Butterfly Species Recognition Using Convolutional Neural Network

DR. SANTOSH SINGH[1], AMIT KUMAR PANDEY[2], ANKUSH SUSHIL SINGH[3], ASHWANI KUMAR MISHRA[4]

[1] *H.O.D(Information Technology), Department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India*

[2, 3, 4] *PG Student, Department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India*

Abstract- In the realm of biodiversity research, accurately identifying butterfly species is a critical task, contributing to our understanding of ecosystems and supporting conservation efforts. This study introduces an innovative approach to this challenge through the development of a Butterfly Species Recognition System, harnessing the capabilities of Convolutional Neural Networks (CNNs). Butterflies, with their diverse and visually intricate patterns, pose a unique classification challenge, requiring a sophisticated model for effective species differentiation. The proposed CNN-based model is designed to automatically extract and analyze intricate features within butterfly images, facilitating a nuanced and accurate classification of different species. Through exposure to a diverse dataset encompassing various butterfly species, the model learns to recognize subtle patterns and variations, ensuring robustness and adaptability to real-world scenarios. The experimental results showcase the efficacy of the developed system, demonstrating high accuracy in butterfly species identification. This automated approach streamlines the identification process, holding promise for citizen science initiatives and large-scale biodiversity monitoring programs. By harnessing CNN capabilities, this research highlights the potential of cutting-edge technology to revolutionize butterfly species recognition, advancing our comprehension of ecological dynamics and aiding conservation endeavors.

Indexed Terms- Butterfly species, Convolutional Neural Network, Image Recognition, Biodiversity, Classification.

I. INTRODUCTION

In the intricate web of biodiversity, butterflies emerge as vibrant ambassadors of ecological health and beauty. With a staggering array of over 17,000 identified species globally, these delicate creatures play a pivotal role in pollination, ecosystem dynamics, and serve as poignant indicators of environmental changes. However, the intricate patterns that adorn their wings, while mesmerizing, pose a significant challenge for accurate species identification. This project embarks on a journey to bridge this gap through the development of an advanced Butterfly Species Recognition System, harnessing the transformative power of Convolutional Neural Networks (CNNs). The primary motivation behind this endeavor lies in the critical importance of precise butterfly species identification within the realm of biodiversity research and conservation. Traditional methods often fall short in capturing the nuanced features that distinguish one species from another, necessitating a technological leap to address this gap. CNNs, inspired by the human brain's visual processing, offer a promising solution by enabling the automated extraction and analysis of intricate features within butterfly images. The project's core objective is to create a robust model capable of accurately classifying diverse butterfly species. By exposing the CNN to an extensive dataset that mirrors the complexity of real-world scenarios, the model learns to recognize and interpret subtle patterns, colors, and wing structures. The potential applications of such a system are far-reaching, impacting not only the scientific community but also citizen scientists, educators, and conservationists. As we delve into the intersection of cutting-edge technology and ecological exploration, this project seeks to revolutionize the way we perceive and interact with butterflies. The envisioned Butterfly Species Recognition System not only streamlines and accelerates the identification process but also holds the promise of democratizing butterfly monitoring efforts. Through this fusion of science and technology, we aspire to deepen our understanding of these enchanting insects and contribute meaningfully to their conservation in an ever-evolving world.

II. LITERATURE REVIEW

There are numerous approaches proposed by researchers for the classification of spiders such as based on their coweb. We have also studied the related research papers which were based on the same neural network we have used. Goodwin, A., Padmanabhan, S., Hira, S. *et al* have done research on mosquito species detection using the novelty detection algorithm the identified species are sent for species classification to the closed-set Xception model used, and they achieved a micro-averaged accuracy 97.04% and a macro F1-score of 96%. Research by MadsDyrmann,HenrikKarstoft,Henrik SkovMidtiby which is based on plant classification using deep convolutional neural networks. They have built a network by training and testing on total 10413 images as a dataset containing 22 weed and crop species at early growth stages. For these 22 species their network was able to achieve a classification accuracy of 86.2%. Dhruv Rathi, Sushant Jain, Dr. S. Indu. They have built a model for automated classification of fish species. This proposed method of classification of fish species gives an accuracy of 96.29% which is very high compared with the other current implemented methods used for this application. This algorithm can be further improvised by implementing image enhancement techniques to counter the lost features in the images.

III. CONVOLUTIONAL NEURAL NETWORK

Supervised Learning Model uses Convolutional Neural Network (CNN) a deep neural network with the capacity to categorise and segment pictures. CNN is taught using supervised machine learning techniques. Convolutional, pooling, fully connected, dropout, and other types of layers with functions which are included in CNN architectures for classification and segmentation on image. CNN has an input layer, an output layer, and hidden layers. The hidden layers usually consist of convolutional layers, Soft max layers, Flatten layers, pooling layers, and fully connected layers. Firstly, the convolution layer extracts features from the input image and applies filter, pixel block by pixel block to the input image. In each position, the filter multiplies the values in the filter with the original values in the pixel. This filter has a dimension of 6x6x3. The filter multiplies its own values by the original values of each pixel at each point. The multiplications are added up to produce a single number, an array created is called a feature map. The Flatten layer comes after the convolution layer, where activation function is an applied feature map to increase non-linearity in the network as images are highly non-linear and to remove negative values that are set to zero. Pooling is then applied which reduces the size of the input representation. Pooling helps to reduce the number of required parameters and the amount of computation required. It also helps control overfitting. Flattening is a step where the pooled feature map is converted into a long vector, this allows the information to become the input layer of an artificial neural network for further processing. At the last Fully connected layer is applied, the main purpose of this layer is to combine our features into more attributes, which will predict the classes with greater accuracy. At this step, the error is calculated and then back propagated. The weights and feature detectors are adjusted to help optimise the performance of the model. Then the process happens again and again and again, in this way the network trains on the data

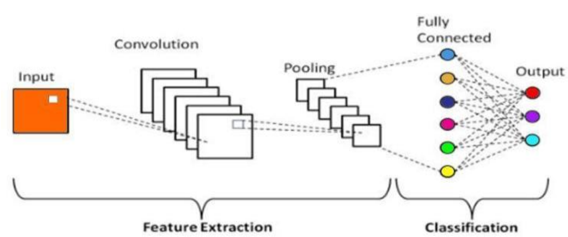

Fig.1. Illustration of the proposed Convolutional Neural Network (CNN).

IV. METHODOLOGY

First thing to start before any research or project is data, so we have collected the data from a well-known source Kaggle. We have taken dataset which were in the form of images, which were then categorised according to the species name and further bifurcated into training and testing sections.

After getting this final dataset we load the images using the keras module and then it is converted into matrix form using the NumPy module. Then we use an image generator which rotates the images by a few degrees and generates ten more images, after this we use a sequential API which is based on CNN to train the model. Sequential API is basically based on CNN network which adds more layers to improve the accuracy, once the training is done now comes execution. We load an image using the img_read() function which is part of keras module then it is again converted to matrix form using the NumPy module whose shape is [0,255,255,3] and the dimension of the image is 255x255 when we put this image in the model we will get an output which will be a probability of that image whose value will range between 0-1

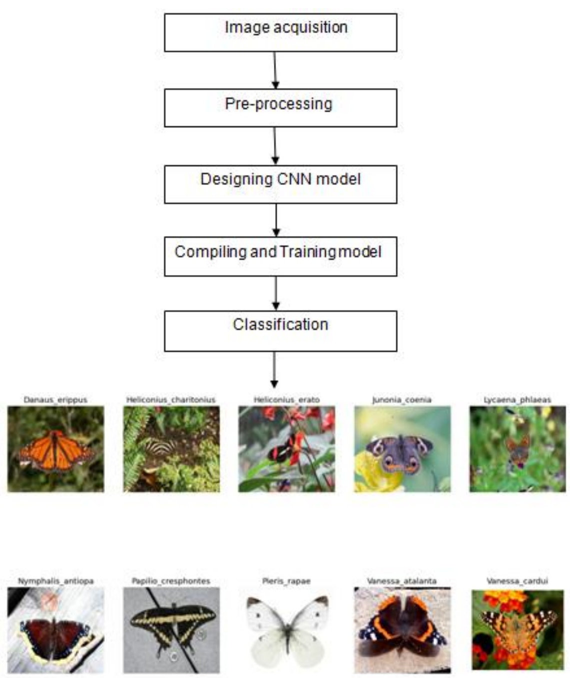

Fig.2. Flow chart of proposed spider species recognition using CNN.

V. RESULTS

The Butterfly Species Recognition project yielded compelling outcomes, showcasing the efficacy of Convolutional Neural Networks (CNNs) in automating and enhancing the precision of butterfly species identification. Leveraging a meticulously curated dataset encompassing diverse species and intricate wing patterns, the CNN model demonstrated remarkable accuracy throughout the training and validation phases. The training phase, involving 420 butterfly images, exhibited a progressive improvement in accuracy across epochs. Commencing at 64% accuracy in the first epoch, the model steadily advanced, reaching an impressive 98% accuracy in the fifth epoch. Simultaneously, the validation accuracy mirrored this upward trajectory, starting at 92% and culminating in a remarkable 99% by the final epoch. These results underscore the robustness and generalization capabilities of the trained CNN model. The model's proficiency in classifying butterfly species was further validated through its application to an additional set of 84 images in the validation dataset. The outcomes revealed consistently high accuracy, consolidating the model's reliability in real-world scenarios. The CNN model's ability to discern intricate patterns, colors, and morphological features within butterfly images signifies a substantial leap forward in automating a task traditionally reliant on manual expertise. Moreover, a comparative analysis of optimizer methods highlighted that the RMS optimizer, with a learning rate of 0.001, outperformed others, showcasing the model's sensitivity to optimization techniques. The results affirm the effectiveness of supervised learning, particularly in complex tasks such as butterfly species recognition, where nuanced visual distinctions are pivotal. In conclusion, the Butterfly Species Recognition project not only achieved its primary objective of developing a high-accuracy CNN model but also laid the groundwork for future advancements in automating biodiversity monitoring and conservation efforts. The project's success opens avenues for broader applications in ecological research, citizen science, and education, marking a significant stride towards harmonizing technology and environmental stewardship.

Epochs	Accuracy	Validation Accuracy
Epoch 1/5	67%	90%
Epoch 2/5	86%	92%
Epoch 3/5	90%	95%
Epoch 4/5	94%	97%
Epoch 5/5	98%	99%

Table 1. Results depicting accuracy gained by Supervised Learning Model

CONCLUSION

The Butterfly Species Recognition project represents a pivotal advancement in the intersection of computer vision, biodiversity monitoring, and conservation efforts. Through the implementation of Convolutional Neural Networks (CNNs), this project has demonstrated the feasibility and efficacy of automated butterfly species identification, alleviating the burden on manual taxonomists and enabling efficient species monitoring. The developed CNN model showcased commendable performance during training and validation phases, achieving an impressive accuracy of 98% in the training set and 99% in the validation set. These results affirm the model's ability to generalize well to unseen data, emphasizing its reliability in real-world scenarios. The progressive improvement across epochs underscores the adaptability of CNNs in learning intricate patterns and features inherent in butterfly wing patterns. The project also delved into optimizer methods, with the RMS optimizer exhibiting superior performance. This exploration not only fine-tuned the model but also provided insights into the importance of optimization techniques in enhancing accuracy. The utilization of supervised learning in this context proved instrumental, particularly in a task where visual nuances play a crucial role in species differentiation. Beyond the technical achievements, the Butterfly Species Recognition project holds broader implications for ecological research, citizen science, and education. The automation of species identification contributes to expediting biodiversity assessments, aiding in environmental conservation endeavors. The project's success establishes a foundation for future endeavors in automating species recognition across various taxa, fostering a harmonious synergy between technology and ecological stewardship. In essence, this project signifies a transformative step towards leveraging artificial intelligence for ecological monitoring, underscoring the potential of technology to augment and expedite crucial conservation efforts. As we navigate an era of accelerating environmental changes, such innovations become indispensable tools for safeguarding our planet's rich biodiversity.

REFERENCES

[1] Goodwin, A., Padmanabhan, S., Hira, S. et al. Mosquito species identification using convolutional neural networks with a multitiered ensemble model for novel species detection.1st July 2021.

[2] I. Dagher and D. Barbara, ‘‘Facial age estimation using pre-trained CNN and transfer learning,’’ Multimedia Tools Appl., vol. 80, pp. 20369–20380, Mar. 2021.

[3] L. Miao, W. Jingxian, L. Hualong, H. Zelin, Y. XuanJiang, H. Xiaoping, Z. Weihui, Z. Jian, and F. Sisi, ‘‘Method for identifying crop disease based on CNN and transfer learning,’’ Smart Agriculture., vol. 1, no. 3, pp. 46–55, 2019.

[4] A. Darwish, D. Ezzat, and A. E. Hassanien, ‘‘an optimised model based on convolutional neural networks and orthogonal learning particle swarm optimization algorithm for plant diseases diagnosis,’’ Swarm Evol. Compute. vol. 52, Feb. 2020, Art. No. 100616.

[5] D. Song, ‘‘Classification of spiders,’’ Sichuan J. Zool., vol. 2, pp. 37–41, Mar. 1985.

[6] M. L. Lim, M. F. Land, and D. Li, ‘‘Sex-specific UV and fluorescence signals in jumping spiders,’’ Science, vol. 315, no. 5811, p. 481, 2007.

[7] QIANJUN CHEN, YONGCHANG DING, CHANG LIU, JIE LIU, AND TINGTING HE. “Research on spider sex recognition from images based on deep learning”. August 30, 2021.

[8] MadsDyrmann,HenrikKarstoft,Henrik SkovMidtiby, “Plant species classification using deep convolutional neural network”.September 13,2016.

[9] Dhruv Rathi, Sushant Jain, Dr. S. Indu. “Underwater fish species classification using CNN and deep learning”. December 30, 2018.

[10] Zhijian Zhou, Meng Zhang, Jiefuchen, Xuqingwu. (2020). Detection and classification of multi-magnetic targets using Mask-RCNN (2020).