector		Colour+Shape								Texture							
Classification algorithms		LR	LD	KNN	DT	RF	GNB	SVM	ETC	LR	LD	KNN	DT	RF	GNB	SVM	ETC
Original IIS	TA(%)	39.13	8.70	37.68	26.08	42.02	31.88	15.94	39.13	44.94	49.43	30.34	42.69	44.94	47.19	31.46	43.82
	CT(s)	0.81	1.02	0.33	0.33	0.64	0.05	0.67	0.25	0.56	0.76	0.06	0.56	0.55	0.03	0.65	0.19
Augmented BIS	TA(%)	36.82	13.96	33.97	26.34	38.09	23.17	11.11	37.77	26.35	39.37	26.67	31.11	40.63	27.61	19.68	37.46
	CT(s)	2.36	2.53	0.33	1.41	5.28	0.13	8.16	0.34	2.31	2.49	0.34	1.38	5.15	0.23	8.09	0.36

Abbreviations: Original IIS, Original imbalanced images set; Augmented BIS, Augmented balanced images set; TA(%), Test accuracy(%); CT(s) Computation time (seconds)

Table 4 Test accuracies of Convolutional Neural Network models

Deep learning model	Val top 1 accuracy	Val top 2 accuracy	Val top 3 accuracy
AlexNet	34.28%	39.68%	57.46%
CNN with 4 convolutional layers and 2 fully connected layers	42.1%	50.25%	61.28%
SqueezeNet	63.81%	69.21%	79.68%

Table 5 Model complexity with validation accuracies of different Convolutional Neural Network models

Deep learning model	Number of trainable parameters	Computational time	Val top 1 accuracy
AlexNet	28,067,625	1 h 32 min	34.28%
CNN with 4 convolutional layers and 2 fully connected layers	889,673	1 h 24 min	42.1%
SqueezeNet	543,387	1 h 05 min	63.81%

Table 6 Test accuracies of pre-trained CNN + SVM models

Pre-trained CNN + SVM	Total number of parameters	Number of trainable parameters	Computational time	Val top 1 accuracy
VGG19	20,029,001	20,029,001	45 min 25 s	19.048%
InceptionV3	21,821,225	21,602,169	51 min 11 s	50.89%
AlexNet	16,008,457	9,449,481	38 min 20s	55.24%
DenseNet121	7,046,729	6,858,505	50 min 53 s	59.36%
SqueezeNet	727,113	727,113	49 min 02 s	60.20%
ResNet50	23,606,153	23,451,209	41 min 42 s	61.93%
Inception-ResNet v2	54,350,569	54,290,025	45 min 54 s	69.45%
VGG16	14,719,305	4617	47 min 01 s	69.52%

Table 7 Model complexity with validation accuracies of different transfer learning models using Softmax classifier

Deep learning model	Number of trainable parameters	Computational time	Val top 1 accuracy
VGG19	20,029,001	50 min 06 s	23.08%
InceptionV3	889,673	52 min 48 s	42.83%
DenseNet121	6,858,505	45 min 51 s	64.44%
SqueezeNet	727,113	50 min 34 s	64.44%
Inception-ResNet v2	54,290,025	49 min 43 s	69.03%
ResNet50	23,451,209	34 min 11 s	69.52%
AlexNet	16,002,633	41 min 48 s	71.75%
VGG16	4617	48 min 11 s	75.87%

Table 8 Test accuracies of Faster-RCNN models

Base model	Computation time	Validation accuracy
VGG16	34 h 24 min	42.5%
Resnet50	36 h 04 min	34.58%

to limited data in the training set the models have been unable to identify exact feature vectors to locate the object precisely from the background area.

However, the highest test accuracy for the target tiger beetle dataset was gained from the transfer learning VGG16 model which was 75.87% (Table 7). In this model as an additional modification all layers above fully connected layers were not trained (freeze) and only the fully connected layers were considered as trainable parameters. When using weights of a pre-trained model on top of the current model, the complexity of the CNN model has not become a significant factor for the improvement of validation accuracies (Table 7). From the above approach, even the models with the least number of trainable parameters have gained the optimum accuracy for the highly diverse limited image tiger beetle dataset. The results depict that better accuracies for vision-based datasets with limited data can be gained from machine learning models with fewer trainable parameters.

Validation accuracies gained for the tiger beetle image dataset from classification algorithms were significantly low. Although the image quantities were increased using image augmentation, there were no significant differences among validation accuracies before and after image augmentation. However, the accuracy gained from texture-based feature extraction was slightly higher than the colour and shape-based feature extraction method. Further, when considering the results gained from different deep learning models it clarifies that complex deep learning architectures such as Faster-RCNN which is also recognized as state-of-the-art deep learning-based object detection and classification method, have provided lower validation accuracy with high computation cost for the target dataset (Table 8). Therefore, when contemplated about computation cost along with validation accuracies it depicts that advanced architectures such as Faster-RCNN models are incapable of providing promising results for a limited image dataset which have images of high background noise with diverse resolution. However, higher validation accuracies have been gained from transfer learning approaches using softmax classifier. This is mainly because the pre-trained models have trained over ImageNet dataset which contains images of tiger beetles as a single collection under the beetle hierarchy. Hence the models have learned some features from the larger dataset previously which has aided to identify features in the new dataset. Therefore, the baseline performance of the model also improves due to knowledge transfer. At the same time, due to transferring of knowledge from a larger data source (ImageNet) and also due to reduction of trainable parameters, the computation costs spent to train target dataset using deep learning CNN models with transfer learning approach has become low when compared to computation costs spent to train target dataset using other deep learning models [12]. For the current study highest accuracy was gained from VGG16 transfer learning model. However, due to reasons like inter-class similarity and availability of limited data, machine learning models have gained validation accuracies below 80%, which has plenty of room for further improvement. The above experiments were conducted to provide a benchmark and the above validation accuracies depict the diversity of the tiger beetle image dataset.

5 Conclusion

The present article reveals a highly precise and diverse dataset created for tiger beetles (Coleoptera, Cicindelinae) of Sri Lanka which contains images of beetles of nine genera of two tribes. Images were tested on different classification algorithms based on different feature extraction techniques (texture, colour, shape), and deep learning models with and without using pre-trained weights. Further, an augmented and balanced dataset was evaluated by extracting features from pre-trained CNN models and classifying them using SVM classifier. Additionally, target image dataset was also evaluated on Faster-RCNN, a state-of-the-art deep learning approach with more complex architecture. The study produced a dataset that was highly specific and challenging for renowned machine learning models and beneficial for automated zoological studies since this is a real-time dataset, consisting of features like high unevenness of data, a limited number of data with huge background noise and having lots of inter-class similarities in images. As benchmark results from the above models, optimum test accuracy was obtained from transfer learning models where the VGG16 transfer learning model gave the highest test accuracy of 75.87%. Attempting to improve accuracies for these types of datasets will assist to overcome limitations in image processing techniques and expand machine learning knowledge. Further, these attempts will be beneficial in expanding biodiversity monitoring systems on a global scale.

Acknowledgements We would like to acknowledge Dr. Agasthya Thotagamuwa of Charles Sturt University, Australia for providing set of Tiger Beetle images required for the study. This work was funded by the National Science Foundation of Sri Lanka [Grant number RG/2017/EB/01].

Data availability Tiger beetle dataset which used for the evaluation can access through following URL https://github.com/lakminia/Tiger-beetle-image-dataset.

Code availability Sample machine learning models can access through following URL https://github.com/lakminia/Tiger-beetle-image-dataset.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

1. Abeywardhana DL, Dangalle CD, Mallawarachchi YW (2019) Automated identification of Coleoptera, Cicindelinae in Sri Lanka by machine learning. In: Asia-Pacific conference 2019 Association for Tropical Biology and ConservationAt: Sri Lanka. Pp 78–79
2. Abualigah L, Shehab M, Alshinwan M, Mirjalili S, Elaziz MA (2020) Ant lion optimizer: a comprehensive survey of its variants and applications. Archives of Computational Methods in Engineering. https://doi.org/10.1007/s11831-020-09420-6
3. Acciavatti RE, Pearson DL (1989) The tiger beetle genus Cicindela (Coleoptera, Insecta) from the Indian subcontinent. Annals of the Carnegie Museum 58:77–353

4. Ali H, Lali MI, Nawaz MZ, Sharif M, Saleem BA (2017) Symptom based automated detection of citrus diseases using color histogram and textural descriptors. Comput Electron Agric 138:92–104. https://doi.org/10.1016/j.compag.2017.04.008
5. Alvarez AJ, Hernandez-Delgado EA, Toranzos GA (1993) Advantages and disadvantages of traditional and molecular techniques applied to the detection of pathogens in waters. Water Sci Technol 27:253–256. https://doi.org/10.2166/wst.1993.0354
6. Bouvrie J (2006) Notes on convolutional neural networks. https://doi.org/10.1016/j.protcy.2014.09.007
7. Caramazza P, Boccolini A, Buschek D, Hullin M, Higham CF, Henderson R, Murray-Smith R, Faccio D (2018) Neural network identification of people hidden from view with a single-pixel, single-photon detector. Sci Rep. https://doi.org/10.1038/s41598-018-30390-0
8. Cheng B, Wei Y, Shi H, Feris R, Xiong J, Huang T (2018) Revisiting RCNN: on awakening the classification power of faster RCNN. In: lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)
9. Bloice DM, Stocker C, Holzinger A (2017) Augmentor: An Image Augmentation Library for Machine Learning. The Journal of Open Source Software. https://doi.org/10.21105/joss.00432
10. Dangalle CD (2018) The forgotten tigers: the arboreal tiger beetles of Sri Lanka. Journal of the National Science Foundation of Sri Lanka 46:241–252. https://doi.org/10.4038/jnsfsr.v46i3.8477
11. Dangalle CD, Dangalle NK, Pallewatta N (2017) Historical and Curent records on the Tiger Beetle,Calomera angulata, Fabricius of Sri Lanka. Journal of Biology and Nature 7:91–99
12. Deng J, Dong W, Socher R, Li L-J, Kai Li, Li Fei-Fei (2010) ImageNet: A large-scale hierarchical image database
13. Englert B, Lam S (2011) The Caltech-UCSD Birds-200-2011 dataset. https://doi.org/10.3182/20090902-3-US-2007.0059
14. Evgeniou T, Pontil M (2001) Support vector machines: Theory and applications. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 2049 LNAI:249–257. https://doi.org/10.1007/3-540-44673-7_12
15. Fowler WW (1912) The Fauna of British India including Ceylon and Burma. Coleoptera. General Introduction and Cicindelidae and Paussidae. Taylor & Francis
16. Gebejes A, Master EM, Samples A (2013) Texture characterization based on Grey-level co-occurrence matrix. In: Conference of Informatics and Management Sciences
17. Gevers T, Smeulders AWM (1999) Color-based object recognition. Pattern Recogn 32:453–464. https://doi.org/10.1016/S0031-3203(98)00036-3
18. Google Colab (2020) Welcome to Colaboratory - Colaboratory. In: Getting Started - Introduction. https://colab.research.google.com/notebooks/intro.ipynb
19. Grm K, Struc V, Artiges A, Caron M, Ekenel HK (2018) Strengths and weaknesses of deep learning models for face recognition against image degradations. IET Biometrics 7:81–89. https://doi.org/10.1049/iet-bmt.2017.0083
20. Gutierrez A, Ansuategi A, Susperregi L, Tubío C, Rankić I, Lenža L (2019) A benchmarking of learning strategies for Pest detection and identification on tomato plants for autonomous scouting robots using internal databases. Journal of Sensors. https://doi.org/10.1155/2019/5219471
21. Hamsher SE, LeGresley MM, Martin JL, Saunders GW (2013) A comparison of morphological and molecular-based surveys to estimate the species richness of Chaetoceros and Thalassiosira (Bacillariophyta), in the bay of Fundy. PLoS One 8:e73521. https://doi.org/10.1371/journal.pone.0073521
22. Hansen OLP, Svenning JC, Olsen K, Dupont S, Garner BH, Iosifidis A, Price BW, Høye TT (2020) Species-level image classification with convolutional neural network enables insect identification from habitus images. Ecology and Evolution 10:737–747. https://doi.org/10.1002/ece3.5921
23. Haralick RM, Dinstein I, Shanmugam K (1973) Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics 6:610–621. https://doi.org/10.1109/TSMC.1973.4309314
24. Horn G Van, Aodha O Mac, Song Y, Cui Y, Sun C, Shepard A, Adam H, Perona P, Belongie S (2018) The iNaturalist species classification and detection dataset. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 8769–8778
25. Hu MK (1962) Visual pattern recognition by moment invariants. IRE Transactions on Information Theory 8:179–187. https://doi.org/10.1109/TIT.1962.1057692
26. Huh M, Agrawal P, Efros AA (2016) What makes ImageNet good for transfer learning? arXiv preprint arXiv:1608.0861
27. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K (2016) SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:160207360. https://doi.org/10.1007/978-3-319-24553-9
28. Jangblad M (2018) Object detection in infrared images using deep convolutional neural networks. Uppasala University,Sweden

29. Kamilaris A, Prenafeta-Boldú FX (2018) A review of the use of convolutional neural networks in agriculture. J Agric Sci 156:312–322. https://doi.org/10.1017/S0021859618000436
30. Khosla A, Jayadevaprakash N, Yao B, Li F-F (2011) Novel dataset for fine-grained image categorization: Stanford dogs. In: Proceedings of the IEEE International Conference on Computer Vision
31. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. pp. 1097–1105
32. Larios N, Deng H, Zhang W, Sarpola M, Yuen J, Paasch R, Moldenke A, Lytle DA, Correa SR, Mortensen EN, Shapiro LG, Dietterich TG (2008) Automated insect identification through concatenated histograms of local appearance features: feature vector generation and region detection for deformable objects. Mach Vis Appl 19:105–123. https://doi.org/10.1007/s00138-007-0086-y
33. Lecun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444. https://doi.org/10.1038/nature14539
34. Lokanath M, Kumar KS, Keerthi ES (2017) Accurate object classification and detection by faster-RCNN. In: IOP conference series: materials science and engineering
35. MacLeod N (2007) Automated taxon identification in systematics: theory, approaches and applications. Crc Press
36. Mao W, Wang FY (2012) New advances in intelligence and security informatics. Zhejiang University Press, Oxford
37. Marques ACR, Raimundo MM, Cavalheiro EMB, Salles LFP, Lyra C, Von Zuben FJ (2018) Ant genera identification using an ensemble of convolutional neural networks. PLoS One 13:e0192011. https://doi.org/10.1371/journal.pone.0192011
38. Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B (2011) How many species are there on earth and in the ocean? PLoS Biol 9:e1001127. https://doi.org/10.1371/journal.pbio.1001127
39. Naviaux R (1991) Les Cicindèles de Thaïlande, étude faunistique (Coleoptera Cicindelidae). Publications de la Société Linnéenne de Lyon 60:209–287. https://doi.org/10.3406/linly.1991.10944
40. Olivas ES, Guerrero JDM, Martinez-Sober M, Magdalena-Benedito J, Lopez AJS (2009) Handbook of research on machine learning applications and trends. IGI Global
41. Pang HW, Yang P, Chen X, Wang Y, Liu CL (2019) Insect recognition under natural scenes using R-FCN with anchor boxes estimation. In: lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics). Pp 689–701
42. Pass G, Zabih R (1999) Comparing images using joint histograms. Multimedia Systems 7:234–240. https://doi.org/10.1007/s005300050125
43. Pearson DL (1988) Biology of tiger beetles. Annu Rev Entomol 33:123–147. https://doi.org/10.1146/annurev.ento.33.1.123
44. Perez L, Wang J (2017) The effectiveness of data augmentation in image classification using deep learning. Convolutional Neural Networks Vis Recognit 11
45. Rahmat T, Ismail A, Aliman S (2019) Chest X-ray image classification using faster R-Cnn. Malaysian Journal of Computing 4:225–236
46. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM (2016) Deep convolutional neural networks for computer-aided detection: CNN architectures, Dataset Characteristics and Transfer Learning IEEE Transactions on Medical Imaging https://doi.org/10.1109/TMI.2016.2528162
47. Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. Journal of Big Data 6:60. https://doi.org/10.1186/s40537-019-0197-0
48. Shu M (2019) Deep learning for image classification on very small datasets using transfer learning. Iowa State University, Ames, Iowa
49. Sun J, Cai X, Sun F, Zhang J (2016) Scene image classification method based on Alex-net model. In: 2016 3rd international conference on informative and cybernetics for computational social systems. ICCSS 2016:363–367
50. Suthaharan S (2016) Machine learning models and algorithms for big data classification. Springer
51. Takefuji Y, Shoji K Effectiveness of ensemble machine learning over the conventional multivariable linear regression models
52. Thotagamuwa A (2018) Using insects as indicators of environmental health: Applications with tiger beetles (Coleoptera, Cicindelidae) of Sri Lanka. Ph.D. Thesis, University of Colombo,Sri Lanka
53. Vapnik VN (1999) An overview of statistical learning theory. IEEE Trans Neural Netw 10:988–999. https://doi.org/10.1109/72.788640
54. Veronese E, Castellani U, Peruzzo D, Bellani M, Brambilla P (2013) Machine learning approaches: from theory to application in schizophrenia. Computational and Mathematical Methods in Medicine

55. Wang M, Wang L, Ye Z, Yang J (2019) Ant lion optimizer for texture classification: a moving convolutional mask. IEEE Access 7:61697–61705. https://doi.org/10.1109/ACCESS.2019.2915553
56. Wu X, Zhan C, Lai YK, Cheng MM, Yang J (2019) IP102: a large-scale benchmark dataset for insect pest recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 8787–8796
57. Yuan Z-W, Zhang J (2016) Feature extraction and image retrieval based on AlexNet. In: Eighth International Conference on Digital Image Processing (ICDIP 2016). Chengdu, China
58. Zhu LQ, Ma MY, Zhang Z, Zhang PY, Wu W, Wang DD, Zhang DX, Wang X, Wang HY (2017) Hybrid deep learning for automated lepidopteran insect image classification. Orient Insects 51:79–91. https://doi.org/10.1080/00305316.2016.1252805