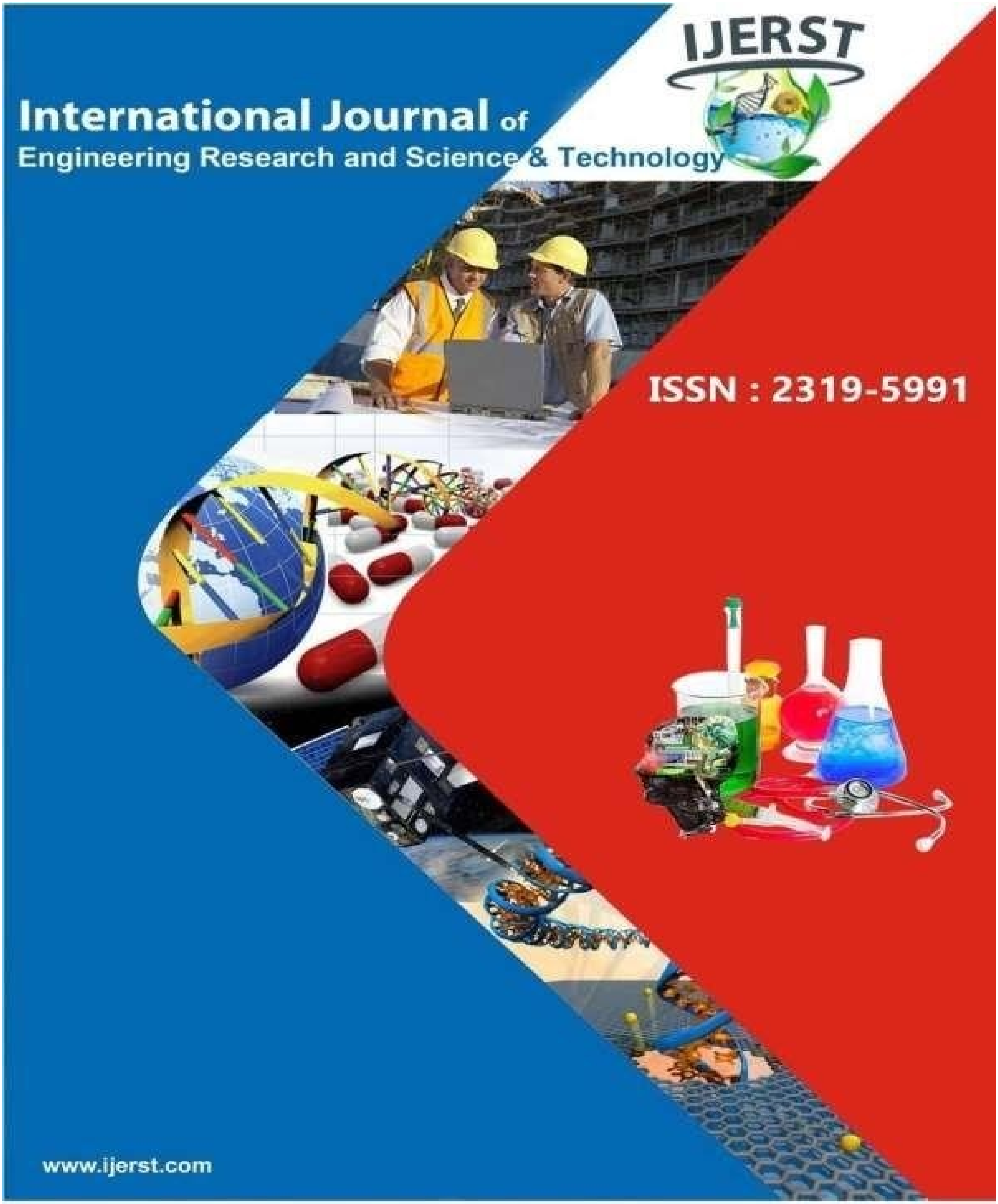


Email: editor@ijerst.com or editor.ijerst@gmail.com

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

Advanced Classification of Butterfly Speciesthrough Image Recognition

Yelgapuri Goutham Rao[1], Madoth Tharun [2], Panyam Guru Bhargav[3], Mr. K. Rajeshwar[4]
[1,2,3] UG Scholar, Dept. of ECE, St. Martin's Engineering College, Secunderabad, Telangana, India, 500100
[4]Assistant Professor, Dept. of ECE, St. Martin's Engineering College, Secunderabad, Telangana, India, 500100
gouthamrao119@gmail.com

Abstract:

Butterflies play a vital role in ecosystems as pollinators and biodiversity indicators, yet accurately identifying species is challenging due to their morphological diversity. Traditionally, butterfly classification has relied on manual identification by experts, which is time-consuming, requires extensive knowledge, and is prone to human error. These limitations make large-scale butterfly monitoring difficult, affecting biodiversity studies and conservation efforts. This research proposes an automated image recognition system for advanced classification of butterfly species, utilizing to identify species based on visual features. By training the model on a large dataset of butterfly images, the system learns to detect subtle differences in wing patterns, colors, and shapes, enhancing classification accuracy beyond traditional methods. Unlike manual identification, the proposed system operates quickly, enabling real-time recognition and classification. This need for an automated approach arises from the demand for scalable, accurate biodiversity monitoring, especially in regions with high species diversity. The proposed system offers significant improvements in speed, accuracy, and accessibility, making it a valuable tool for ecologists, researchers, and citizen scientists engaged in butterfly studies, biodiversity monitoring, and conservation planning.

Keywords: Butterfly Classification, Image Recognition, Bio-Diversity Monitoring, Automated Species Identification, Real-Time Classification, Machine Learning, Visual Features.

1. INTRODUCTION

Butterflies are crucial pollinators and serve as bio-indicators of environmental health, making their study essential for biodiversity conservation. India, home to over 1,800 butterfly species, has a rich diversity influenced by varied climates and ecosystems. However, accurately identifying butterfly species is a complex task due to their morphological similarities, seasonal variations, and mimicry. Traditional classification methods rely on manual observation by entomologists, a process that is not only time-consuming but also requires extensive expertise. According to studies, nearly 20-30% of butterfly species remain misidentified due to human error, affecting ecological research. With the advent of image recognition and deep learning techniques, automated classification of butterflies has emerged as a promising solution to overcome these challenges. Leveraging artificial intelligence (AI) in biodiversity research enhances species identification accuracy while facilitating large-scale monitoring. The development of an advanced image recognition system for butterfly classification enables real-time identification, significantly contributing to ecological studies, species conservation, and citizen science initiatives. Such a system is vital in regions like India, where biodiversity monitoring needs scalable and efficient technological interventions.

2. LITERATURE SURVEY

Butterflies are insects in the order Lepidoptera. The number of butterfly species in the world varies from 15,000 to 21,000 according to a recent estimate in [1]. Due to various species, high similarity and the characteristics of the distinction of butterflies is not evident, the identification and classification of butterflies have the problems of low accuracy and slow recognition. Furthermore, the number of taxonomists and trained technicians has decreased dramatically. The discrimination between butterfly species requires expertise and time, which is not always available, but after the development of software that identifies butterfly species by extracting features from images, the need for experts will reduce. There are two main problems in the existing butterfly species identification research based on computer vision techniques. First, collecting the butterfly dataset is difficult, identifying is time-consuming work for entomologists, and the number of butterflies included in the butterfly dataset is not comprehensive. Second, the butterfly pictures used for training are all pattern pictures with obvious morphological features, lacking the ecological pictures of butterflies in the nature. Furthermore, the differences between two pictures are obvious that makes the combination of research and production difficult and the recognition accuracy is low [2]. Therefore, it is of great importance to make research on the automatic identification of butterflies and improve its accuracy and efficiency. The development of tools for automating the identification of butterfly species has an important contribution to the literature.
Kaya et al. [3] proposed two different local binary patterns (LPB) descriptors for detecting special textures in images. The first one is based on the relations between the sequential neighbors of a center pixel with a specific distance and the other one is relies on determining the neighbors in the same orientation through the central pixel parameter. They tested their descriptors to identify butterfly species on lab-based images of 140 butterflies collected in Van city of Turkey. The highest accuracy they obtained for classifying butterflies by the artificial neural network is 95.71%.

Kaya and Kaycı [4] developed a model for butterfly identification based on color and texture features by the artificial neural network using the gray level co-occurrence matrix with different angles and distances. The accuracy of their method reached 92.85%.

Xie et al. [5] improved a learning model for the classification of insect images using advanced multiple-task sparse representation and multiple-kernel learning techniques. They tested the proposed model on 24 common pest species of field crops and compared it with some recent methods. Feng et al. [6] improved an insect species identification and retrieval system based on wing attributes of moth images. Their retrieval system is based on CBIR architecture which is not about having a final answer but just providing a match list to the user [7]. The final decision can be made by an expert. Another study for pest recognition was done by Faithpraise et al. using k-means clustering algorithm and correspondence filters [8].

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

Leow et al. [9] developed an automated identification system of copepod specimens by extracting morphological features and using ANN. They used seven copepod features of 240 sample images and estimated an overall accuracy of 93.13%. Silva et al. [10] aimed to investigate the best combination of feature selection techniques and classifiers to identify honeybee subspecies. They found the best pair as the combination of Naïve Bayes classifier and the Correlation-Based feature selector in their experimental results among seven combinations of feature selectors and classifiers.

Wang et al. [11] designed an identification system of insect images at the order level. They extracted seven features from insect images. However, the authors manually removed some attached to insects such as pins. Their method has been tested on 225 specimen images from nine orders and sub-orders using artificial neural network with an accuracy of 93%. Abeysinghe et al. [12] introduced a fully automated framework to identify snake species using convolutional Siamese networks. Although the current dataset of snakes is small, they achieved satisfactory results.

Surabhi et al. [13] proposed the use of Deep Convolutional Neural Networks (CNNs) a deep learning (DL) technique for accurate and efficient recognition and classification of butterfly species based on their visual features and patterns. This study is focused specifically to automate the classification of butterfly species like Tigers and Emigrants category, since they are more prevalent in the Indian region. This study has been considered a probabilistic data driven approach, as they have curated imaging dataset specific only to the Indian regions. This study used techniques like web scrapping to gather and curate imaging dataset. In this study application of CNN based method is analyzed.

Yasmin et al. studied the review of butterfly species classification to portray a summary of various proposed methods of butterfly species identification as well as classification strategies [14]. Various methods essentially image processing, computer vision, traditional machine learning, neural networks, deep learning, transfer learning, etc. have been employed until now to detect and classify butterflies. Rajeena et al. aimed to assist science students in correctly recognizing butterflies without harming the insects during their analysis [15]. This article discussed transfer-learning-based neural network models to identify butterfly species. The datasets are collected from the Kaggle website, which contains 10,035 images of 75 different species of butterflies. From the available dataset, 15 unusual species were selected, including various butterfly orientations, photography angles, butterfly lengths, occlusion, and backdrop complexity. When they analyzed the dataset, they found an imbalanced class distribution among the 15 identified classes, leading to overfitting. The proposed system performs data augmentation to prevent data scarcity and reduce overfitting. The augmented dataset is also used to improve the accuracy of the data models.

3. PROPOSED METHODOLOGY

Step 1: Butterfly Image Dataset

The dataset consists of butterfly images categorized into different species. Each image is labeled according to its respective butterfly class. The dataset is stored in a directory structure where each folder represents a species.

Step 2: Image Preprocessing

Images are read using OpenCV (imread), resized to a fixed dimension (resize), flattened to standardize pixel intensity values, and reshaped to match the input format of the neural network. The processed image data and corresponding labels are saved as .npy files for efficient loading.

Step 3: Image Augmentation

To enhance dataset diversity, augmentation techniques such as rotation, shear, and horizontal flipping are applied using ImageDataGenerator. This helps improve model generalization and prevent overfitting by artificially increasing the training set.

Step 4: Existing CNN with SGD Classifier

A Convolutional Neural Network (CNN) is trained using the Stochastic Gradient Descent (SGD) optimizer. The network includes multiple convolutional, pooling, and fully connected layers. SGD is used with a fixed learning rate to optimize weight updates during training.

Step 5: Existing CNN with ADAM Classifier

The same CNN architecture is trained using the Adam optimizer, which adapts learning rates dynamically for each parameter. Adam often converges faster than SGD and improves accuracy by handling sparse gradients more effectively.

Step 6: Proposed Adam & Valid Padding Classifier

A modified CNN model is introduced with "valid" padding in convolutional layers and the Adam optimizer. This model structure aims to enhance feature extraction and classification accuracy by preserving spatial information in the feature maps.

Step 7: Performance Comparison & Accuracy vs. Epoch Graph

The performance of all models is compared using accuracy, precision, recall, and F1-score. Confusion matrices and bar graphs are plotted for visual comparison. An Accuracy vs. Epoch graph illustrates model convergence over training epochs.

Step 8: Prediction on Test Images (Proposed Model)

The final trained model (Adam & Valid Padding) is used to classify unseen test images. The model predicts class labels for new butterfly images, demonstrating its generalization capability and real-world applicability as shown in Fig. 1

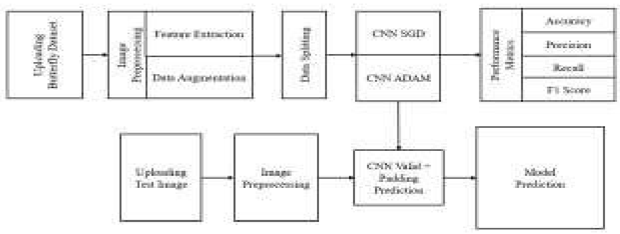


Fig. 1: Block Diagram

4. EXPERIMENTAL ANALYSIS

Fig. 2: Upload of Butterfly Species image Dataset and Its Analysis in the GUI Interface.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

Fig. 2 illustrates the process of uploading the Butterfly Species image dataset into the GUI interface for analysis. It shows how the dataset is loaded and provides a visual overview of the dataset's contents. Key statistics about the dataset, such as the number of images, distribution across different categories, and sample images from each category, are displayed. This setup facilitates the understanding and verification of the dataset before any further processing or analysis.

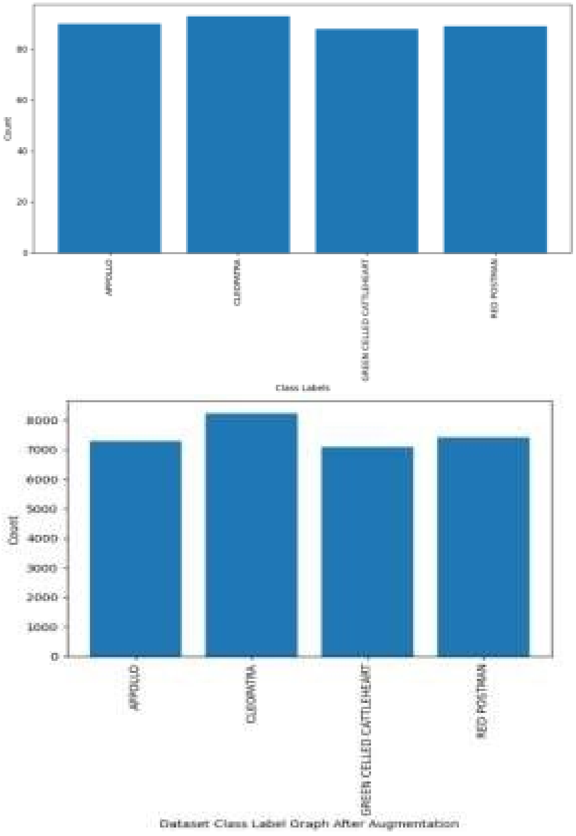


Fig. 3: Categories and After Augmentation Count Plot of the Dataset.

Fig. 3 presents the categorization of the Butterfly Species image dataset after data augmentation. The plot illustrates the number of images in each category before and after augmentation. It clearly shows how augmentation techniques have balanced the dataset, ensuring that each category has a sufficient number of images for training the model. The visual distribution highlights the success of augmentation in addressing class imbalances.

Fig. 4: Image Preprocessing and Image Augmentation in the GUI.

Fig. 4 showcases the GUI interface where image preprocessing and augmentation techniques are applied to the Butterfly Specimage dataset. It includes transformations like rotation, flipping, and zooming, which are used to augment the dataset. The interface allows users to apply these techniques in real-time, providing visual feedback of the augmented images and ensuring that the data is appropriately prepared for model training.

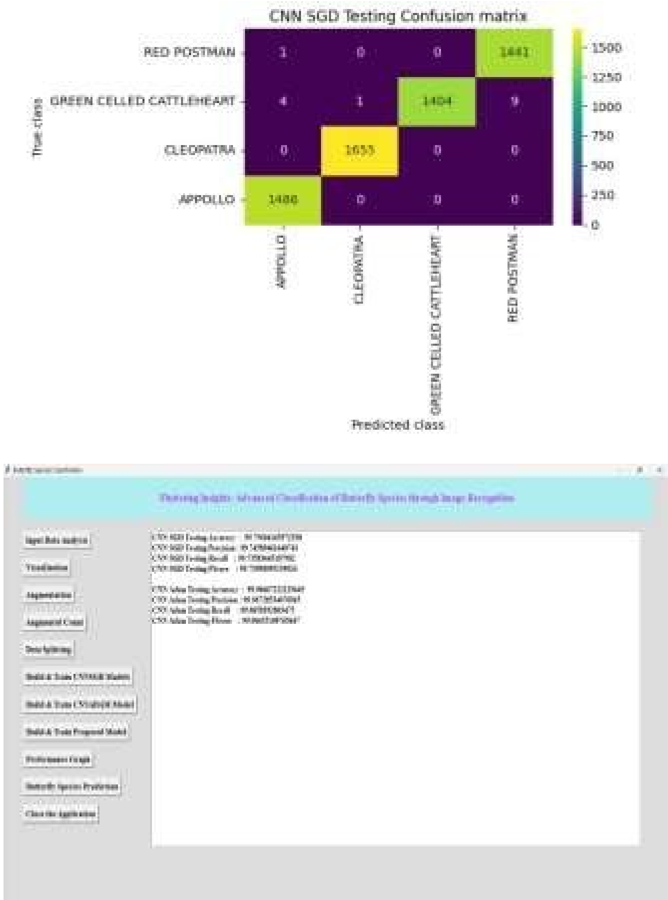
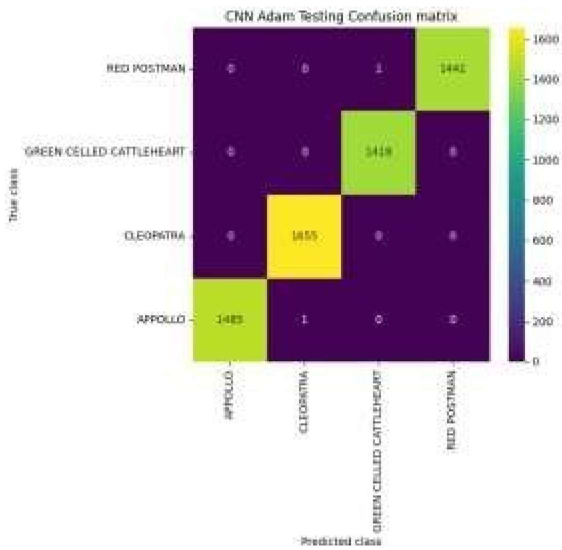


Fig. 5: The Performance Metrics and Confusion Matrix Plot of CNN with SGD Classifier Model.

Fig. 5 displays the performance metrics and confusion matrix plot of the Convolutional Neural Network (CNN) model trained with the Stochastic Gradient Descent (SGD) classifier. The confusion matrix further supports these metrics by visualizing the model's ability to correctly classify each category of the dataset, showing a near-perfect classification performance.



International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

Fig. 6: The Performance Metrics and Confusion Matrix Plot of CNN with Adam Classifier Model.

Fig. 6 highlights the performance metrics and confusion matrix of the CNN model trained with the Adam optimizer. The results reflect the exceptional performance The confusion matrix confirms the model's high accuracy, as it shows minimal misclassifications across the categories.

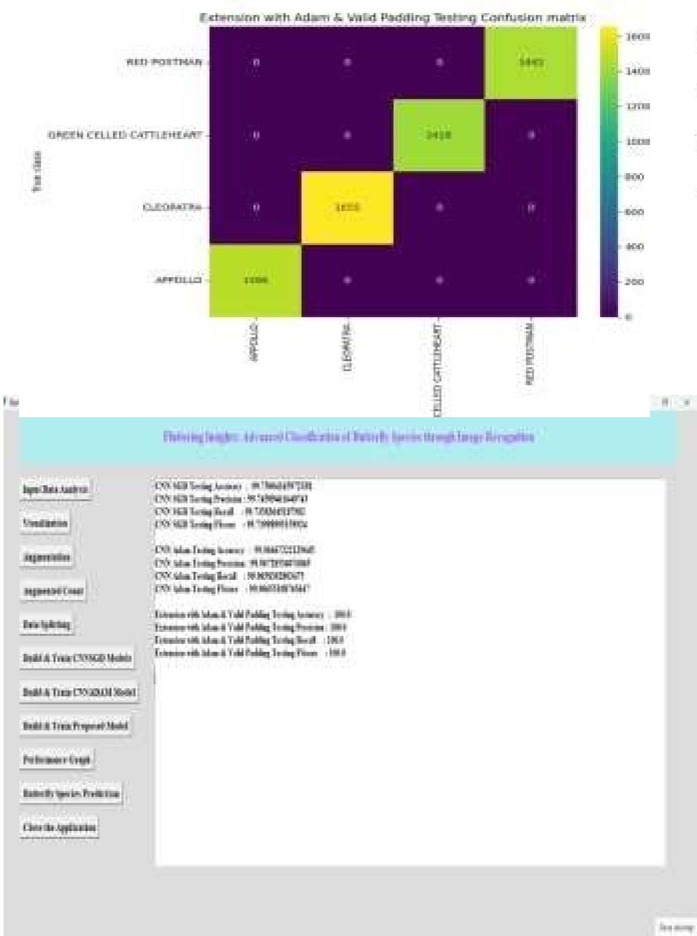


Fig.7: The Performance Metrics and Confusion Matrix Plot of CNN with Adam and Valid Padding Classifier Model.

Fig. 7 presents the performance metrics and confusion matrix of the CNN model trained with the Adam optimizer and valid padding. The confusion matrix, paired with the metrics, demonstrates that the model has achieved perfect classification on the test data, with no misclassifications in any of the categories.

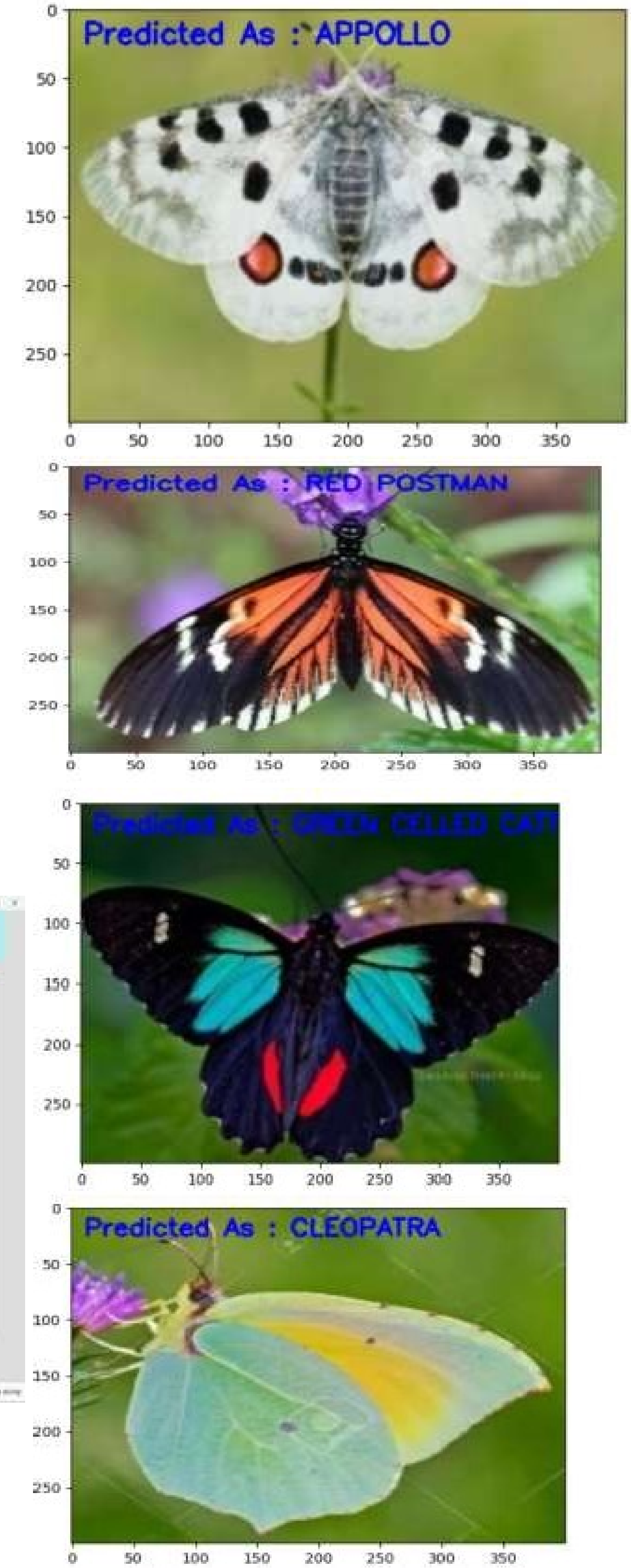


Fig. 8: The Model Prediction on the Test Data.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

Fig. 8 displays the predictions made by the model on the test dataset. The figure shows a series of predicted labels alongside the actual labels for the test images. The model successfully classifies each image, matching the predicted label with the correct category, further confirming its high performance.

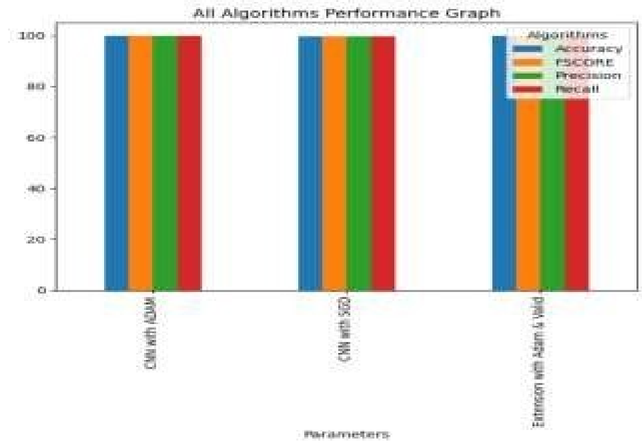
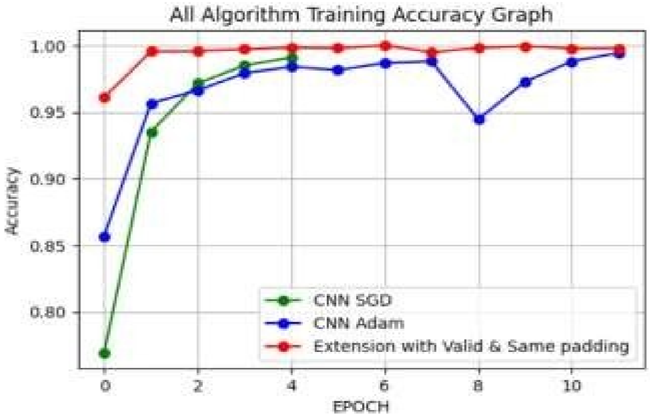


Fig. 9: The Performance Comparison Graph and Epoch vs. Loss



Graph of All Models.

Fig. 9 provides a comparison of the performance of all models used in this project. The performance comparison graph visually contrasts the testing accuracy, precision, recall, and F-score of each model (CNN with SGD, CNN with Adam, and CNN with Adam and Valid Padding). The Epoch vs. Loss graph tracks the loss values during training for each model, providing insights into the training progress and model convergence.

5. CONCLUSION

This Project successfully implements an image classification model to identify butterfly species using machine learning techniques. Image preprocessing and data augmentation improve model performance by enhancing feature extraction. A convolutional neural network (CNN) is trained with different optimization techniques to achieve high accuracy. The results demonstrate that the chosen approach is effective in distinguishing butterfly species based on visual patterns. The model provides reliable predictions, making it useful for applications in biodiversity monitoring, ecological studies, and species identification.

REFERENCES

[1] N.E. Stork, How many species of insects and other terrestrial arthropods are there on Earth?, Annu. Rev. Entomol. 63 (2018) 31–45, https://doi.org/10.1146/annurev-ento-020117-043348.

[2] M. Pinzari, M. Santonico, G. Pennazza, E. Martinelli, R. Capuano, R. Paolesse, et al., Chemically mediated species recognition in two sympatric Grayling butterflies: Hipparchia fagi and Hipparchia hermione (Lepidoptera: Nymphalidae, Satyrinae), PLoS One 13 (6) (2018), https://doi.org/10.1371/journal.pone.0199997.

[3] Y. Kaya, O.F. Ertug˘rul, R. Tekin, Two novel local binary pattern descriptors for texture analysis, Appl. Soft Comput. 34 (2015) 728–735, https://doi.org/10.1016/j.asoc.2015.06.009.

[4] Y. Kaya, L. Kaycı, Application of artificial neural network for automatic detection of butterfly species using color and texture features, Visual Comput. 30 (2014) 71–79, https://doi.org/10.1007/s00371-013-0782-8.

[5] C. Xie, J. Zhang, R. Li, J. Li, P. Hong, J. Xia, P. Chen, Automatic classification for field crop insects via multiple-task sparse representation and multiple-kernel learning, Comput. Electron. Agric. 119 (2015) 123–132, https://doi.org/10.1016/j.compag.2015.10.015 .

[6] L. Feng, B. Bhanu, J. Heraty, A software system for automated identification and retrieval of moth images based on wing attributes, Pattern Recogn. 51 (2016) 225–241, https://doi.org/10.1016/j.patcog.2015.09.012.

[7] M. Martineau, D. Conte, R. Raveaux, I. Arnault, D. Munier, G. Venturini, A survey on image-based insect classification, Pattern Recogn. 65 (2017) 273–284, https://doi.org/10.1016/j.patcog.2016.12.020.

[8] F. Faithpraise, P. Birch, R. Young, J. Obu, B. Faithpraise, C. Chatwin, Automatic plant pest detection and recognition using k-means clustering algorithm and correspondence filters, Int. J. Adv. Biotechnol. Res. 4 (2013) 189–199. http://sro.sussex.ac.uk/id/eprint/49042.

[9] L.K. Leow, L.-L. Chew, V.C. Chong, S.K. Dhillon, Automated identification of copepods using digital image processing and artificial neural network, BMC Bioinf. 16 (2015) S4, https://doi.org/10.1186/1471-2105-16-S18-S4 .

[10] F.L.d. Silva, M.L. Grassi Sella, T.M. Francoy, A.H.R. Costa, Evaluating classification and feature selection techniques for honeybee subspecies identification using wing images, Comput. Electron. Agric. 114 (2015) 6877, https://doi.org/10.1016/j.compag.2015.03.012.

[11] J. Wang, C. Lin, L. Ji, A. Liang, A new automatic identification system of insect images at the order level, Knowl.-Based Syst. 33 (2012) 102–110, https://doi.org/10.1016/j.knosys.2012.03.014 .

[12] C. Abeysinghe, A. Welivita, I. Perera, Snake image classification using Siamese networks, in: Proceedings of the 2019 3rd International Conference on Graphics and Signal Processing, 2019, pp. 8–12, https://doi.org/10.1145/3338472.3338476.

[13] T. Surabhi, B. Sachin and C. Advait, "Deep Convolutional Neural Networks for Automated Butterfly Species Recognition and Classification," *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, 2023, pp. 778-783, doi: 10.1109/ICIRCA57980.2023.10220696.

[14] R. Yasmin, A. Das, L. J. Rozario, Md. E. Islam, “Butterfly detection and classification techniques: A review”, Intelligent Systems with Applications, Volume 18, May 2023, 200214. doi: https://doi.org/10.1016/j.iswa.2023.200214 .