www.biodicon.com Biological Diversity and Conservation

ISSN 1308-8084 Online ISSN 1308-5301 Print

Research article
DOI: 10.46309/biodicon.2025.1515641
18/1 (2025) 1-12

An expert system for honeybee species identification and information retrieval

Swati SHILASKAR *[1], Shripad BHATLAWANDE[1], Shafaque SHEIKH[1], Sahil SALVE[1], Tayyab SHAIKH[1]
ORCID: 0000-0002-1450-2939; 0000-0001-8405-9824; 0009-0004-0082-5851; 0009-0000-5085-2234; 0009-0008-3927-2261

[1] Department of Electronics and Telecommunication Engineering, Vishwakarma Institute of Technology Pune, 666, Upper Indiranagar, 411037 Pune, Bibwewadi, India

Abstract

Detecting honeybee species is important for ecological and agricultural research, as it helps researchers understand their behavior, population, movement pattern, and pollination habits. This paper proposed a honeybee Identification system categorizing five honeybees namely: *Apis cerena indica, Apis mellifera, Apis florea, Apis dorsata,* and *Trigona.* Input images of honeybees were preprocessed to improve quality and eliminate any noise. Data augmentation methods were used to increase the dataset size, ensuring effective model training. The architecture of VGG16 which is popular for the image classification tasks identified the morphological characteristics present in the data set that were more relevant for species categorization rather than behavior analysis. Additionally, Rectified Linear Unit (ReLU) and Softmax layers were added, increasing the model's efficiency. The support Vector Machine model was trained to classify 5 classes of honeybees. After training, the model made accurate predictions of different honeybee species with high levels of precision and recall. This system performed exceptionally well in species classification, providing advancements in ecological and agricultural studies, through the implementation of VGG16 and SVM.

Keywords: ecosystem conservation, Honeybees, Computer vision, Deep learning, VGG16, SVM

---------- * ----------

Bal Arısı türlerinin tespiti ve bilgi erişimi için uzman sistem

Özet

Bal arısı türlerinin tespit edilmesi, araştırmacıların davranışlarını, popülasyonlarını, hareket şekillerini ve tozlaşma alışkanlıklarını anlamalarına yardımcı olduğundan ekolojik ve tarımsal araştırmalar için önemlidir. Bu makale, beş bal arısını kategorize eden bir bal arısı Tanımlama sistemi önermiştir: *Apis cerena indica, Apis mellifera, Apis florea, Apis dorsata* ve *Trigona*. Bal arılarının girdi görüntüleri, kaliteyi artırmak ve gürültüyü ortadan kaldırmak için önceden işlendi. Veri kümesi boyutunu artırmak ve etkili model eğitimi sağlamak için veri artırma yöntemleri kullanıldı. Görüntü sınıflandırma görevleri için popüler olan VGG16 mimarisi, veri setinde bulunan ve davranış analizinden ziyade tür kategorizasyonuyla daha alakalı olan morfolojik özellikleri tanımladı. Ayrıca Rektifiye Linear Unit (ReLU) ve Softmax katmanları eklenerek modelin verimliliği artırıldı. Destek Vektör Makinesi modeli, 5 sınıf bal arısını sınıflandırmak için eğitildi. Eğitimin ardından model, farklı bal arısı türlerine ilişkin yüksek düzeyde hassasiyet ve hatırlamayla doğru tahminler yaptı. Bu sistem, VGG16 ve SVM'nin uygulanması yoluyla ekolojik ve tarımsal çalışmalarda ilerlemeler sağlayarak tür sınıflandırmasında olağanüstü iyi performans gösterdi.

Anahtar kelimeler: ekosistemin korunması, Bal arıları, Bilgisayarla görme, Derin öğrenme, VGG16, SVM

* *Corresponding author*: Tel.: +912024202180; Fax.: +912024202290; E-mail: swati.shilaskar@vit.edu
 Received: 13.07.2024; Published: 15.04.2025 BioDiCon. 1159-130724

1. Introduction

In recent years, there has been a growing interest in leveraging computer vision techniques for honeybee identification, driven by the need to monitor bee populations efficiently and effectively. The advantages of pollination for people and the environment are in danger as bee numbers worldwide decrease due to which there is a pressing need for advanced methods to monitor bees species efficiently.

Honeybees are a diverse group of insects belonging to the class genus *Apis*. Their role is crucial for honey production and pollination. *Apis mellifera* is the most well-known species, usually called Western honeybee, spread all over and domesticated for its keeping of honey and agricultural pollination. Others are the Asian honeybee (*Apis cerena*), giant honeybee (*Apis dorsata*), and dwarf honeybee (*Apis florea*). Each has unique behaviors as well as adaptations for instance; *Apis dorsata* constructs large, exposed nests on tree branches while *Apis cerena* has a high resistance to certain diseases and parasites. Although bee pollination's role in advancing sustainable development goals by fostering biodiversity and food security is well known, there are still many undiscovered advantages that bees offer. It's crucial to remember that although image analysis can identify morphological characteristics that are particular to a species, it cannot immediately disclose behaviors that are specific to that species, such as *Apis dorsata's* aggression or *Apis florea's* restricted flight range. While identifying species is a first step in comprehending ecological responsibilities, this research introduces a computer vision-based method that does not directly evaluate behavioral patterns. Bees have the potential to support at least 30 SDG targets and 15 of the 17 Sustainable Development Goals (SDGs) [1]. Ecological research, apiary management, and conservation initiatives all heavily rely on information and species identification systems for honeybees. Knowledge of the distribution, behavior, and ecological roles of honeybee species within ecosystems depends on accurate species identification. Manual procedures, which are labor-intensive, time-consuming, and need specialized taxonomic skills, were the norm for species identification in the past. Recently, though, technological advancements have created new opportunities for automating this process, especially in the areas of deep learning, computer vision, and remote sensing. Through the use of image processing techniques, these systems can distinguish between distinct species and subspecies of honeybees by analyzing complex morphological traits such as wing venation patterns. This technique provides a productive and non-invasive way to identify species, eliminating the requirement for labor-intensive human labor and a high level of taxonomic knowledge. DeepABIS deep learning method is used for identifying honeybee subspecies using wing photos of eight different species of honeybees e.g. Carnica, Siciliana, etc. The models that are used to train these datasets are CNN models such as MobileNetV2, B-CNN, and Inception-ResNet. By removing intricate features, DeepABIS gained an accuracy of 93.95% for (top-1) and 99.61% for (top-5). A smartphone application and website named CloudABIS is used to input the image and for the graphical user interface [2]. Xilotheques in the 21st century have a unique opportunity to advance with AI and deep learning for enhanced forest species identification. Challenges include limited image datasets, lack of automated identification focus, and uneven taxonomic distribution. Solutions involve digital sample photography, standardized data sharing, and prioritizing sampling methods for comprehensive image collections, positioning xilotheques as crucial hubs for forest species research [3].

To monitor honeybee colony conditions such as brood rearing and swarming, the study presented a system that uses temperature data from one sensor per hive. It successfully recognizes these states by using artificial neural networks for pattern identification. Limitations include possible algorithm complexity and reliance just on temperature data, despite the promising nature of the approach. Still, the strategy looks promising for better colony management of honeybees [4]. An imaging device to track the activity of honeybees at hive entrances is shown in this work. It tracks individual bees using infrared technology and character-encoding tags, obtaining great accuracy in bee identification (86%) and character recognition (98%) respectively. The method shows its potential for effectively investigating honeybee behavior by offering insights into everyday foraging behavior [5]. The Hungarian algorithm and Kalman Filter are used in this study's real-time imaging system to track honeybee activity at hive entrances, providing precise bee detection and background reduction. A 4G LTE router is used to send data to a distant server for analysis, allowing for long-term observation of colony health and behavior, including the impacts of pesticides. This economical system provides important information on colony's health and honeybees' activity [6]. The findings of research describes an audio recording technique near hive entrances for early swarming mood detection in bee colonies. Mel Frequency Cepstral Coefficients (MFCCs) and their derivatives are analyzed to identify significant coefficients for worker bee and drone bee categorization. The study surpasses the typical MFCC coefficients' accuracy of up to 90% by improving detection accuracy to slightly above 95% with the chosen set of signal attributes through the use of an autoencoder neural network [7]. This study used computer vision and deep learning to automate the identification of bumble bee species. Using 89,000 photos from 36 different species, it assesses four models; InceptionV3 achieves 91.6% accuracy in 3.34 milliseconds. Bee Machine, a user-friendly identification tool that, and highlights the technology's revolutionary potential is presented [8]. The methodology used geometric morphometrics to analyze forewing vein patterns to produce a quick and simple approach for identifying populations, ecotypes, subspecies, or hybrids of honeybees. Reference samples comprising 187 honeybee colonies from 25 subspecies spanning four evolutionary lineages were taken from the Morphometric Bee Data Bank. While evolutionary lineage identification turned shown to be very consistent, subspecies identification accuracy varied, with African honeybees demonstrating less consistency. For analysis, the acquired data were exported to the Identify programmer. All things considered; the study helps create effective instruments for identifying different subspecies of honeybees [9]. The study introduced a hierarchical multi-task structural learning algorithm designed to enhance plant species identification on a large scale. A visual tree is utilized to organize numerous plant species in a coarse-to-fine manner, facilitating the identification of interrelated learning tasks and guiding hierarchical classifier training. By effectively controlling error propagation between levels, the developed tree classifiers demonstrate superior discrimination power for large-scale plant species identification. Experimental results showcase competitive performance in terms of identification accuracy and computational efficiency compared to existing approaches. Additionally, the proposed algorithms have potential applications in large-scale image classification tasks [10]. A developed Neural Network was applied to detect two honeybee colony states. It detected 11 possible swooning cases [11]. The research presents an alternative solution for the automated identification of tree species using macroscopic woodcut images. They propose a convolutional network trained from scratch achieving 93.6% top-1 accuracy and further improve results with a pre-trained ResNet50 model reaching 98.03%. Despite successful application, the immediate challenge lies in building larger datasets, with ongoing efforts focused on a database of Costa Rican tree species. The importance of xylotheques in various fields underscores the significance of this research [12]. Findings of the research demonstrates the effectiveness of image processing in identifying fish species with a high degree of accuracy. However, limitations exist in the scope of fish species covered and the static nature of the image acquisition process. Future research should focus on expanding the range of freshwater fish species studied and developing dynamic image processing methods to enable real-time and online identification, thereby improving practical applicability in production settings [13]. To classify datasets of different bee species, this paper uses the DeepABIS deep learning system that uses MobileNetV2, B-CNN, and Inception-ResNet. By using these models, they got accuracies of 85%, 87%, and 92%, respectively. They have also made a website using Flask and PHP for the backend for inputting the image. The novelty of this paper is anomaly detection for species identification tasks where the number of different existing species is often unknown so that's why detecting outliers is important [14]. The main part of paper that makes it novel is the region of interest locator network can be trained with a minimum number of hand-annotated images; it can handle crowdsourced data with a wide range of image quality; it doesn't require manual image cropping before genus/species identification; and it can accurately classify multiple bees in a single image. A dataset from the Bee Spotter website is used for this paper, comprising 15,347 crowd-sourced images of bees annotated for species by an expert. The method used for both bee identification and classification tasks is Faster R-CNN with a Resnet 101+FPN backbone. The model's accuracy on average is 91% [15]. The dataset which contains 5000 images of different bee subspecies like Italian bee, Russian bee, Western Honey bee, etc is used in this paper with health conditions using a two-layer CNN. The identification of the bees is done in this paper. The CNN model has 86.5% accuracy for subspecies and 84.9% accuracy for health classification. Data Preprocessing is done by SMOTE Data Balancing. To enhance generalization, data augmentation techniques are applied to the images. Relu and SoftMax are used as activation functions [16]. The work proposes a method for the recognition of pollen-bearing honeybees in hive entrance videos to monitor foraging behavior. Tests were conducted on a newly annotated dataset with several models, including baseline classifiers and Convolutional Neural Networks. Although the simple CNN performed the best, its predictions were somewhat degraded due to misalignments and the visual noise of manual annotations. Networks shallower in depth did not lead to better performance due to limited data and computational demands. However, in the future, large, high-quality datasets should be developed by automatic image collection and validation to enhance model performance and applicability in real practice [17]. Deep learning-based models for monitoring honeybee hives to detect healthy bees, pollen-bearing bees, and anomalies such as Varroa parasites, ants, hive robberies, and small hive beetles are reflected in the research. Transfer learning with pre-trained deep neural networks (DNNs) and support vector machines (SVMs) under a variety of feature sets are tested on three datasets consisting of 19,393 images. The models reach as high as 99.07% accuracy, proving to be useful in smart beekeeping and real-time hive monitoring. Although no DNN was best across all datasets, the accuracy of these models and the short classification time make them promising tools for improving the safety of bees and enhancing the efficiency of beekeeping. Future work will adapt these models to other hive conditions and evaluate the performance of DNNs in further detail [18]. A bee hive health monitoring system using image processing with Convolutional Neural Networks has been proposed in the work. It includes two models: one for detecting bees and another for health classification. In health classification, it exceeded existing methods with an accuracy of 95%, while the detection of bees was lower, at 82%. While health classification is quite fast, at more than 500 FPS, bee detection is slower at 2-3 FPS, which makes the system unsuitable for real-time usage. Future work will then focus on improving speed and accuracy with the potential use of Faster R-CNN. This system promises to reduce stress in monitoring bees and support global food production through better bee health [19]. Findings describes a process of classifying images of pollen-bearing bees using Convolutional Neural Networks, with the goal of future implementation on low-cost FPGA hardware. A new dataset of images of bees was taken at hive entrances. This paper has investigated multiple configurations of CNNs, up to 3 hidden layers with up to 15 filters per layer, and with filter sizes ranging from 15x15 to 3x3. The most appropriate configuration found is a three-hidden-layer CNN with a 7-7, 5-5, 3-3 configuration, which reaches 94% of accuracy in classification, balancing the computational requirements. This trained CNN will be implemented on FPGA for the real-time video detection of pollen-bearing bees. To identify the regions of interest, a histogram of oriented gradients and background subtraction techniques will be applied [20]. The color of the bee depends on the form of the bee, whether the bee is a drone, worker bee, or queen bee. In these 3 forms of bees, the color and pattern variation is observed. The color variation ranges from yellow to black. Yellow workers are the result of the cross of the yellow queen with a black drone. Black drones are produced by yellow-laying workers. The body color is an expression of the gender of the bees [21]. The genetics based on the origin of the brood influences the coloration of abdominal tergites. The pattern on the abdomen varies based on temperature and the surrounding ecosystem [22]. A morphometric study of honeybee species *Apis cerena* is carried out in [23]. This study relies heavily on color patterns and detailed visual features of honeybees. AI-based technique using ResNet for classification of bumble bees from non-bumble bees is mentioned in [24]. 91.6 % accuracy of classification was achieved in this work. Geometric Morphometric Analysis was used to classify honeybees based on shape morphology. Deviations in the intersections of the wing vein angles was used as the distinguishing parameter [25].

Honeybees are the most important pollinators. They are a vital part of the forest ecosystem and agriculture. Identification and classification of honeybees are necessary to maintain a record of the honeybees and for track the movement of their colony. Honeybees can be identified based on morphology, foraging behavior, and defensive mechanisms. However, in the forest surrounding it is difficult and time-consuming to observe in detail the color shades and visual patterns on the body of a honeybee. Only an expert and well-trained individual will be able to identify such cases. This work uses computer vision, machine learning, and deep learning to identify the honeybee based on image input. Based on the identification results, the system fetches details about the honeybee species from the dataset. The system generates the details which include scientific name, distribution, and habitat. In this work, a more inclusive environment is developed with an expert system for honeybee species identification from the captured image. The computer vision-based model is trained by well-identified input images. The honeybee identification must be carried out with their habitat in the background. Hence honeybee images with a background of natural surroundings like flowers and leaves are used for testing the model in the proposed work. Table 1. Displays the work that has already been done for classification utilizing various algorithms

Table 1. Comparison of prior art

Reference No.	Database/Dataset used in literature	Algorithm
[13] (2020)	Datasets for Fish Species Datase 1 : 9942 images, 166 genera, and 881 species. Dataset 2: 300 images with exactly 60 images per population (30 males, and 30 females).	MobileNetV2, Tensorflow, Keras, B-CNN, Inception-ResNet
[1] (2022)	Total of 9887 images which includes 8 classes for Honeybees.	ResNet 50, MobileNet V2, Inception Net V3, and Inception ResNet V2
[2] (2018)	9000 photographs of some 1800 species of hard commercial timber in the world	MnasNet, ResNet, Wide ResNet

The research outcomes of prior work are consisely presented in Table 1 that shows a variety of methods for classifying species using deep learning. As an instance, the Study [13] used MobileNetV2, B-CNN, and Inception-ResNet to classify a large, heterogeneous dataset of fish species, emphasizing image analysis for an aquatic species group that is distinct to a population and cross-gender. Although Study [1] examined honeybee classification it used larger classes of honeybees with numerous model architectures, such as ResNet50 and InceptionNet V3, with the goal of recognizing general species rather than differentiating between subspecies. Similarly, to identify timber species, Study [2] used MnasNet, ResNet, and Wide ResNet, which necessitated algorithms tailored to plant morphology. In contrast, this research optimizes model accuracy to capture subtle morphological distinctions essential for ecological and agricultural research by combining VGG16 with SVM in a unique way to specifically identify five honeybee classes. This customized strategy sets this methodology apart from previous studies' more generalized classification objectives and promotes it as a focused option for honeybee biodiversity monitoring.

2. Materials and methods

The paper presents a thorough approach to computer vision-based honeybee identification that includes image capture, preprocessing, feature extraction, and classification steps. To improve feature extraction, image data is preprocessed using techniques like resizing, normalization, and Gaussian blur. However, as the input photos and extracted characteristics only offer morphological information, excluding behavioral analysis.

The dataset includes 5 types of honeybees and honeybees are classified into 5 classes as given below:
1. *Apis cerena indica* (Indian Honey bee)
2. *Apis dorsata*
3. *Apis florea*
4. *Apis mellifera*
5. *Trigona*

2.1 Details of the Information System

Honeybees are colonial insects having a complex social hierarchy with one queen, many worker bees, and drones. The fact that they can repurpose information about food sources through waggle dance is amazing for them. Their pollinating activities make them crucial parts of ecosystems due to their contribution towards the reproduction of many flowering plants hence maintaining biodiversity including agricultural productivity. A description of the class of honeybees and the difference between them in terms of features and location is shown in Table 2. Honey bees are climate indicators. The information system is designed to fetch information about the identified honeybee. This expert system records their presence, location, day, date, and Time of occurance. A scientific study of this pattern of occurance helps in monitoring and assessing the health of the environment, and predict future changes. This will aid in taking measures to take positive measures towards biodiversity conservation.

Table. 2. Desciption of honey bees classified in this work

Sr	**Names**	**Category**	**Features / characteristics**	**Location**
1	*Apis cerena dica*	Subspecies	Slightly Larger, relatively non-gressive, and rarely exhibiting arming behavior. Has 4 stripes on the domen	India, Pakistan, Nepal, Myanmar, ngladesh. Srilanka, Thailand, inland Asia
2	*Apis mellifera*	Species	Western Honey bee, Lifespan 30-60 ys. Has 3 abdominal stripes	Africa, Europe and the Middle st.
3	*Apis florea*	Species	Dwarf honeybees, small in size	South-east Asia
4	*Apis dorsata*	Species	very large (17-20 mm), golden in lor. Abdominal stipes yellow to black.	South-east Asia
5	*Trigona*	Genus	Stingless bees, Black spiny feet	Mexico, Columbia, South nerica, and Venezuela

2.1.1 Honeybee identification architecture

The honeybee category detection project using computer vision starts by gathering honeybee images and then goes through preprocessing tasks like resizing and normalization. To increase the variety in the dataset, techniques like rotation and flipping are applied for data augmentation. The VGG16 architecture, which is a pre-trained convolutional neural network (CNN), is used to extract features from the images. Non-linearity for classification is introduced with Relu activation functions, and a SoftMax layer generates probability distributions for different species. Ultimately, the model that has been trained predicts the honeybee species by analyzing input images using the features and probabilities it has learned. Figure1. shows the block architecture of the honeybees species identification system.

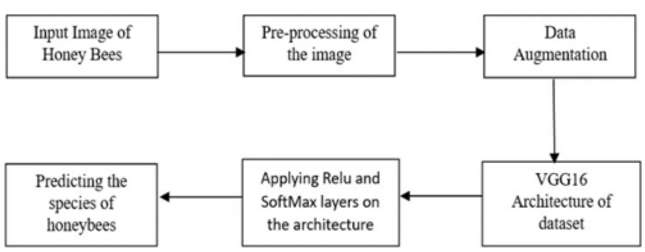

Figure 1. Honeybee Identification system

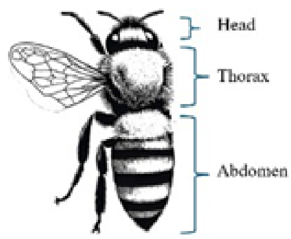

Figure 2. Major external organs of a honeybee

The anatomy of a honeybee has 3 major parts namely the head, thorax, and abdomen as shown in Figure 2. Abdomen is the most visible part of the honeybee image. The color, size, and shape of the stripes on the abdomen and the size and shape of the abdomen are important for species identification. The image processing and feature extraction techniques are used to capture the underlying distinguishing pattern for a particular species.

2.1.2 Dataset/Preprocessing details

The dataset is collected from the iNaturalist website. iNaturalist is a well-known website and community-based database where people can share their sightings of different species of plants and animals, including honeybees. The dataset contains images of 5 types of honeybees namely *Apis cerena indica, Apis mellifera, Apis florea, Apis dorsata*, and *Trigona*. Each species contains approximately 200 images, therefore, overall, 1070 images are used in the dataset as shown in Figure 3.

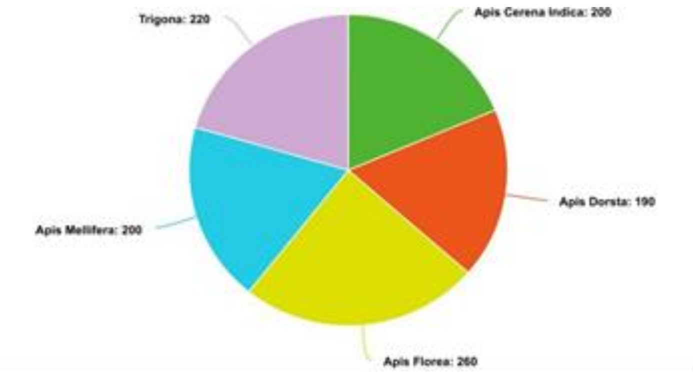

Figure 3. Dataset taken for each subspecies

Preprocessing: During dataset preprocessing, the first step is to label and resize images to ensure consistency in representation. Gaussian blur is applied to remove noise and to enhance feature extraction for SVM after resizing, while VGG16 only requires resizing. Normalization adjusts pixel values and data augmentation helps diversify the dataset, improving model robustness. These steps improve dataset preparation and ultimately boost model performance in machine learning or deep learning tasks.

Figure 4. Honeybees images with and without natural surrounding

Figure 4 depicts the images of honeybees belonging to catoried mentioned in this paper. The images include honeybees perched on flowers or leaves as it is their natural surroundings. These images are then used for dataset preprocessing. Figure 6 depicts the preprocessed images of *Apis mellifera* where Gaussian blurring has been performed.

Figure 5. Labeled images for 5 classes

Figure 6. Gaussian Blur filtered images of *Apis mellifera type*

2.1.3. Classification of honeybees

This work implements two techniques for classification. The machine learning technique uses SIFT feature classification using a Support Vector Machine (SVM). The deep learning technique uses the VGG-16 model for class identification.

SIFT features and SVM-based classification: SIFT is a feature extraction algorithm used to detect the key points and descriptors of the given image. It is renowned for its robustness to variations in illumination, scale, and rotation which makes it suitable for extracting key points and descriptors from grayscale images. Edges, corners, and texture patterns, these types of local features are extracted from the image. SIFT first identifies key point locations in the image based on gradient information and then computes descriptors around these key points to capture their local image information. Equation 1 describes the process of extracting key points from an image using the SIFT method.

$$Z = -\frac{\partial^2 D^{-1}}{\partial x^2}\frac{\partial D}{\partial x} \qquad (1)$$

(a) (b) (c)

Figure 7. SIFT feature detection of morphological parameters of honeybee image

It can be seen from Figure 7 a and b that the distinct organs of the honeybee are mapped and key points are detected. The wings, thorax, and abdomen key points are identified based on pixel intensities and neighborhood. Figure 7 c shows three types of stipes patterns on the abdomen of honeybee. The key points identified in this image carry important information about the underlying pattern. SIFT features are invariant to rotation, scale, blur, Illumination, warping, and noise. Hence the system trained on SIFT features is robust and can identify the species even under the following adverse conditions. It is observed that honeybees curl their abdomens while in flight and under restraint. Their body part could get occluded by flowers or leaves. Light intensity could be low due to shadow or low ambient conditions. Honebees beat their wings 230 times per second while flying. Rhythmic thorax pulsation and wing movement make it difficult to capture sharp images of honeybees. The image of a moving honeybee could get blurred especially for low shutter speed cameras. Figure 8 shows features for *Trigona* genus. and Figure 9 exhibit the key points on wings of *Apis mellifera* class honeybee detected after applying the SIFT feature extraction method.

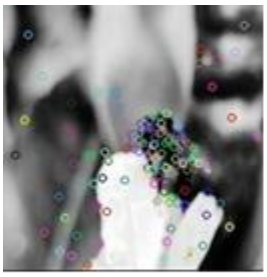

Figure 8. Keypoints of *Trigona* class detected by SIFT

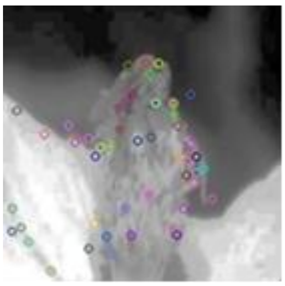

Figure 9. Keypoint of *Apis mellifera* class detected by SIFT

Images are divided into training and testing datasets. From each class, 20% of the images were reserved for testing and 80% of the images were used for training. The SIFT features are extracted from the image dataset. Extracted features are then reduced by using the k-means algorithm. The reduced feature set represents a one-row vector for each image. The feature set is generated for all classes and provided to an SVM model for classification training. The SVM model is trained to predict the class of honeybees by analyzing the keypoint descriptors extracted from the images. This combination of SIFT for feature extraction and SVM for classification increases the robustness of SIFT in handling variations and the effectiveness of SVM in classification between different species based on the extracted features.

VGG 16 Model for classification: The paper proposes a deep learning-based classification system where the VGG16 model has been implemented. Figure 10. represents the system architecture of VGG16. The VGG16 is a convolutional neural network architecture developed by the Visual Geometry Group at the University of Oxford. It is known for its deep structure, which includes 16 layers, hence the name "VGG16." The classifier model used to classify honey bee images is based on transfer learning. It allows to use a pre trained deep learning model trained on huge dataset. The model is pre trained on a diverse and rich feature representation for wide range of images. This feature representation is transferred to classify a small number of training images involving 5 classes of honeybees. This improves classification performance and training time as compared to training the model from scratch [18].

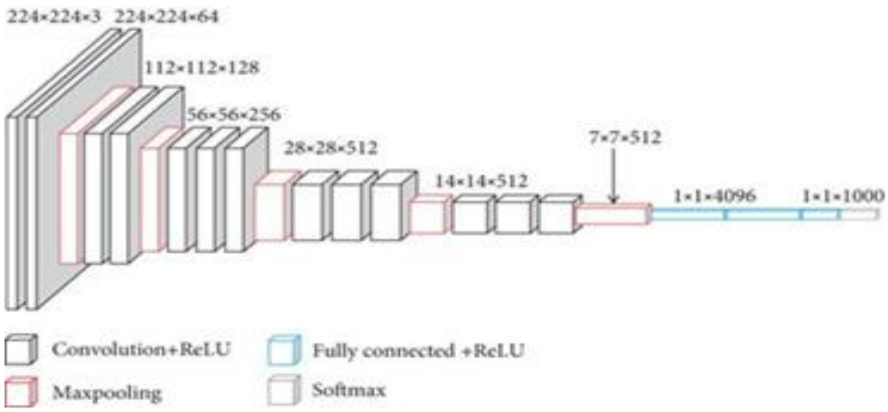

Figure 10. System Architecture of VGG16 [18] weights='imagenet': This parameter specifies that the pre-trained weights of the VGG16 model trained on the

ImageNet dataset should be used. ImageNet is a large dataset with millions of labeled images across thousands of categories, and pre-training on it can give the model good generalization capabilities. input_shape = (256, 256, 3): This argument defines the expected shape of the input images that will be fed into the model. It specifies 256 pixels in height, 256 pixels in width, and 3 color channels (RGB). include_top=False: This parameter indicates that the fully connected layers (also known as the "top" layers) of the VGG16 model should not be included. The fully connected layers are typically responsible for the final classification, and by excluding them, you can use the convolutional base of the VGG16 for feature extraction in a different task.

Setting conv_base.trainable = False has the following effects:

1. Freezing Weights: It prevents the weights of the convolutional layers within conv_base from being updated during the training process. This means that the learned feature representations from ImageNet will remain unchanged.
2. Faster Training: By freezing the majority of the model's weights, you significantly reduce the number of parameters that need to be adjusted during training. This leads to faster training times and less computational resources required.
3. Transfer Learning: This technique is often used in transfer learning, where you leverage the knowledge of a pre-trained model on a different task. By freezing the base layers, you preserve the generic feature extraction capabilities and focus on fine-tuning the new layers you add on top for your specific task. Transfer learning in VGG16 is shown in Figure 11 and Figure 12. represents the working of VGG16.

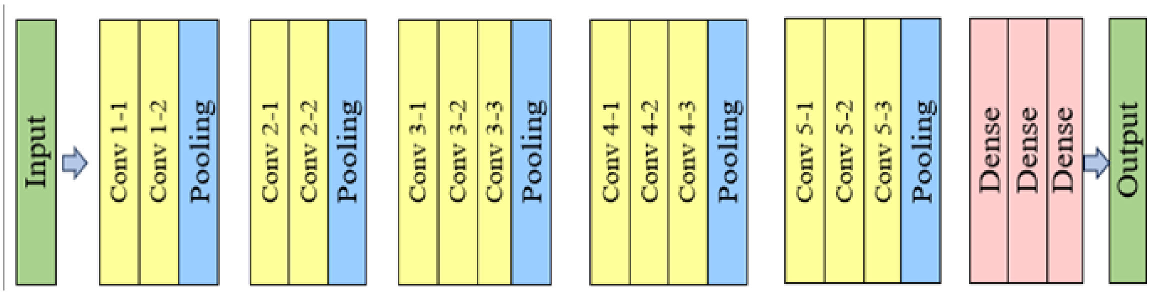

Figure 11. Transfer learning in VGG16

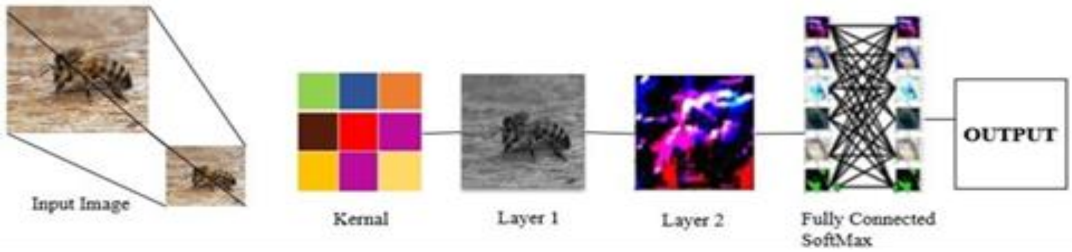

Figure 12. Working of VGG16

Algorithm: Identification Using VGG16
Input: Honey Bee image
Output: Identified *Apis cerena indica, Apis mellifera, Apis florea, Apis dorsata, and Trigona*

1. Split the dataset into train, test, and valid
2. for image in the folder
3. Normalize and resize the image
4. Apply 2D CNN with filter size 3
5. Applying f*f filters from 256 to 32
6. Activation Function using R(Z) = max(0,Z)
7. Neuron Dropout by a factor of 0.2
8. Flat 2D output matrix to 1D
9. end for

Algorithm 1 displays the Activation Function (relu) along with neuron dropout by a factor of 0.1 to avoid overfitting. Max-Pooling followed by converting the feature matrix into a dense layer is applied along with the softmax function and uses Adam optimizer while compiling the model.

2.1.4 *Implementation of a website to integrate the trained model*

Flask API was used to integrate the trained model into the self-designed website. HTML, CSS, and JavaScript are used to design the website's front end. The goal of creating a user-friendly website is to integrate the model and to provide a better experience to the user. The website is designed to record the predicted class, sighting date, Time alongwith the location. The website gives details of natural habitat and typical characteristics of the predicted class of honeybee.

3. Results

The paper focuses on identifying honeybees, encompassing five classes namely: *Apis cerena indica, Apis mellifera, Apis florea, Apis dorsata,* and *Trigona*. A dataset comprising 1000 images underwent preprocessing procedures including Gaussian blur and histogram equalization to enhance image quality. Histogram equalization enhances image contrast, while Gaussian blur reduces noise and simplifies details. The use of VGG16, a Convolutional Neural Network (CNN) architecture, helps in identifying and classifying species in this study. This paper introduces an innovative approach for honeybee species identification employing VGG16 and SVM algorithms. SVM shows an accuracy of 90%. Visualizations showing training and validation accuracy, as well as loss metrics, are included. The model accurately predicts different bee classes, giving a clear visual representation of the classification performance shown in Figure 13 and Figure 14. The classification model achieves high accuracy in species identification, resulting in reliable differentiation based on physical traits visible in images. However, these findings are not behavior-related. For instance, *Apis dorsata* is known for its aggressive nature and extended flight range, qualities that images cannot convey. This methodology is an effective tool for honeybee class identification, indirectly supporting ecological studies but not replacing direct behavior analysis methods.

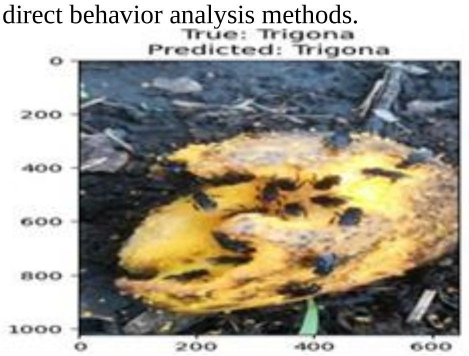

Figure13. Results predicted by the model for *Trigona* Species

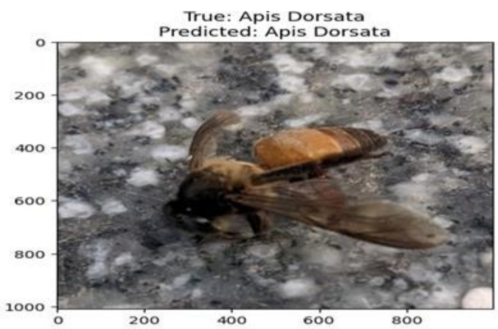

Figure14. Results predicted by the model for *Apis Dorsata* Species

The Train accuracy and validation accuracy graph for training and validation is shown in Figure 15 and Figure 16. represents Training Loss and Validation Loss for VGG16. It is seen that as the epochs increase, the model gets trained the accuracy improves and losses reduce. The x-axis shows a number of epochs.

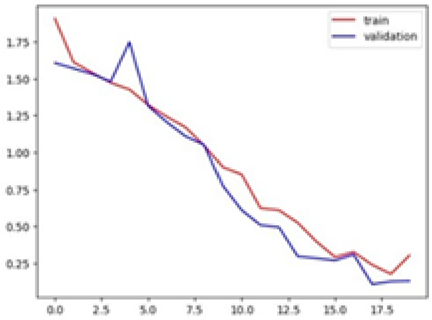

Figure15. Train accuracy and validation accuracy

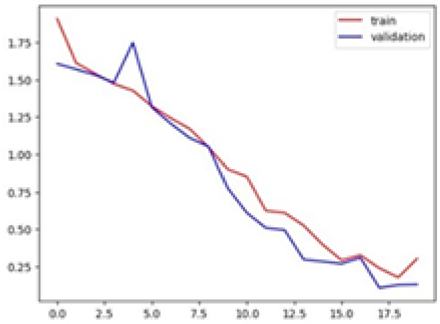

Figure 16. Train loss and validation loss

The dataset used in this research includes pictures of flowers and leaves with honeybees on them, which makes extracting features difficult and lowers accuracy. Therefore, it is important to include clearer images in the training dataset to overcome this limitation and improve the feature extraction methods' effectiveness. Table.3. represents performance for the SVM and VGG16 Models. According to the results obtained, it is seen that the model VGG16 has a better performance than SVM.

Table .3. SVM and VGG16 classification performance (in %)

Classification Performance	**SVM Classifier**	**VGG16 Classifier**
Accuracy	90.00	99.6
Precision	95.00	98.00
Recall	83.33	96.21
F1 Score	87.40	97.17
AUC Score	90.00	95.00

Figure17. depicts the information system for honeybee species in a self-designed website. The model is trained and then integrated into a website as shown. The input image of a honeybee is identified as the class *Apis cerena indica* successfully.

Figure17. Model integrated into the self-designed website

Figure18. Information system for honeybees subspecies

Figure18 represents the Information System developed in this work details the honeybee habitat and the characteristics of the identified honeybee. This system is aimed to aid the conservation of forests. The natural inclination of honeybees towards specific types of flower pollens drives the colonies to settle and grow at the location. The expert system identifies the honeybee and provides the information that the forest conservators can use for constructive measures. It acts as an information system even for a learner involved in keeping records of the forest health and ecosystem.

4. Conclusions and discussion

This research endeavors to contribute to the field of honeybee species identification by proposing a robust classification system that can distinguish between five different subspecies: *Apis cerena indica*, *Apis dorsata*, *Apis mellifera*, *Apis florea*, and *Trigona*. Through the integration of VGG16 and SVM algorithms, our approach demonstrates promising accuracy in identifying these subspecies. Furthermore, the seamless integration of the trained model into a user-friendly website interface enhances accessibility and usability for stakeholders interested in honeybee classification. Accurately identifying different species of honeybees is crucial for beekeepers, researchers, and policymakers. It helps with biodiversity conservation, monitoring ecosystems, and promoting sustainable agriculture. Therefore, this study not only introduces a practical technological answer but also emphasizes the importance of accurate class recognition which is vital for taking measures to protect the honeybee population and maintain ecological harmony.

This research could be advanced by broadening the dataset to include more species of honeybees and a wider range of geographical areas, investigating more complex deep learning architectures, incorporating real-time data collection techniques, adding more features like genetic markers or behavioral traits, utilizing transfer learning techniques, and expanding the application to larger ecological studies like habitat monitoring and biodiversity assessments. Further techniques such as direct observation, are required for in-depth behavioral insights. Future work can integrate sensors or movement that would help track complement image-based species identification for a fuller understanding of behavior.

Acknowledgements

Authors would like thank Vishwakarma Institute of Technology, Pune for the encouragement to carry out research towards environment protection awareness. The authors would also like to thank all who have helped directly or indirectly to carry out this research.

References

[1] Patel, V., Pauli, N., Biggs, E., Barbour, L., & Boruff, B. (2020). *Why bees are critical for achieving sustainable development*. Ambio, 50(1), 49–59. https://doi.org/10.1007/s13280-020-01333-9

[2] De Nart, D., Costa, C., Di Prisco, G., & Carpana, E. (2022). *Image recognition using convolutional neural networks for classification of honey bee subspecies*. Apidologie, 53(1). https://doi.org/10.1007/s13592-022-00918-5

[3] Figueroa-Mata, G., Mata-Montero, E., Valverde-Otarola, J. C., & Arias-Aguilar, D. (2018). *Automated Image-based Identification of Forest Species: Challenges and Opportunities for 21st Century Xylotheques*. https://doi.org/10.1109/iwobi.2018.8464206

[4] Kviesis, A., & Zacepins, A. (2016). *Application of neural networks for honey bee colony state identification*. https://doi.org/10.1109/carpathiancc.2016.7501133

[5] Chen, C., Yang, E. C., Jiang, J. A., & Lin, T. T. (2012). *An imaging system for monitoring the in-and-out activity of honey bees*. Computers and Electronics in Agriculture, 89, 100–109. https://doi.org/10.1016/j.compag.2012.08.006

[6] Ngo, T. N., Wu, K. C., Yang, E. C., & Lin, T. T. (2019). *A real-time imaging system for multiple honey bee tracking and activity monitoring*. Computers and Electronics in Agriculture, 163, 104841. https://doi.org/10.1016/j.compag.2019.05.050

[7] Libal, U., & Biernacki, P. (2024). *MFCC Selection by LASSO for Honey Bee Classification*. Applied Sciences, 14(2), 913. https://doi.org/10.3390/app14020913

[8] Spiesman, B. J., Gratton, C., Hatfield, R. G., Hsu, W. H., Jepsen, S., McCornack, B., Patel, K., & Wang, G. (2021). *Assessing the potential for deep learning and computer vision to identify bumble bee species from images*. Scientific Reports, 11(1). https://doi.org/10.1038/s41598-021-87210-1

[9] Nawrocka, A., Kandemir, R., Fuchs, S., & Tofilski, A. (2017). *Computer software for identification of honey bee subspecies and evolutionary lineages*. Apidologie. https://doi.org/10.1007/s13592-017-0538-y

[10] Fan, N. J., Zhou, N. N., Peng, N. J., & Gao, N. L. (2015). *Hierarchical Learning of Tree Classifiers for Large-Scale Plant Species Identification*. IEEE Transactions on Image Processing, 24(11), 4172–4184. https://doi.org/10.1109/tip.2015.2457337

[11] Karaboga, D., & Kaya, E. (2018). *Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey*. Artificial Intelligence Review, 52(4), 2263–2293. https://doi.org/10.1007/s10462-017-9610-2

[12] Figueroa-Mata, G., & Mata-Montero, E. (2020). *Using a Convolutional Siamese Network for Image-Based Plant Species Identification with Small Datasets*. Biomimetics, 5(1), 8. https://doi.org/10.3390/biomimetics5010008

[13] Li, L., & Hong, J. (2014, August). Identification of fish species based on image processing and statistical analysis research. In 2014 IEEE International Conference on Mechatronics and Automation (pp. 1155-1160). IEEE. DOI: 10.1109/ICMA.2014.6885861

[14] Buschbacher, K., Ahrens, D., Espeland, M., & Steinhage, V. (2020). *Image-based species identification of wild bees using convolutional neural networks.* Ecological Informatics, 55, 101017. https://doi.org/10.1016/j.ecoinf.2019.101017

[15] Kelley, W., Valova, I., Bell, D., Ameh, O., & Bader, J. (2021). *Honey sources: neural network approach to bee species classification.* Procedia Computer Science, 192, 650–657. https://doi.org/10.1016/j.procs.2021.08.067

[16] Karthiga M; Sountharrajan S; Nandhini S S; Suganya E; Sankaranath S. (2021*). A Deep Learning Approach to classify the Honeybee Species and health Identification*. https://doi.org/10.1109/icbsii51839.2021.9445173

[17] Rodriguez, I. F., Megret, R., Acuna, E., Agosto-Rivera, J. L., & Giray, T. (2018). *Recognition of Pollen-Bearing Bees from Video Using Convolutional Neural Network.* https://doi.org/10.1109/wacv.2018.00041

[18] Berkaya, S. K., Gunal, E. S., & Gunal, S. (2021). *Deep learning-based classification models for beehive monitoring.* Ecological Informatics, 64, 101353. https://doi.org/10.1016/j.ecoinf.2021.101353

[19] Braga, D., Madureira, A., Scotti, F., Piuri, V., & Abraham, A. (2021). *An Intelligent Monitoring System for Assessing Bee Hive Health*. IEEE Access, 9, 89009–89019. https://doi.org/10.1109/access.2021.3089538

[20] Sledevic, T. (2018). *The Application of Convolutional Neural Network for Pollen Bearing Bee Classification.* https://doi.org/10.1109/aieee.2018.8592464

[21] Woyke, J. (1998, March). Differences in body colour expression between European and Asian honeybees. In Proceeding of Fourth Asian Apicultural Association International Conference (pp. 23-28).

[22] Bubnič, J., & Prešern, J. (2024). *Quantifying Abdominal Coloration of Worker Honey Bees*. Insects, 15(4), 213. https://doi.org/10.3390/insects15040213

[23] Dar, S. A., & Ahmad, S. B. (2017). *Morphometric variations and expression of body colour pattern of honeybee, Apis cerana F.* in Kashmir. Journal of Entomology and Zoology Studies, 5(4), 364–371. https://www.entomoljournal.com/archives/2017/vol5issue4/PartE/5-3-129-616.pdf

[24] Bhuiyan, Tanvir, Ryan M. Carney, and Sriram Chellappan. "Artificial intelligence versus natural selection: Using computer vision techniques to classify bees and bee mimics." Iscience 25, no. 9 (2022). DOI: 10.1016/j.isci.2022.104924

[25] Kekeçoğlu, Meral, Merve Kambur, Münir Uçak, Tuğçe Çaprazlı, and Songül Bir. "Biodiversity of honey bees (Apis mellifera L.) in Turkey by geometric morphometric analysis." Biological Diversity and Conservation 13, no. 3 (2020): 282-289. nhttps://doi.org/10.46309/biodicon.2020.773984