

PAPER • OPEN ACCESS

Patang Abhidhani - Convolution Neural Network based Butterfly Research Survey

To cite this article: R L Priya *et al* 2021 *J. Phys.: Conf. Ser.* **1916** 012010

View the [article online](#) for updates and enhancements.

You may also like

- [What do Bloch electrons in a magnetic field have to do with Apollonian packing of circles?](#)
Indubala I Satija
- [Ecological Green Space Planning for Butterflies Habitat in Residential Area](#)
H Maryam and Q Pramukanto
- [Diversity of Butterfly species \(Lepidoptera: Rhopalocera\) attracted to Carrion Trap at Harau Valley Nature Reserve](#)
E Bibas, H Herwina, Dahelmi et al.

ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Retraction

Retraction: Patang Abhidhani - Convolution Neural Network based Butterfly Research Survey (*J. Phys.: Conf. Ser.* **1916** 012010)

Published 23 February 2022

This article (and all articles in the proceedings volume relating to the same conference) has been retracted by IOP Publishing following an extensive investigation in line with the COPE guidelines. This investigation has uncovered evidence of systematic manipulation of the publication process and considerable citation manipulation.

IOP Publishing respectfully requests that readers consider all work within this volume potentially unreliable, as the volume has not been through a credible peer review process.

IOP Publishing regrets that our usual quality checks did not identify these issues before publication, and have since put additional measures in place to try to prevent these issues from reoccurring. IOP Publishing wishes to credit anonymous whistleblowers and the [Problematic Paper Screener](#) [1] for bringing some of the above issues to our attention, prompting us to investigate further.

[1] Cabanac G, Labbé C and Magazinov A 2021 arXiv:[2107.06751v1](#)

Retraction published: 23 February 2022



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Published under licence by IOP Publishing Ltd

Patang Abhidhani - Convolution Neural Network based Butterfly Research Survey

Priya R L¹, Yash Mate¹, Gaurav Tirodkar¹, Neelam Somai¹, Gayatri Patil¹ and Nikhil Bhopale²

¹Computer Department, Vivekanand Education Society's Education of Society, Chembur Mumbai.

²Green Works Trust

priya.rl@ves.ac.in, 2017.yash.mate@ves.ac.in, 2017.gaurav.tirodkar@ves.ac.in, 2017.neelam.somai@ves.ac.in, 2017.gayatri.patil@ves.ac.in, greenworkstrust2016@gmail.com

Abstract. There has been a great loss of biodiversity worldwide and efforts have to be taken in order to restore it. To restore biodiversity, it's crucial to know the cause for its decline and the role played by the insects in it. In India, butterflies are not given much importance in the conservation species. Since there are no names for butterflies in regional languages, the only access for comprehensive information for people's awareness is really nothing. This paper aims at providing details of research done on different butterfly species, their existence once in different parts of the world, and technologies employed for solving the problems like identification of different species, distribution, restoration, and rejuvenation of diversity.

Keywords: Butterfly, Convolution Neural Network, pollination, species, biodiversity, ecology, genus.

1. Introduction

India is a land of rich biodiversity, where insects, birds, and all animals have been flourishing for decades. Worldwide, humid tropical regions are known for high biodiversity and high endemism of flora and fauna which are under increasing threat of species extinction due to deforestation and habitat loss caused by the increasing demands of human populations. With the increase in the human population, the natural habitats of butterflies have considerably decreased. Due to the very high sensitivity of butterflies to the changes in the environment, their abundance, and advanced taxonomy, butterflies can be said to be the ideal indicators of habitat disturbance [1]. It is found that even a few changes in the environment can lead to perturbations in the population of butterfly species [1].

Butterflies are very graceful insects that provide economic as well as ecological benefits to society. They are very valuable as pollinators when they go around from one plant to another while gathering nectar. They are very good indicators of natural factors too [2].



Butterflies are an indispensable part of the ecosystem where they carry the pollen grains along with them when they migrate in search of nectar. This helps fruits, vegetables, and flowers to produce new seeds. The majority of plants require pollinators especially bees and butterflies for reproduction. Pollination produces genetically modified plant species that are more resistant to changing climatic conditions and in turn give rise to fitter butterflies in future generations. Butterflies are at the lowest level in the food chain. They are a food source of birds, spiders, lizards, rats, and other animals of the food chain. If butterfly populations diminish, the impact will be seen higher up in the food chain and can affect the entire ecosystem of nature.

Butterflies are commonly found in huge numbers on the planet and improve the flora of the surrounding through pollination. This builds up the need for research in this field and has been increasing exponentially. Consequently, there has been a remarkable decline in the diversity and population of these spectacular insects. The survey aims at the detailed study of the various species of butterflies and their impact on biodiversity. It focuses on the integrated system for data regarding the various existing species in various regions making all the data required for research available at a single source. Analysis of the data available will help in creating awareness of the importance of conservation and would help the researchers to find solutions for the same problem efficiently. The verified data resource will help in maintaining the integrity of the system. The distribution of butterfly species will be personalized according to the geographical location of the users and will be based on various sectors of occupations.

2. Literature Survey:

Butterflies are one of the most well-known groups of insects. It is a wonderful sight to nature lovers for its beauty and amazing colors. Butterflies are considered to be very good pollinators and are very sensitive to temperature, humidity, rainfall, solar radiation, air temperature, wind speed, etc. However, butterfly fauna is not always reported from many parts of the Indian subcontinent.[1]

India is a house for nearly 1400 different species of butterflies, out of which 330 can be located in the Western Ghats alone, and of which 37 are endemic. Skipper butterflies are quite common in the northern and northeastern districts of India like Kashmir, Simla, Manali where they were also spotted in huge numbers in Thirthahalli Panchayat town in Karnataka.[2]

North-Eastern region alone is the home for 137 different species of Skipper genus including but not limited to purple Lancer *Salanoemia fuscicornis*, Red-vein Lancer *Pyroneura asana burmana* Evans [3]

There is relatively more research work on the classification of butterfly specimen photos rather than ecological photos. In the paper by Dongjin Xin, the authors have presented their work on classification based on ecological photos that present a new classification that combines the squeeze-and-excitation (SE) module, dilated residual network and spatial attention (SA) module [4].

In this paper [5], the author uses a pre-trained model of CNN architecture called GoogleNet for butterfly species identification. The species used for research were Grey Pansy, Chocolate Grass Yellow, Black Veined Tiger and Plain Lacewing and the accuracy obtained was 97.5% overall on one hundred and twenty images of four types of butterflies.

Another research [6] uses Faster R-CNN to build automatic butterfly detection and identification of species in various ecological environments, and the average classification accuracy obtained is up to 70.4%.

The paper assesses the ability of classification of butterfly species by using texture features of different butterfly images. The author uses two texture descriptors called local binary patterns (LBP) and grey-level co-occurrence matrix (GLCM) for comparison and 19 different species of the pieridae butterfly family, comprising 190 butterfly images were used for the same. The accuracies were 96.45% and 98.25% with LBP and GLCM respectively [7].

Chorge has presented a study of survey butterflies of Maharashtra National Park (MNP) in Mumbai, which would provide baseline data, for planning an effective management study of insect diversity for the development of Maharashtra Nature Park. This park has more than 14,000 plants, 40 species of butterflies, and 84 species of birds.[8-10]

The main objective of the research is to identify the species of butterflies in the Himalayan Ranges, right from Jammu and Kashmir to Uttarakhand that is threatened. Observational data from 2013 to 2016 which encompasses ecological attributes like local status, distribution, and habitat utilization are taken into consideration. To highlight the gravity of the situation on the loss of biodiversity and degradation of habitats, butterfly species from Northern Himalayan regions and those not included in the Wild Life Protection Act 1972 were studied. [11]

The paper describes the significant aspects of the written contributions of the Seventh International Symposium, '2010 and beyond for Lepidoptera'. Some authors have investigated the impact of micro-climate, habitat quality, and topographical variation on the large-scale existence of butterflies and moths. The paper also illustrates measures for preserving habitat quality and ways to monitor the population trends. [12]

Factors related to the successful conservation of butterflies are studied by comprehensively conducting tests by varying the butterfly populations and habitats. The short-term fluctuations in the butterfly population may be attributed to variations in climate changes whereas the density of non-migratory species is mainly influenced by the availability of optimum food-plants as preferred by its larvae. Therefore, larval habitat quality is the prime factor of successful conservation. [13]

The paper describes a Convolutional Neural Network-based approach for the identification of butterfly species. Transfer Learning on VGG-16, VGG-19, and ResNet architectures have been utilized to obtain the highest accuracy of 84.8%. The dataset collected from a butterfly field in Turkey contained about 17,000 images of 10 different species. [14]

The research highlights the composition and diversity of 138 butterfly species found in six forest ranges of Nagarhole National Park (NNP) located in Karnataka from 2014 to 2015. Five butterfly families namely Lycaenidae, Nymphalidae, Pieridae, Hesperidae, and Papilionidae were studied. The composition was measured using the Shannon Diversity Index and Fischer Alpha value. Nymphalidae, among the six, had the highest composition of species (47). The research also provides insights into habitat restoration by using metrics like Simpson and Shannon indices to measure the evenness between the forest ranges in NNP. [15]

A research was conducted in the Netherlands to estimate the overall long-term changes in the butterfly species of that region. The modified list length method was used to estimate the distribution of Dutch butterfly species from 1890 to 2017, which then was used to calculate the MSI. Also, MSI's for three

different habitats: grassland, woodland, and heathland were also calculated, and an estimated decline of species distributions (> 80%) was observed with the help of the logistic regression method. [16]

A research study was conducted in the regions of Amazon to understand the impact of logging for timber production on the butterfly species in that region. Total of 5608 butterflies from 160 species across 60 sites, 40 of which were logged. The mean number of fruits feeding butterflies over 48 hours was found to be 40% higher in logged sites than the unlogged sites, and the richness found was 30% higher. The composition of species changed drastically between unlogged and particularly logged forest, mostly at higher logging intensities, thus highlighting the significance of logging activities while predicting the species distribution and richness. [17]

With the accelerating global extinction, it becomes important to establish reserves for biodiversity protection. The effectiveness of military training areas in Pennsylvania, USA was studied for conservation by evaluating the population of a very rare butterflies, the eastern regal fritillary (*Speyeria idalia idalia*). Mixed effect logistic regression and zero-inflated Poisson distribution for the count of species were used. The population trajectories exhibited a non-linear change in trajectory overtime. Trajectories since 1998 indicate three periods of increase (1998–2001, 2007–2009, 2011–2014) and two periods of decline (2001–2007, 2009–2011), where the period of 2011–2014 was observed as a recovery period, with the increasing trend beginning to level by 2014. [18]

The Mediterranean region of Andorra, Menorca (north-east Spain) and Catalonia was studied wherein the results of the seven species characteristics on the growth rate approximation were tested using linear regressions. Of the 82 regularly recorded species, 15 species exhibited a positive rate (22.7%) whereas 5 were stable (7.6%), and 42 exhibited a negative rate (69.7%). Habitat specialization and voltinism have the greatest influence, whilst other characteristics have a marginal or zero effect. [19] Table 1 shows Comparison with Other Existing Systems

Table 1. Comparison with Other Existing Systems (Data collected from selective surveys are discussed here).

Authors & Articles	Year	Research Objective	Data Collection	ML / AI methods	Validation Method or Evaluation measures	Pros	Cons
[4]	2020	Classification of butterfly species	Chinese butterfly dataset- 467 images and the Leeds butterfly dataset - 832 images, Total-1299 butterfly images	Squeeze-and-excitation (SE) and Spatial attention (SA) modules	Accuracy and loss of training dataset and validation for each epoch is done	Usage of ecological butterflies' images	1. Less number of image of butterflies in nearly all categories 2. Some butterflies have only a single image with a background completely different from

							others. Thus, these different aspects can cause target detection errors
[5]	2019	Study of butterfly species identification	120 Images of 4 species of butterflies- Chocolate Grass Yellow, Black Veined Tiger, Plain Lacewing and Grey Pansy	Google-Net	Overall accuracy for testing - 97.5%	Classification accuracy was recorded successfully as 97.5%.	Only 120 images were used for training the CNN model.
[6]	2019	Recognition of butterfly	Total of 5,695 butterfly photos, wherein there are 1,425 ecological photos, 111 different species, containing 17 species of butterflies.	Faster R-CNN	IOU, map, Average classification accuracy- 70.4%	With the use of Faster RCNN, the speed of recognition was increased.	Dataset expansion had a lot of scopes, like left-right rotation, upside-down, cooling, noise adding
[7]	2014	Automatic detection of different butterfly species using texture features	190 different butterfly images having 19 species of the Pieridae butterfly family.	Feature descriptors like local binary patterns (LBP) and grey-level co-occurrence matrix (GLCM)	identification accuracy of 98.25% and 96.45% for GLCM and LBP feature descriptors respectively	The classification accuracy of 98.25% for 70-30% training-test partition.	Images chosen were of the same background and only 190 images were used.
[14]	2020	Automatic Identification for	44,659 images of 104 different butterfly	CNN-VGG16, VGG19 ResNet	Training and testing accuracies have been	The ResNet training accuracy is very good at	Images here have a few problems such as the shooting

Field butterflies	species	obtained for all three CNN models.	84.8%	angle, position of butterflies, butterfly distance, occlusion, and complexity of background
----------------------	---------	---	-------	--

3. Proposed System

A web/mobile application will be developed with functionalities like image classification and expert verification of butterfly species. There can also be miscellaneous functionalities like research done on butterflies, their existence history, and benefits. A full-fledged application will be developed after testing it on different parameters and for different regions of India.

3.1 Block Diagram

Figure 1 represents the conceptual diagram of the proposed model. A custom-made dataset has been created using various books on butterflies, social media, and surveys. The data is preprocessed to remove outliers and the veracity of data is maintained to train the system with higher efficiency and obtain better results. The features most important for the decline of butterfly species have been identified and an image classification model will be developed which will be later integrated with an application.

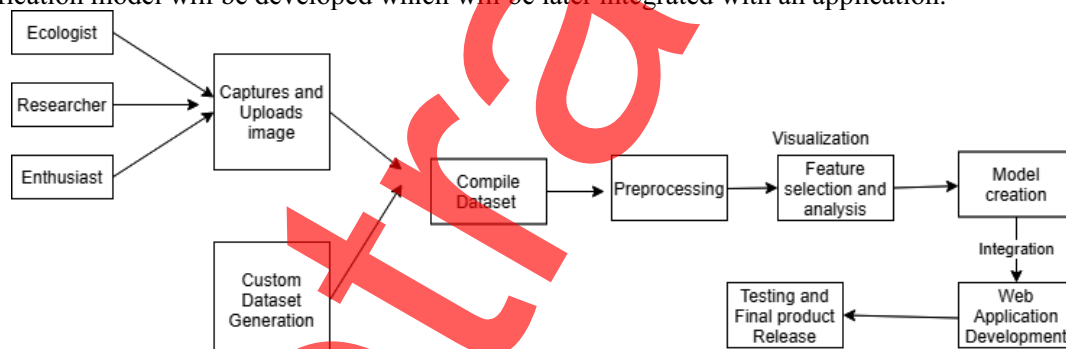


Figure 1. Block Diagram

3.1.1 Dataset Creation. The biggest lacuna in current systems is the availability of data. There isn't any focus given to Skipper Butterflies hence, there aren't any sources available easily of authentic data. Data collection will be a big and perhaps the toughest task of all. To make sure the data collected was authentic a variety of research papers, books and journals have to be referred to. Data Features used important in the case of Butterflies are:

- Genus,
- Species,
- Subspecies,
- Male and Female count,
- Location
- Year of Identification

3.1.2 Preprocessing. Taxonomical and population-based analytical methods are employed to visualize the distribution of the declining population of butterflies and their fragile habitats. This visualization would ensure a better, in-depth insight into the regions that need to be given special attention.

3.1.3 Feature Selection Analysis and Model creation. Deep Learning Methods for classification and segmentation of the butterfly from its surrounding are proposed. Neural Networks are efficient in providing accurate results owing to their complex architectures. A particular form of Neural Network called Convolutional Neural Network which is exquisite in carrying out tasks related to Visual data like Images and Videos would be designed. Leveraging the method of Transfer Learning with fine-tuning with Drop-Out and Batch Normalization layers whilst manipulating the hyper-parameters would lead to optimum results. State-of-the-art CNN architectures like ResNets, VGG, DenseNets would be fine-tuned and some custom additional layers would be added to the network. The performance of the network would be evaluated using the AUC-ROC curve. Accordingly, the learning-rate, optimizer, and loss function would be modified to ensure legitimate and optimized results.

For tabular data that deals with a population-based study, forecasting the population of the future species to a specific species residing in a particular habitat would be carried out with Time Series analysis. Machine learning-based forecasting methods like ARIMAX, SARIMAX with exponential smoothing to understand the trends and seasonality are proposed. Regression models that are resistant to outliers and deviations like Ensembling methods and Gradient Boosting are going to be investigated. Along with that, Deep Learning based solutions offered by Long Short-Term Memory (LSTMs) would also be evaluated using the Mean Squared Error metric.

Finally, the factors influencing the decline in the species can be evaluated using Feature Selection and Tree-Based approaches that highlight the feature importance of every attribute. A Hybrid Approach for feature selection by the amalgamation of traditional approaches like Filter, Wrapper, and Embedded Feature Selection methods are proposed to generalize the results obtained from the analysis to a wide range of populations. Other statistical methods to determine the correlation between the input attributes and the target variable are formulated using Pearson's Correlation, Z-Score, and Chi-Square Tests. After identification of the factors, the ecological authorities can take measures to ameliorate the degrading butterfly population.

3.1.4 Web application development. For the ease of accessibility, a minimalistic yet cutting-edge User Experience is proposed by integrating the results obtained from the analysis of population and visual data. The figure below illustrates the stages of the development pipeline.

All this information will be readily available on the website so that users across the globe will be able to access this information at the same time in a legible format easily in one single format. The attractive UI will enhance the user experience.

3.1.5 Testing and Deployment. From an avid butterfly connoisseur to an erudite research scientist working on ecology; this website would help them in navigating to even the remotest locations and do a detailed analysis of the ecology of butterflies residing there. Endemic and Migratory species play a historical role in maintaining the ecological balance of the biosphere. Merely hovering over the map would display the counts of the species with their corresponding exhaustive description.

Images of Butterflies in natural settings are ubiquitous. However, a manual visual understanding of some exotic species of butterflies yields nebulous results. In such cases, it becomes categorical to design and develop a system for the detection and recognition of rare species.

On top of this as more and more users start using this platform the model will start working with even better accuracy and also with the help of user reviews and suggestions the functionality of the website will be improved.

4. Conclusion

The study of various research papers and articles helped in the proper understanding of the various species of butterflies, the various factors which affect them, and the technological aspects and assistance taken for the conservation of those species. The parameters like the local population distribution, climatic changes, habitat quality, have a high impact on the population distribution and growth of butterflies. These insights will be highly helpful for farmers, butterfly researchers, ecologists, and government authorities. The focus is to provide a full-fledged website for butterfly species in India. This project can be extended for all Asian butterfly species. Further, a comparative study can be developed for all declining or rare species of butterflies all over Asia.

References

- [1] Dey, Pritam&Payra, Arajush&Mondal andKrishnendu (2017). A study on butterfly diversity in Singur, West Bengal, India, *e-planet*,**15**, 73-77.
- [2] ENJeevan, KLNaik, HMAshashree and HASayeswara, Butterfly Diversity and Status in Mandagadde of Shivamogga, Karnataka, India, *International Journal of Applied Biology and Pharmaceutical Technology*, **4** (4), ISSN: 0976-4550.
- [3] Gogoi and Monsoon. (2013), Notes on some skipper butterflies (Lepidoptera: Hesperidae) from Panbari Forest and its adjoining areas, Kaziranga-KarbiAnglong, upper Assam, India,*Journal of Threatened Taxa*. 5. 4759-4768. 10.11609/JoTT.o3340.4759-68.
- [4] Xin, D, Chen, YW, Li and J, Fine-Grained Butterfly Classification in Ecological Images Using Squeeze-And-Excitation and Spatial Attention Modules, *Appl. Sci.* 2020, **10**, 1681.
- [5] Arzar, Nur&Sabri, Baity&Johari, Nur&Amilah, Anis&Noordin, Mohd& Ibrahim andShafaf. (2019), Butterfly Species Identification Using Convolutional Neural Network (CNN),*10.1109/I2CACIS.2019.8825031*, 221-224.
- [6] R Zhao, C Li, S Ye, and X Fang, "Butterfly Recognition based on Faster R-CNN", *Journal of Physics*, **1176**(3)
- [7] Kaya, Yilmaz&Kayci, Lokman&Tekin, Ramazan&Ertugrul and Omer. (2014). Evaluation of texture features for automatic detecting butterfly species using an extreme learning machine. *Journal of Experimental & Theoretical Artificial Intelligence*. 26, 10.1080/0952813X.2013.861875, pp267-281.
- [8] Chorge, Sachin&Rajbhar, Anil &Prashant and Gokarankar. (2008). Butterfly diversity of Maharashtra Nature Park.
- [9] Mehra, Deepika&Kirti, Jagbir&Sidhu and Avtar. (2017). Biodiversity and conservation status of Butterflies in Western Himalaya, India: An appraisal. *Journal of Entomology and Zoology Studies*. **5**,pp 1171-1177.
- [10] A Catalogue of the Hesperidae from Europe, Asia, and Australia in the British Museum(Natural History) By Evans, W H (William Harry), 1876-1956.British Museum (Natural History). Department of Entomology. [Lepidoptera].
- [11] Mehra, Deepika&Kirti, Jagbir&Sidhu and Avtar. (2017). Biodiversity and conservation status of Butterflies in Western Himalaya, India: An appraisal. *Journal of Entomology and Zoology Studies*. **5**. Pp1171-1177.

- [12] Dover, J & Bourn, Nigel & Lewis, Owen & Shreeve and Tim. (2015). The ecology and conservation of butterflies and moths. *Journal of Insect Conservation*. 19. 10.1007/s10841-015-9770-9, pp183-184.
- [13] Thomas, JA, Simcox, DJ & Hovestadt, T, Evidence-based conservation of butterflies. *J Insect Conserv* doi.org/10.1007/s10841-010-9341-z **15**, 241–258 (2011).
- [14] Ayad Saad, Almryad and Hakan Kutucu, Automatic identification for field butterflies by convolutional neural networks, *Engineering Science and Technology, an International Journal*, **23**(1), February 2020, pp 189-195
- [15] H. Anandakumar and K. Umamaheswari, A bio-inspired swarm intelligence technique for social aware cognitive radio handovers, *Computers & Electrical Engineering*, vol. **71**, pp. 925–937, Oct. 2018. doi:10.1016/j.compeleceng.2017.09.016
- [16] R. Arulmurugan and H. Anandakumar, Early Detection of Lung Cancer Using Wavelet Feature Descriptor and Feed Forward Back Propagation Neural Networks Classifier, *Lecture Notes in Computational Vision and Biomechanics*, pp. 103–110, 2018. doi:10.1007/978-3-319-71767-8_9
- [17] Gabriela Montejó-Kovacevich, Matthew GHethcoat, Felix KSLim, Charles J Marsh, Dayan Bonfantti, Carlos APeres and David P Edwards, Impacts of selective logging management on butterflies in the Amazon, *Biological Conservation*, **225**, September 2018, pp 1-9
- [18] Konstantina Zografou, Mark T. Swartz, Virginia PTilden, Erika NMckinney, Julie AEckenrode and Brent JSewall, Severe decline and partial recovery of a rare butterfly on an active military training area, *Biological Conservation*, **216**, December 2017, Pages 43-50
- [19] Yolanda Melero, Constantí Stefanescu and Joan Pino, General declines in Mediterranean butterflies over the last two decades are modulated by species traits, *Biological Conservation* **201** (2016) 336–342
- [20] Richard AHenderson, Jed Meunier and Nathan SHoloubek, Disentangling effects of fire, habitat, and climate on an endangered prairie-specialist butterfly, *Biological Conservation* **218**, February 2018, pp 41-48