IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 3, June 2026, pp. 2685~2697
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i3.pp2685-2697

A novel deep learning-based pollinator species classification using the multimodal species-image regularization framework

Pooja Hadimane[1], Ashoka Kukkuvada[2]
[1]Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology (Affiliated to Visvesvaraya Technological University, Belagavi), Davangere, India
[2]Department of Information Science and Engineering, Bapuji Institute of Engineering and Technology (Affiliated to Visvesvaraya Technological University, Belagavi), Davangere, India

Article Info

Article history:

Received Dec 11, 2024
Revised Mar 9, 2026
Accepted Apr 22, 2026

Keywords:

Climate change
Image-species prototype regularization
Multimodal contrastive learning framework
Multimodal species-image regularization
Pollination

ABSTRACT

Pollinators, such as bees, butterflies, and other insects, are essential to maintaining biodiversity and ensuring agricultural productivity, with over 80% of flowering plants and 75% of global food crops relying on them for successful reproduction. However, pollinator populations are facing significant declines due to environmental changes, habitat destruction, and climate change, posing substantial risks to ecosystems and global food security. This paper introduces the multimodal species-image regularization (MSIR) framework for automating the classification of pollinator species using both binary classification (pollinator vs. non-pollinator) and multiclass classification (bee genera). The framework integrates multimodal data, including visual images of pollinators and species details such as genus, family, and environmental factors, to improve accuracy and scalability. The system leverages the multimodal contrastive learning framework (MCLF) to align both image and species-detail features into a unified embedding space, enabling mor effective classification. Additionally, the framework applies image-species prototype regularization (ISPR) and species-detail prototype regularization (SDPR) to further enhance the classification accuracy by regularizing the tunable weights based on prototype alignment. The proposed deep learning (DL) model is evaluated against traditional machine learning (ML) methods, such as random forest (RF), and demonstrates superior performance on key metrics, including accuracy, precision, recall, and F1-score.

Corresponding Author:

Pooja Hadimane
Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology
(Affiliated to Visvesvaraya Technological University, Belagavi)
Davangere, India
Email: poojah_12@rediffmail.com

1. INTRODUCTION

Pollinators play an essential role in maintaining biodiversity and ensuring agricultural productivity, with over 80% of flowering plants and 75% of global food crops relying on pollination. However, pollinator populations are under threat due to environmental changes, habitat destruction, and climate change, which pose significant risks to both ecosystems and food security. Traditional methods of identifying and classifying pollinator species are labor-intensive and limited by the decreasing number of taxonomists. These challenges call for advanced solutions that leverage artificial intelligence (AI) and deep learning (DL) to automate the classification process, providing accurate and scalable systems for monitoring pollinator species. This paper introduces a novel DL framework that integrates multimodal data, including visual and environmental factors, to enhance pollinator classification and vulnerability assessment, addressing a critical gap in ecological studies and conservation efforts.

In natural environments, many animals are interspecifically biotic with plants, a case of herbivores and pollinators. Plants having defensive systems are placed under heavy selective pressures from herbivores that eat plant material on a large scale by herbivores. These types of defenses vary in characteristics based on the chemical sensitivity and the feeding behaviors of herbivores of various species. Pollinators, on the other hand, by means of mutualistic interactions with plants, will select some physical trait and several signals of increased success of the pollen transport. Selection pressures by different pollinators can maintain the different evolution in plant populations and species since they possess different physical features and sensory preferences. The ecological components are interconnected differently; the soil type and the plant-animal interactions are some examples of this. In a process of phenotypic plasticity, the soil type has a huge influence on the plant physiology and morphology, and subsequently, it impacts the plant-pollinator and plant-herbivore interactions. Massive deaths of insects have been reported even in the so-called pristine environments. These reports contribute to the existing issue of biodiversity on the globe. There is mounting evidence to indicate that insect losses are dramatic both in terms of taxonomy and geography. Simple processes in the ecosystem, such as pollination, decomposition, and management of biological pests, have many species that rely on them, such as human beings. The loss of these processes is a serious threat to the stability of the ecosystems and biodiversity [1]–[3]. Insect pollinators are vital to the welfare of the terrestrial ecosystem as well as human life. By a count of 8%, of all the 1,400 crop species grown worldwide to provide food to humans and plant-based industrial products, animal pollination is required. Pollinators are required to ensure that ecosystems are preserved, food production, and natural resources are preserved through ecosystem services. As highly valuable resources, both managed and wild populations of pollinators are facing severe pressures with a significant consequence to the world's food security. The decline in diversity and the number of insects is one of the most significant issues that humankind faces today. The insects offer special impediments to population surveillance due to their small size and the laborious nature of more traditional mechanisms of their collection and identification. It is increasingly becoming difficult to identify the pollinators, particularly bees, taxonomically, due to the dwindling number of insect taxonomists. Moreover, one cannot find many readily accessible resources to educate the general population and future generations on the ability to identify different insects and pollinator species.

Other factors that result in the loss of pollinators have been statistically modeled of opportunistic data and field studies, such as pesticide exposure, land use, and climate change. Nonetheless, one cannot tell how relevant each of them is concerning the cause of decreases at varying sizes unless there is long-term abundance data. In addition, the primary concern of the regulation of the wild pollinators in the United Kingdom (UK) and the rest of the countries is the subsidies offered to farmers under the agri-environment scheme. These subsidies promote the application of agri-environment methods, including field margins that are high in flowers and hedge preservation. The beneficial effects have been realized in these efforts in reducing local scales with respect to the pollinator variety, activity, and pollination services, but more likely, they are short-lived termini of flower visits, not permanent population increase [4].

Agricultural technology has advanced significantly, but there are still no systems designed expressly for the automated detection and classification of pollinators. The majority of current approaches target insects and pests; thus, creative solutions are needed to solve this urgent pollinator-related issue. Identification and categorization of pollinator species are essential for ecological studies and agricultural uses. Because they allow many flowering plants to reproduce, pollinators—insects like bees, butterflies, and moths—are crucial to maintaining biodiversity and ecosystems. For several reasons, these species must be precisely classified and identified [5].

Conservation of biodiversity: over 80% of flowering plants rely on pollinators for reproduction, highlighting their critical role in maintaining biodiversity. By accurately identifying and categorizing pollinator species, scientists can monitor fluctuations in species populations and enhance their comprehension of ecological diversity. Conservation initiatives aimed at protecting threatened species and ensuring the overall health of ecosystems are critical. Agricultural productivity: pollinators are responsible for the successful production of approximately 75% of the world's food crops. The essential components for human nutrition and food security encompass fruits, vegetables, nuts, and seeds. Accurate classification and monitoring of pollinator species contribute to assessing the health of these populations, ensuring their ability to provide essential services. Understanding the efficiency of various species as pollinators for specific crops enables farmers and agriculturalists to enhance their pollination strategies, thereby increasing both crop yields and quality. Ecosystem services and functionality: pollinators play a critical role beyond agriculture by supporting essential ecosystem services, including the restoration of natural habitats and the enhancement of genetic diversity among plant species. Accurate identification and classification of pollinator species enable researchers to monitor ecosystem functioning and assess the condition of natural environments.

The equilibrium of ecosystems is contingent upon this knowledge, particularly in the context of environmental changes such as pollution, habitat degradation, and climate change. Climate change impact assessment: the evaluation of climate change effects on ecosystems necessitates accurate classification and identification of pollinator species. The distribution and behavior of pollinators are influenced by fluctuations in temperature, precipitation, and various climatic factors. Monitoring species diversity and population dynamics enables researchers to detect early signs of climate-induced stress in ecosystems and implement protective measures accordingly.

Pollination management and conservation strategies: effective management of pollinator populations is essential for sustaining their pollination roles in both natural ecosystems and agricultural environments. The identification and classification of pollinator species contribute to the formulation of targeted conservation strategies. These strategies encompass the preservation of critical pollinator habitats, habitat restoration efforts, and the creation of environments that support pollinator populations. This approach ensures the availability of pollinators, which is essential for the maintenance of natural ecosystems and the sustainability of the food supply. Control of invasive species: the management of invasive species that pose a threat to native pollinator populations is significantly reliant on the precise identification of pollinator species. Invasive species have the potential to alter pollination networks, leading to disruptions in local ecosystems and the outcompeting of native pollinators. Management plans aimed at conserving native pollinators and maintaining ecological balance can be developed through the identification and monitoring of these species. Figure 1 shows an example of a plant species *(Araliaceae—Oreopanax capitatus (Jacq.) Decne. and Planch.)* with phenotypic generalization visited by diverse small insects. Figure 1(a) shows the butterfly, Figure 1(b) shows the stingless bees and lepidoptera, Figure 1(c) shows the fly, and Figure 1(d) shows the wasp.

(a) (b) (c) (d)

Figure 1. Example of a plant species *(Araliaceae—Oreopanax capitatus (Jacq.) Decne. and Planch.)* with phenotypic generalization visited by diverse small insects for (a) butterfly, (b) stingless bees and lepidoptera, (c) fly, and (d) wasp

Pollinator species identification and categorization are changing as a result of advances in AI, machine learning (ML), and DL. These technologies improve accuracy and make it possible to accurately identify species that can be challenging to differentiate manually by analyzing vast datasets. Real-time results are critical to large-scale ecological studies and agriculture, and AI and ML can help by automating the identification process. Compiling behavioral, environmental, and visual data together makes it easier to comprehend pollinator species and how they interact with their surroundings. These components are essential for forecasting future patterns, evaluating how pollinators may be affected by climate change, and developing conservation plans. Global conservation efforts may be supported by AI-driven systems that guarantee consistency, minimize human error, and can be scaled to monitor pollinator numbers across large territories [6]–[9].

The pollination vulnerability index (PVI) is one important tool for assessing the potential decline in pollination services. Several plants, notably those that are key to agricultural yield and the ecosystem, rely on pollination services for their reproduction process. Over the past few years, DL has become a major trend in tackling complex aspects of classification, especially those requiring large and varied types of datasets, where this technique and its variants have become especially popular. The fact that DL models are capable of providing both binary and multi-classification, in which data is divided into two classes, respectively, and one of multiple classes, respectively, is one of the strengths. DL in the ecological monitoring context can be used to effectively and efficiently solve problems of identifying and classifying species. The models will be able to learn more detailed patterns in visual and textual data using layers of abstraction, meaning that they will be more accurate than the other methods of ML. DL is especially adaptable, which is why it is especially appropriate in solving problems such as pollinator species identification, where a binary question (pollinator vs. non-pollinator) needs to be answered, but a multiclass one (what genus is it) has to be posed simultaneously [10]. Pollinators are critical to maintaining biodiversity and supporting global food production, as over 80% of flowering plants and 75% of the world's food crops rely on animal pollination. Despite their importance, pollinator populations, particularly bees, are declining at alarming rates due to factors like habitat destruction, pesticide use, and climate change. Traditional methods for identifying and monitoring pollinator species are labor-intensive, rely heavily on expert taxonomists, and lack scalability, which limits their ability to track pollinator populations on a large scale. This creates a pressing need for automated, accurate, and scalable solutions to monitor pollinator diversity and health. Advances in AI and DL offer promising tools to overcome these limitations, but current models often fail to integrate multiple data types (such as visual and environmental data) and lack the precision required for fine-grained species classification.

This research is motivated by the urgent need to develop a robust, AI-driven framework that can not only automate pollinator classification but also improve the accuracy of predictions by integrating multimodal data, addressing both binary and multiclass classification challenges in an ecological context. Research contribution includes:

i) Multimodal species-image regularization (MSIR) framework: we introduce a novel DL framework, the MSIR, that integrates both visual and species-detail data to improve classification performance. This framework enables the system to handle both binary classification (pollinator vs. non-pollinator) and multiclass classification (classifying bee genera) more effectively than traditional methods.
ii) Prototype-driven regularization: the framework employs image-species prototype regularization (ISPR) and species-detail prototype regularization (SDPR) to enhance the model’s accuracy. These techniques align the tunable weights of the model based on both image and species prototypes, leading to better generalization across different species and environmental conditions.
iii) Superior performance metrics: the proposed model outperforms traditional ML methods, such as random forest (RF), achieving an accuracy of 98.34%, significantly surpassing the baseline of 92%. This showcases the potential of DL models to revolutionize pollinator monitoring by offering more precise and scalable solutions.

2. RELATED WORK

In the recent past, advancements in DL and AI have enhanced the capacity to recognize and categorize insects and other species within the ecology. Such advances underscore the importance of such technologies in the agricultural and food sector, and the conservation of biodiversity. For example, in research [11], investigated object recognition algorithms in video footage to apply AI systems to detect thistle caterpillars *(Vanessa cardui)*. With a 2,416-photo dataset, the YOLOv5 object identification architecture was trained with scratch and transfer learning models. The results can be treated positively as it indicates that DL is an efficient method to detect insects. Selvaraj *et al.* [12] applied a deep convolutional neural network (DCNN) to identify the diseases on bananas using a dataset of over 18,000 pre-screened banana photos and found that transfer learning can be effective in agriculture. The level of accuracy achieved ranged between 70% and 99%. The purpose of subsequent research [13] was to predict the occurrence of insects by maintaining a gaze at the environmental conditions, such as relative humidity and air temperature. An ecological monitoring using ML in combination with the environment demonstrated a 76.5% detection accuracy. In a paper authored by Marković *et al.* [14], the problem of drone-based insect and pest detection has been explored, which further proves that AI technology can be used in various conditions when applied to monitoring. With the MobileNet architecture, the research paper in [15] had an accuracy of 97.5 in defining and categorizing pests and tomato diseases among insects. This indicates that convolutional neural networks (CNNs) can have applications in agriculture, combined with transfer learning, with possible benefits. This strategy complies with the recent methods of employing small and efficient models to be applied in agriculture on-time. Technological developments have resulted in the innovation of new ways of automated detection of insects. A classical model of the combination of DL and optimization methods is the automated insect detection and classification approach [16] software. It extracts features through DenseNet121 and the Pelican optimization algorithm. To determine insects, Assiri *et al.* [17] proposed an optimizer, which is an artificial ecosystem-based model that was fused with MobileNetV2. This highlights the rising trend of hybrid models incorporating a number of AI techniques. Besides, there is an evident proof using the advanced neural networks in pest control, as it is demonstrated in the reference [18] when the farmland fertility algorithm is built to automatically detect the rice pests. The models that were used to extract features to perform the classification process were the Elman recurrent neural network (ERNN) and neural architecture search network-large (NASNetLarge). Likewise, in researchers [19], [20] has employed the method of innovative insect detection and classification (that is, a combination of DenseNet and Gannet optimization algorithm) to show the further development of insect-detecting models. AI has gone a long way in its effort to assimilate into the pollination of plants. It has been suggested that a CNN and AI can be applied to reliably identify flowers that can be pollinated by drones or robots. Also, Hiraguri *et al.* [21] enhanced the automation of pollinator identification by developing CNN models of bee species that pollinate blueberries and combining them with log Mel-spectrogram representations. Furthermore, Ferreira *et al.* [22] enhanced the automated insect surveillance, both in terms of insect recognition using DL models and in terms of providing a training dataset that is enabled by a specifically developed time-lapse camera system. This is one of the ideal examples of a DL algorithm and real-time data gathering techniques. Likewise, Bjerge *et al.* [23] developed a multisensory imaging system that can be applied to the conventional insect traps to enhance the quality of the requested images to be effective in the categorization of the species under precise taxonomic classification, and Brandt *et al.* [24] proposed a CNN-based detection system to the real-time classification of the live fruit fly species, showing the application of AI in pest monitoring of the types of species, such as olive and Mediterranean fruit flies, which are highly dangerous to global agriculture.

DL and AI have achieved a lot in the sphere of insect recognition and classification, where most studies have largely been focused on the environmental or visual image. When it comes to multimodal data integration, more research is needed to classify the species of pollinators and calculate the risk of pollination. Through the integration of environmental and visual data, this research will fill its existing gap and generate a comprehensive and accurate approach to the definition of pollinator species and the assessment of the PVI. It enhances better categorization and presents informative data on farm and conservation activities. Although AI and DL have made a massive leap in the insect identification process, most researchers have studied visual or environmental data in isolation, restraining the overall precision and usability of both in ecological research [25]. The existing models, though useful in identifying specific species of pests or in diagnosing agricultural diseases, do not reflect the variation of the pollinator species classification and may need to have a more comprehensive view of understanding the ecological interactions and multimodal data. Also, there is a lack of literature on the combination of environmental and visual data to determine the susceptibility of pollination populations to climate change and habitat loss. The existence of this gap shows that additional and more extensive models are necessary to include ecological and environmental information to classify the pollinators and their vulnerability. The absence of automated and scalable systems to track species of pollinators, which are important in biodiversity and in agriculture, is also a critical gap that is critical.

3. PROPOSED METHODOLOGY

The proposed MSIR framework for multi-modal image classification explores fine-tuning the multimodal contrastive learning framework (MCLF) model for image classification by integrating both its visual and textual encoders. Initially, it defines a loss function that minimizes the discrepancy between predicted probabilities and actual class labels using contrastive learning objectives. Two primary frameworks are examined: the image-feature-based framework, which inputs visual features into a tunable classifier, and the language-encoder-based framework, which adjusts the text encoder with learnable token embeddings but faces alignment issues between visual and textual features. To overcome this, the analysis introduces prototype-driven visual and text regularization methods. These methods generate class prototypes from the visual and textual encoders, respectively, and regularize the tunable weights by minimizing the Kullback-Leibler divergence between these prototypes and the classifier weights. The overall loss function combines the cross-entropy loss with these regularization terms, weighted by balance factors. Figure 2 presents the proposed workflow.

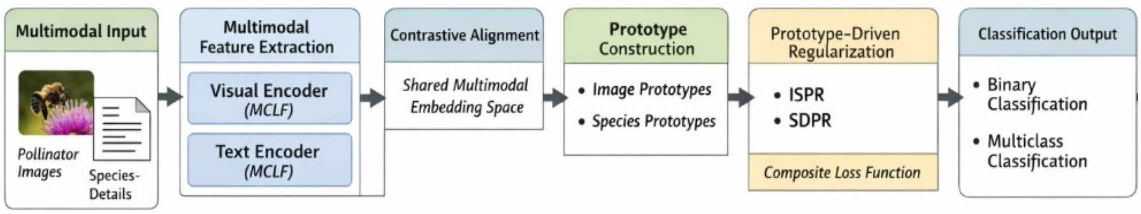

Figure 2. Proposed workflow

This approach aims to enhance model performance by ensuring better alignment and generalization across both modalities. Figure 3 shows the proposed architecture. The MSIR framework integrates visual and species-detail information for robust pollinator classification. Step 1: pollinator images and corresponding species-detail attributes are provided as multimodal inputs. Step 2: image features are extracted using the visual encoder of the MCLF, while species details are encoded using its text encoder. Step 3: a contrastive learning objective aligns visual and textual embeddings in a shared feature space. Step 4: class-level image and species-detail prototypes are constructed to represent each class. Step 5: ISPR and SDPR align classifier weights with their respective prototypes, reducing cross-modal feature misalignment. Step 6: the network is optimized using a composite loss combining classification and prototype-based regularization. Step 7: the trained model performs binary and multiclass pollinator classification.

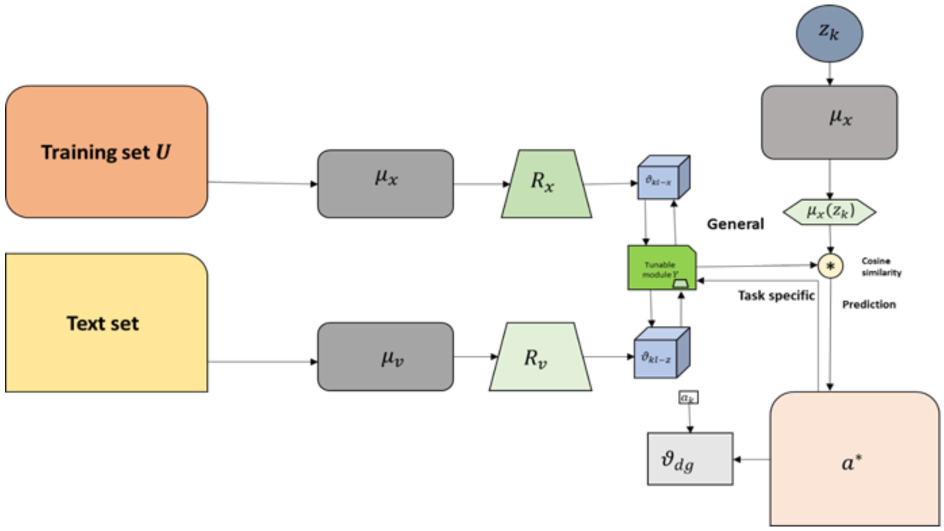

Figure 3. Proposed architecture

3.1. Multimodal contrastive learning framework model

The MCLF represents a significant breakthrough in ML by demonstrating how natural language supervision can be used to train state-of-the-art visual models. Unlike traditional computer vision systems, which rely on predefined categories of objects, MCLF learns from a vast dataset of image-text pairs, 400 million in total, collected from the internet. The model is trained to predict which text is associated with which image using a contrastive learning approach, rather than relying on labelled datasets like ImageNet. This allows the MCLF model to generalize across various tasks and perform zero-shot transfer, where it can classify images that it has never seen before without additional training.

MCLF achieves its versatility by jointly training an image encoder and a text encoder. During training, the model maximizes the cosine similarity between paired image and text representations while minimizing similarity for mismatched pairs. This results in a shared embedding space for both images and text, which can be leveraged to perform classification tasks in a zero-shot manner. Instead of training a classifier with labelled images, MCLF can use the textual descriptions of target classes to create a classifier on-the-fly. This ability to use natural language as a supervisory signal, MCLF, enables it to outperform or match fully supervised models in many computers vision tasks, including object recognition, optical character recognition (OCR), and even action recognition in videos. The MCLF model's success highlights the potential of scaling up learning from web-scale natural language data. By training on massive, diverse datasets, MCLF demonstrates improved robustness and transfer learning capabilities compared to models trained on limited, labelled datasets. Its zero-shot performance is competitive across a wide range of tasks and domains, including fine-grained image classification and localization tasks. This opens up new possibilities for developing more generalized AI systems that can learn from diverse data sources without the need for extensive task-specific datasets.

3.1.1. Preliminary analysis

By assuming a common approach, a pre-training mechanism that learns to associate images with textual descriptions using contrastive learning objectives, which is fine-tuned for classifying images denoted by δ, by considering a training set, $U = \{(z_k, a_k)\}_{k=1}^{P}$, here z is an image and a is the one-hot encoding, a tunable parameter denoted by y along the MCLF model, $\tau = \{\tau_x, \tau_v\}$ wherein τ_x and τ_v depict the visual encoder, the learning objective leads to minimizing a loss denoted. However, $r(z_k, y, \mu)$ depicts the probability for z_k the softmax class, denoted by the confidence vector, the fine-tuning vectors applied on the various optimization strategies for y along with the various calculation techniques for r respectively in (1).

$$\vartheta_{dg}(U, y, \mu) = -\frac{1}{P}\sum_{k=1}^{P} a_k logr(z_k, y, \mu) \tag{1}$$

3.1.2. Image-species classification framework

Linear probe network denoted by δ is the basic method used for the image-feature-based fine-tuning framework, which initially extracts the visual features? $\tau_x(z_k)$ for the image z_k along with the image feature τ_x in (2). Then the visual features are fed into the tunable weights denoted by y, considered a usual classifier to evaluate the probability. The loss function is denoted. Here $exp\ (y.\mu_x(z_k))_l$ depicts the exponential logic for the $l-th$ class and represents the dot product in (3).

$$\alpha_{dg-x}(U, y, \mu_x) = -\frac{1}{P}\sum_{k=1}^{P} a_k logr(z_k, y, \mu_x) \tag{2}$$

$$r(z_k, y, \mu) = \frac{exp\ (y.\mu_x(z_k))}{\sum_l exp\ (y.\mu_x(z_k))_l} \tag{3}$$

3.1.3. Species attribute encoding framework

The prompt learning framework acts as the baseline method for the species attribute encoding framework (SAEF), and the tunable weight. y considered as learnable token embeddings for each class of the text encoder. This further concatenates y with the embeddings of each class text a^v, the text is fed into the text encoder μ_v to derive the text features represented as $\mu_v(y, a^v)$. This reduces the similarity between the visual features $\mu_x(z_k)$ and the species features for ground truth class text in (4). The loss function is determined here in (5).

$$\alpha_{dg-x}(U, y, \mu_x) = -\frac{1}{P}\sum_{k=1}^{P} a_k logr(z_k, a^v, y, \mu_x, \mu_v) \tag{4}$$

$$r(z_k, a^v, y, \mu_x, \mu_v) = \frac{exp\ (cos(\mu_v(y,a^v),\mu_x(z_k)))}{\sum_l exp(cos\ (\mu_v(y,a^v),\mu_x(z_k)))_l} \tag{5}$$

Wherein $cos(.,.)$ represents the cosine similarity calculation as $\sum exp(\cos\ (\mu_v(y, a^v), \mu_x(z_k)))_l$ denotes the exponential prediction for the $l-th$ class. The language-encoder-based framework fine-tunes the text input and then inherits the mechanism δ. The performance is poor because of the misalignment between the characteristics that the text and visual encoders extract.

3.1.4. Negative log-likelihood loss

Upon fixing the visual representation layer μ_x with the text encoder μ_v along with the cosine classifier with the adjustable weight y upon the visual encoder μ_x. Based on the visual encoder μ_x. The cross-entropy loss is given. The key difference between the proposed and the visual encoder-based framework is that the bias term associated with it is y essential for the regularization process, propose prototype-driven visual regularization and prototype-driven text regularization to utilize the visual and text encoder of the MCLF model.

3.1.5. Image-species prototype regularization

Based on the visual encoder μ_x, the image-based prototype model R_x used for the o the class as the visual features for all the samples z_k within the class o. However, M_o denotes the number of samples in class o with $R_{x,o}$ is the $o-th$ row of R_x. Then, regularize the tunable weight. y with the visual prototype. Here O depicts the total number of classes, y_o is the $o-th$ row of y and ϑ_{kl} denotes the Kullback-Leibler divergence between two vectors by (6) and (7).

$$R_{x,o} = \frac{1}{M_o}\sum_{(a_k)_o=1} \mu_x(z_k) \tag{6}$$

$$\vartheta_{kl-x} = \frac{1}{O}\sum_{o=1}^{O} \tau_{kl}(R_{x,o}||y_o) \tag{7}$$

3.1.6. Species-detail prototype regularization

Based on the text encoder μ_v, the text-based prototype R_v for the $o-th$ class as the text feature of the class text a^v, that is expressed in (8). Here, a_o^v is considered the text for the $o-th$ class, by using the text encoder, a prompt is designed to prefix the a_o^v. Similarly, the text-based prototype regularization is given in (9). Wherein O is the class size, y_o is the $o-th$ row of o and τ_{kl} represents the Kullback-Leibler divergence between two vectors. Depending on the cross-entropy loss, the visual and text-based regularization, the total loss ϑ_{loss} denoted is given in (10). However, here α_1, α_2, and α_3 depict the balance factors.

$$R_{v,o} = \mu_v(a_o^v) \tag{8}$$

$$\vartheta_{kl-v} = \frac{1}{O}\sum_{o=1}^{O} \tau_{kl}(R_{v,o}||y_o) \tag{9}$$

$$\vartheta_{loss} = \alpha_1\vartheta_{df} + \alpha_2\vartheta_{kl-dis} + \alpha_3\vartheta_{kl-v} \tag{10}$$

4. RESULTS AND DISCUSSION

The dataset used for pollinator species classification consists of detailed field observations of insect visits to flowering plants, along with associated environmental and land-use information. Each data record includes contextual variables such as the date, latitude, longitude, site description, and weather, which provide a rich set of features for model training. The accuracy, precision, recall, and F1-score are the key performance metrics used to evaluate the models. Accuracy measures the overall correctness of the classifications, while precision focuses on the proportion of true positive classifications among all positive classifications. Recall evaluates the model's ability to identify all relevant instances, and the F1-score provides a balance between precision and recall. These metrics collectively help assess the effectiveness of the models in accurately classifying pollinator species.

4.1. Dataset details

The data set shall contain 5,687 records of insect specimens, the subject matter of which will be the pollinator species. An in-depth review of the dataset indicates that pollination species of known species count 5,573, and non-pollinators count 114 (when considering the genus *Botanophila* that consists of mostly seed-feeding flies). The data contain insects belonging to three major orders, namely, Hymenoptera (5,348 records), Diptera (332 records), and Coleoptera (7 records). Hymenoptera, predominantly species of the family Apidae is home to key pollinators, honeybees (*Apis* genus) and bumblebees (*Bombus* genus), which make up a large part of the dataset. The other genus to note is *Melissodes, Lasioglossum*, and *Halictus*, which are all useful in the pollination of crops and wild flowers. The analysis conducted at the species level shows that *Apis mellifera* (the European honeybee) has the largest percentage with 4,548 entries, indicating that it is important in agricultural pollination. Also, species such as *Melissodes trinodis, Bombus griseocollis*, or *Bombus ternarius* are the most significant pollinators of native plants and crops. The *Botanophila* genus, 114 records, in contrast, consists almost solely of non-pollinating species that are more accurately described as seed feeders and which contribute little to pollination. The image data of our dataset is shown in Figure 4.

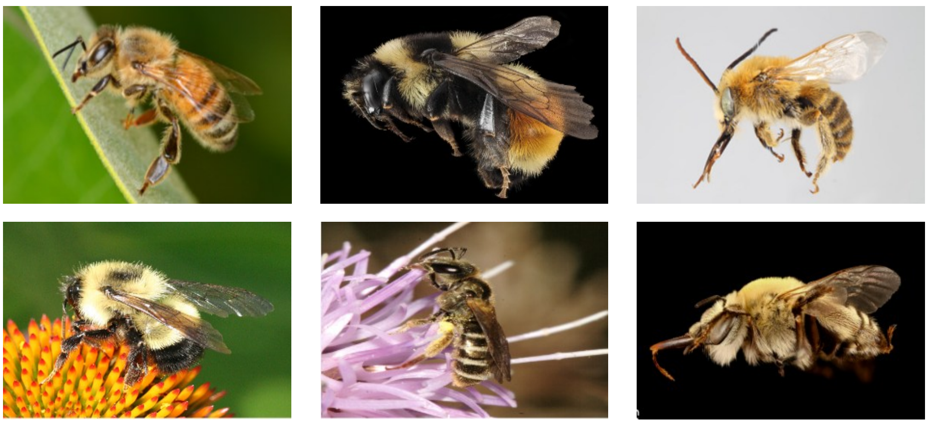

Figure 4. Image dataset

4.2. Results

4.2.1. Binary classification

A binary classification would be seen as a division of the data set into two different categories, including pollinator vs. non-pollinator. It makes this problem easier to classify and is aimed at determining whether an insect is considered to play a role in pollination or not. As an illustration, genera such as *Apis*, *Bombus,* and *Lasioglossum*, all common pollinators, can be termed as pollinators. Conversely, any genus not typical of pollination (had the pollinating data available in the dataset would be termed a non-pollinator. This binary classification exercise aims at estimating whether a particular insect, judged by its characteristics (such as insect family, tribe, and species), will be described either as part of the pollinator or as a non-pollinator. The model is trained to discriminate between these two classes of features given, and it is a single binary label, which is 0 (non-pollinator) or 1 (pollinator).

The chart is a comparison between the accuracy of two models, the RF and our proposed DL network, with pollinator species classification. DL model is highly more effective than the RF with an approximate of 98.34% accuracy when compared to 92. The finding shows the deeper applicability of the DL method to the correct classification of the pollinator species under our multimodal approach. Figure 5 represents the analysis of accuracy.

The chart compares the precision of the RF model and our proposed DL network for pollinator species classification. The DL model demonstrates significantly higher precision, around 95%, compared to the 85% achieved by RF. This indicates that the DL model is more effective in accurately identifying relevant pollinator species while minimizing false positives. Figure 6 shows the precision comparison.

The chart compares the recall of the RF model and our proposed DL network for pollinator species classification. The proposed DL network exhibits a higher recall, around 97%, compared to the 92% achieved by RF. This indicates that the DL model is more effective at identifying most of the relevant pollinator species, minimizing false negatives in the process. Figure 7 shows the recall comparison.

The chart compares the F1-score of the RF model and our proposed DL network for pollinator species classification. The proposed DL network achieves a significantly higher F1-score, around 95%, compared to the 87% achieved by the RF. This indicates that the DL model strikes a better balance between precision and recall, making it more effective overall in accurately classifying pollinator species. Figure 8 shows the F1-score comparison.

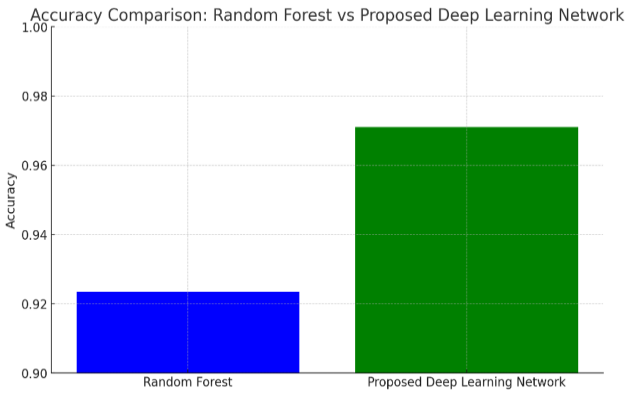

Figure 5. Accuracy comparison

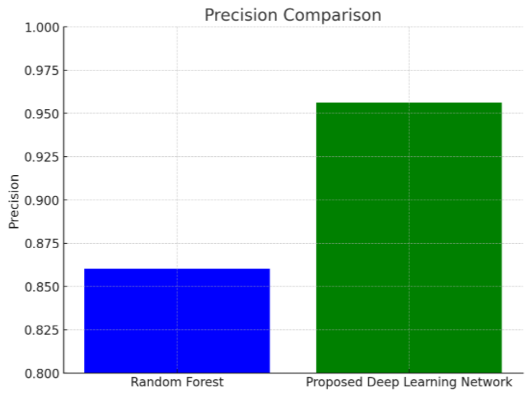

Figure 6. Precision comparison

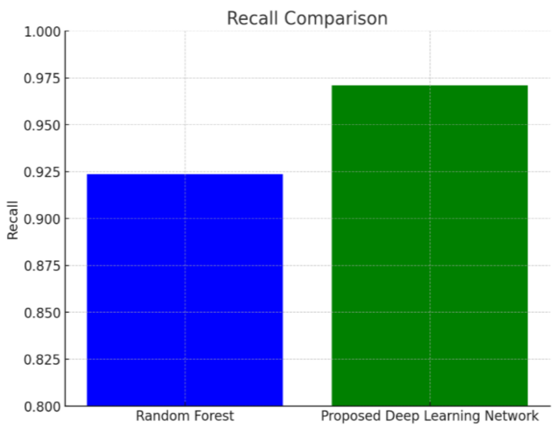

Figure 7. Recall comparison

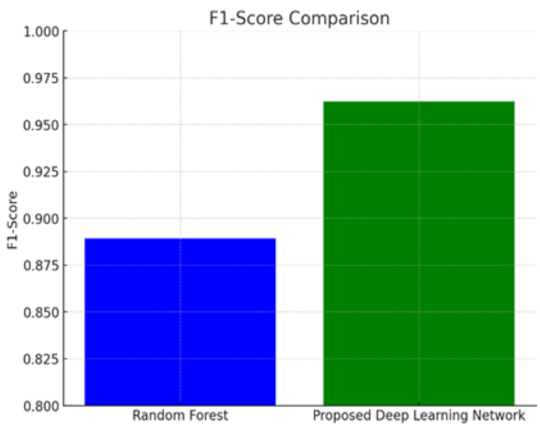

Figure 8. F1-score comparison

4.2.2. Multiclassification

In a multiclass classification task, the goal is to classify each bee into one of several possible categories, where each category represents a specific bee genus. In this scenario, the model assigns each insect to a particular genus, such as *Apis*, *Bombus*, or *Lasioglossum*. Instead of grouping the data into just two categories, like in binary classification, each bee genus is treated as its own class. For instance, class 0 could represent *Apis* (the honeybee genus), class 1 could represent *Bombus* (the bumblebee genus), class 2 could represent *Lasioglossum* (the sweat bee genus), and so on for the other genera. The model's objective is to predict the genus of each bee based on its features, such as family, species, or tribe. The model will output one class, identifying the genus to which the bee belongs. To evaluate the model's performance, metrics like precision, recall, and F1-score are calculated for each class, showing how well the model correctly predicts each bee genus. Figure 9 shows the multiclass classification.

The chart shows the accuracy, recall and F1-scores of each of the bee genera in the framework of a multiclass classification problem, the objective of which is to classify bee species correctly into their corresponding genera. The genus that are represented are *Agapostemon, Andrena, Apis, Lasioglossum*, and so on, with scores 0 to 1, showing the ability of the classification model. In the vast majority of genera, including *Apis, Agapostemon, Ceratina, Dufourea, Halictus, Megachile*, and *Melissodes*, the model has very good performance, with almost perfect precision, recall, and F1-scores. This implies that the model is very useful in the detection of these genera, with few classification errors and mostly predicts most of these samples.

Nonetheless, the performance of the model with regard to a few genera, such as *Andrena* and *Hylaeus*, is slightly disproportionate. In particular, *Andrena* has low recall and F1-score when compared to its precision. It means that although the model is effective in recognizing *Andrena* when it makes a prediction, it cannot recognize all real examples of this genus, which is why the recall decreases. The same trend is observed with *Hylaeus*; in that case, recall is somewhat lower, indicating the model is missing some real *Hylaeus* bees in its model. The confusion matrix is displayed in Figure 10.

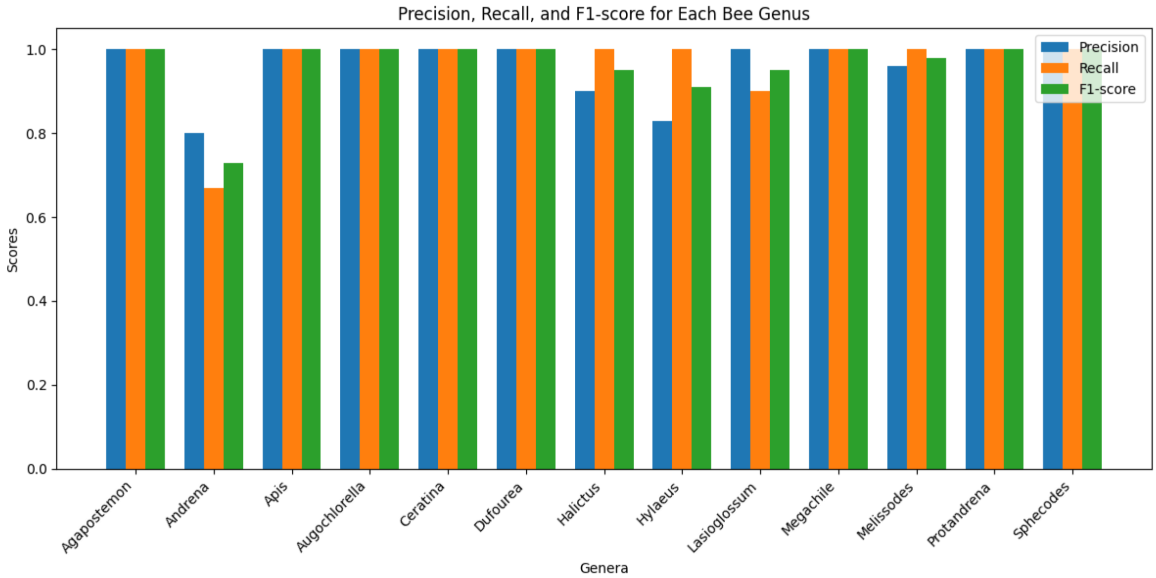

Figure 9. Multiclass classification

The confusion matrix indicates that the model is doing really well at classifying the genus *Apis*, with 919 correct and no misclassifications. This means that the model is accurate and has good recall. There are also other genera, such as *Melissodes* and *Lasioglossum*, in which the performance is high (27 and 18 correct classifications, respectively). There are minor misclassifications, e.g., *Lasioglossum* is confused with *Andrena*. There is some misclassification in the genus *Andrena*, where only 4/6 cases were correctly determined, and 1/6 of each case was misplaced in the genus *Lasioglossum* and the genus *Ceratina*. Smaller genera such as *Agapostemon*, *Augochlorella*, and *Dufourea* have been placed into place, usually despite having fewer samples and indicate the overall good multiclassification capability of the model.

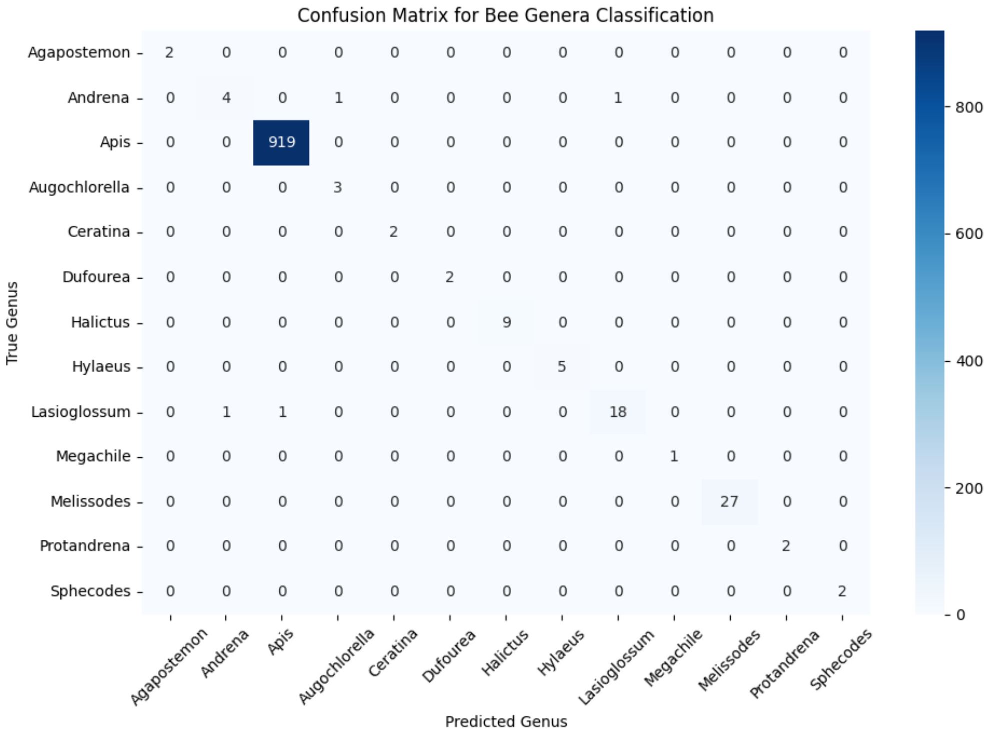

Figure 10. Confusion matrix

5. CONCLUSION

This paper introduces a novel DL framework that effectively improves the automated classification and identification of pollinator species through integrating both visual and environmental data. Pollinators are essential to maintaining biodiversity and supporting agricultural productivity, as over 80% of flowering plants and 75% of global food crops depend on them. However, pollinator populations are in steep decline due to environmental changes, habitat destruction, and climate change, posing significant threats to ecosystems and food security. Traditional methods for identifying pollinators are labor-intensive, time-consuming and heavily reliant on taxonomists, which limits their scalability. To overcome these challenges, we developed a scalable, AI-driven solution that outperforms traditional ML models like RF. MSIR combines both image and environmental data, resulting in superior classification accuracy and generalization. The designed framework demonstrates a significant improvement in pollinator species classification, achieving an accuracy of 98.34%, compared to the 92% accuracy achieved by RF models. These results highlight the potential of AI-based approaches in real-time ecological monitoring and conservation efforts, providing essential tools to mitigate the decline of pollinator populations. A key direction for future work is the integration of the PVI into this designed framework. PVI offers a critical measure for assessing the risk of pollination service declines based on species' vulnerability to environmental factors such as climate change, habitat fragmentation, and pesticide exposure. By incorporating PVI into the existing model, we aim to enhance its predictive power by not only identifying pollinator species but also evaluating the risk levels for these populations in various ecological contexts. Scalability and real-time deployment remain challenges due to the computational cost of multimodal learning. Future work will focus on larger datasets and integrating the PVI for risk-aware pollinator assessment.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to everyone who has supported and contributed to this research project. In particular, the authors extend their deepest gratitude to our advisor for his tireless guidance, invaluable insights, and encouragement throughout the research process.

FUNDING INFORMATION

No funding is raised for this research.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Pooja Hadimane	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Ashoka Kukkuvada	✓	✓			✓	✓		✓	✓	✓				

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

Author declares no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available in Frontiers in Plant Science at http://doi.org/10.3389/fpls.2023.1081050, reference number [22].

REFERENCES

[1] E. Barreto *et al.*, "Macroevolution of the plant–hummingbird pollination system," *Biological Reviews*, vol. 99, no. 5, pp. 1831–1847, Oct. 2024, doi: 10.1111/brv.13094.

[2] Y. Gao *et al.*, "Application of machine learning in automatic image identification of insects-a review," *Ecological Informatics*, vol. 80, May 2024, doi: 10.1016/j.ecoinf.2024.102539.

[3] P. deMaynadier *et al.*, "Insect pollinators: the time is now for identifying species of greatest conservation need," *Wildlife Society Bulletin*, vol. 48, no. 3, Sep. 2024, doi: 10.1002/wsb.1537.

[4] C. P.-Kairath *et al.*, "Insect pollination in deep time," *Trends in Ecology & Evolution*, vol. 38, no. 8, pp. 749–759, Aug. 2023, doi: 10.1016/j.tree.2023.03.008.

[5] A. R. S. Parmezan, V. M. A. Souza, A. Seth, I. Žliobaitė, and G. E. A. P. A. Batista, "Hierarchical classification of pollinating flying insects under changing environments," *Ecological Informatics*, vol. 70, Sep. 2022, doi: 10.1016/j.ecoinf.2022.101751.

[6] M. N. Ratnayake, D. C. Amarathunga, A. Zaman, A. G. Dyer, and A. Dorin, "Spatial monitoring and insect behavioural analysis using computer vision for precision pollination," *International Journal of Computer Vision*, vol. 131, no. 3, pp. 591–606, Mar. 2023, doi: 10.1007/s11263-022-01715-4.

[7] M. M. Moreira and L. Freitas, "Review of the pollination system by small diverse insects," *Neotropical Entomology*, vol. 49, no. 4, pp. 472–481, Aug. 2020, doi: 10.1007/s13744-020-00779-6.

[8] M. E. Saunders, L. K. Kendall, J. B. Lanuza, M. A. Hall, R. Rader, and J. R. Stavert, "Climate mediates roles of pollinator species in plant–pollinator networks," *Global Ecology and Biogeography*, vol. 32, no. 4, pp. 511–518, Apr. 2023, doi: 10.1111/geb.13643.

[9] R. Nath, H. Singh, and S. Mukherjee, "Insect pollinators decline: an emerging concern of Anthropocene epoch," *Journal of Apicultural Research*, vol. 62, no. 1, pp. 23–38, Jan. 2023, doi: 10.1080/00218839.2022.2088931.

[10] S. Ba, C. Zhao, Y. Liu, and Q. Fang, "Constructing a pollination network by identifying pollen on insect bodies: consistency between human recognition and an AI model," *Biodiversity Science*, vol. 32, no. 6, 2024, doi: 10.17520/biods.2024088.

[11] E. Önler, "Real-time pest detection using YOLOv5," *International Journal of Agricultural and Natural Sciences*, vol. 14, no. 3, pp. 232–246, 2021.

[12] M. G. Selvaraj *et al.*, "AI-powered banana diseases and pest detection," *Plant Methods*, vol. 15, no. 1, Dec. 2019, doi: 10.1186/s13007-019-0475-z.

[13] S. H. Sreedhara, V. Kumar, and S. Salma, "Efficient big data clustering using ad hoc fuzzy c means and auto-encoder CNN," in *Inventive Computation and Information Technologies*, 2023, pp. 353–368, doi: 10.1007/978-981-19-7402-1_25.

[14] D. Marković, D. Vujičić, S. Tanasković, B. Đorđević, S. Ranđić, and Z. Stamenković, "Prediction of pest insect appearance using sensors and machine learning," *Sensors*, vol. 21, no. 14, Jul. 2021, doi: 10.3390/s21144846.

[15] N. M.-Gonzales and M. J. Brewer, "A special collection: drones to improve insect pest management," *Journal of Economic Entomology*, vol. 114, no. 5, pp. 1853–1856, Oct. 2021, doi: 10.1093/jee/toab081.

[16] L. Yang, L. Yu, S. Tao, Z. Yang, W. Gao, and Y. Ren, "Identification of tomato pests and diseases based on transfer learning," *Journal of Physics: Conference Series*, vol. 2025, no. 1, Sep. 2021, doi: 10.1088/1742-6596/2025/1/012076.

[17] M. Assiri, E. S. A. Elhameed, A. Kumar, and C. Singla, "Automated insect detection and classification using Pelican optimization algorithm with deep learning on internet of enabled agricultural sector," *SN Computer Science*, vol. 5, no. 5, May 2024, doi: 10.1007/s42979-024-02893-3.

[18] M. Aljebreen, H. A. Mengash, F. Kouki, and A. Motwakel, "Improved artificial ecosystem optimizer with deep-learning-based insect detection and classification for agricultural sector," *Sustainability*, vol. 15, no. 20, Oct. 2023, doi: 10.3390/su152014770.

[19] H. A and B. S. P, "Leveraging deep learning and farmland fertility algorithm for automated rice pest detection and classification model," *KSII Transactions on Internet and Information Systems*, vol. 18, no. 4, Apr. 2024, doi: 10.3837/tiis.2024.04.008.

[20] E. A. Al-Shahari, G. Aldehim, N. S. Almalki, M. Assiri, A. Sayed, and M. M. Alnfiai, "Innovative insect detection and classification for the agricultural sector using gannet optimization algorithm with deep learning," *IEEE Access*, vol. 12, pp. 108041–108051, 2024, doi: 10.1109/ACCESS.2024.3438308.
[21] T. Hiraguri, T. Kimura, K. Endo, T. Ohya, T. Takanashi, and H. Shimizu, "Shape classification technology of pollinated tomato flowers for robotic implementation," *Scientific Reports*, vol. 13, no. 1, Feb. 2023, doi: 10.1038/s41598-023-27971-z.
[22] A. I. S. Ferreira, N. F. F. da Silva, F. N. Mesquita, T. C. Rosa, V. H. Monzón, and J. N. M.-Neto, "Automatic acoustic recognition of pollinating bee species can be highly improved by deep learning models accompanied by pre-training and strong data augmentation," *Frontiers in Plant Science*, vol. 14, Apr. 2023, doi: 10.3389/fpls.2023.1081050.
[23] K. Bjerge, J. Alison, M. Dyrmann, C. E. Frigaard, H. M. R. Mann, and T. T. Høye, "Accurate detection and identification of insects from camera trap images with deep learning," *PLOS Sustainability and Transformation*, vol. 2, no. 3, Mar. 2023, doi: 10.1371/journal.pstr.0000051.
[24] D. Brandt *et al.*, "Low cost machine vision for insect classification," in *Intelligent Systems and Applications*, 2024, pp. 18–34. doi: 10.1007/978-3-031-47715-7_2.
[25] M. Tannous, C. Stefanini, and D. Romano, "A deep-learning-based detection approach for the identification of insect species of economic importance," *Insects*, vol. 14, no. 2, Jan. 2023, doi: 10.3390/insects14020148.

BIOGRAPHIES OF AUTHORS

Pooja Hadimane qualified for her Bachelors of Engineering (B.E.) degree and master's degree (M.Tech.) in CSE from STJIT Ranebennur, affiliated to VTU, Belagavi, in 2013 and 2015, respectively. She is pursuing her Ph.D. in Computer Science and Engineering at BIET, Davangere, and also worked as an assistant professor in the Department of Computer Science and Engineering at Bapuji Institute of Engineering and Technology, Davangere. She has attended many workshops and induction programs conducted by various institutes. Her areas of interest are machine learning, artificial intelligence, data mining, computer vision, and image processing. She can be contacted at email: poojah_12@rediffmail.com.

Dr. Ashoka Kukkuvada is a professor in the Department of Information Science and Engineering at Bapuji Institute of Engineering and Technology, Davangere. He is a qualified Bachelor of Engineering in Computer Science and Engineering and a master's degree in VLSI and Embedded Systems. Ph.D. from Visveswaraya Technological University in 2018. His areas of interest are data mining, computer vision, machine learning, artificial intelligence, and image processing. He can be contacted at email: ashokakukkuvada@gmail.com.