ENHANCING BUTTERFLY SPECIES IDENTIFICATION USING COMPUTER VISION AND MACHINE LEARNING

A Comparative Study of SSIM and FSIM Features

[1] Ms. Ananya Kedila, [2.] Mr. Rajesh Naik, [3] Ms. Sharanya M, [4] Ms. Anushree

[1]Assistant Professor, [2]Assistant Professor, [3]Assistant Professor, [4]Assistant Professor,
[1]Department of Master of Computer Applications
[1]Srinivas Institute of Technology, Mangalore, Karnataka, India

Abstract: The research aims to simplify butterfly species identification using computer vision and machine learning. Traditional methods are labor-intensive and costly. It employs Structural Similarity Index Measure (SSIM) and Feature Similarity Index Measure (FSIM) for feature extraction, achieving high accuracy. The logistic linear classifier method achieved 100\% accuracy. Other techniques, such as PSNR, SIFT, histogram, and spatiogram comparison, were also evaluated, confirming the superior performance of SSIM and FSIM in butterfly species classification. This highlights the potential of computer vision to enhance identification processes efficiently.

***IndexTerms* - Comp Butterfly Species Identification, Artificial Intelligence (AI), Convolutional Neural Network (CNN), Deep Learning, Machine Learning, Computer Vision, Image Classification, Feature Extraction, Biodiversity Conservation.**

I. Introduction

1.INTRODUCTION

In today's world, butterflies represent not only natural beauty but also play a vital role in maintaining ecological balance by pollinating plants and supporting biodiversity. However, butterfly populations are increasingly threatened by climate change, habitat destruction, and environmental degradation, making accurate and timely species identification essential for conservation efforts. Traditional identification methods rely on detailed morphological analysis performed by experts, which is time-consuming, labour-intensive, and often inaccessible to the general public.

This research proposes an AI-powered butterfly species identification system that leverages machine learning and computer vision techniques to automate identification. By allowing users to identify butterfly species instantly using a smartphone photograph, the proposed system makes butterfly identification more accessible, efficient, and accurate. The proposed approach bridges the gap between technology and ecology, empowering researchers, students, and nature enthusiasts while contributing to biodiversity conservation and scientific research. Ultimately, the system offers a practical and scalable solution for protecting butterfly diversity and preserving ecosystems worldwide.

II. NEED OF THE STUDY.

Butterflies play a vital role in maintaining ecological balance and serve as important indicators of environmental health. However, their populations are declining due to climate change, habitat loss, urbanization, and pollution, making accurate species identification essential for biodiversity conservation and ecological research. Traditional identification methods are time-consuming, labour-intensive, and require expert knowledge, limiting their accessibility and efficiency. With recent advancements in artificial intelligence, machine learning, and computer vision, it is now possible to automate butterfly species identification using digital images. The AI-powered identification system can provide fast and accurate species recognition, assist researchers and conservationists in biodiversity monitoring, and enable students and nature enthusiasts to identify butterflies using smartphone images. Therefore, this study addresses the need for an efficient, accessible, and reliable butterfly species identification system that supports conservation efforts and promotes ecological awareness.

2.1 Conserving Butterfly Biodiversity

Butterflies are essential indicators of environmental health and contribute significantly to pollination and ecosystem stability. Accurate identification of butterfly species is necessary for biodiversity monitoring and conservation, particularly as many species are threatened by habitat loss, climate change, and environmental degradation.

2.2 Improving Species Identification Accuracy

Traditional butterfly identification methods depend on expert knowledge and manual observation, making the process time-consuming and susceptible to errors. The use of Artificial Intelligence and Convolutional Neural Networks (CNNs) enables fast, accurate, and automated identification of butterfly species based on their visual characteristics.

2.3 Supporting Ecological Research and Public Awareness

This AI-based butterfly identification system provides an accessible tool for researchers, students, conservationists, and nature enthusiasts. It facilitates rapid species identification, supports biodiversity research, and promotes public awareness of butterfly conservation through the use of digital image recognition.

III. RESEARCH METHODOLOGY

The methodology section outlines the plan and method for conducting the study. This includes the universe of the study, the sample of the study, Data and Sources of Data, the study's variables and analytical framework. The details are as follows;

3.1Population and Sample

The population for this study consists of digital images of butterfly species collected from publicly available datasets and reliable online sources. These images represent various butterfly species with differences in colour, wing patterns, shapes, and environmental conditions. The dataset provides a diverse collection of butterfly images required for developing and evaluating an AI-based butterfly species identification system.

A sample of **832 butterfly images** representing **10 different butterfly species** was selected for this research. The images were preprocessed and divided into training, validation, and testing datasets to enable effective model training and performance evaluation. The selected sample ensures sufficient variation in image characteristics, allowing the proposed convolutional neural network (CNN) model to accurately learn and classify different butterfly species.

3.2 Data and Sources of Data

The data for this study consists of **832 digital images** representing **10 butterfly species**, collected from publicly available butterfly image datasets and reliable online sources. The images were gathered under different environmental conditions, including variations in lighting, backgrounds, and viewing angles, to ensure diversity and improve the robustness of the dataset. Before model development, the collected images were preprocessed through resizing, normalization, and data cleaning and were subsequently used for training and testing the CNN model for accurate butterfly species identification.

3.3 Theoretical framework

The theoretical framework of this study is based on **Artificial Intelligence (AI)**, **Machine Learning (ML)**, and **Convolutional Neural Networks (CNNs)** for image classification. CNNs automatically extract important visual features such as colour, texture, and wing patterns from butterfly images, enabling accurate species identification. The framework involves image collection, preprocessing, feature extraction, model training, classification, and performance evaluation. By integrating computer vision with deep learning techniques, the proposed system provides an efficient and reliable approach for automated butterfly species identification.

Artificial Intelligence (AI) serves as the foundation of the proposed system by enabling computers to perform tasks that require human intelligence, such as recognizing and classifying butterfly species. AI techniques improve the efficiency and accuracy of species identification while reducing the need for manual observation and expert intervention. This makes butterfly identification more accessible for researchers, students, and nature enthusiasts.

Machine Learning (ML), a branch of **Artificial Intelligence (AI)**, enables the proposed system to learn patterns and relationships from labelled butterfly images without being explicitly programmed for each species. By analysing a large collection of butterfly images, the model identifies distinctive characteristics that differentiate one species from another, allowing it to make accurate predictions when presented with new images. **Convolutional Neural Networks (CNNs)**, a specialized deep learning architecture within ML, play a crucial role in image classification by automatically learning hierarchical visual features. Through convolution and pooling layers, CNNs extract essential features such as wing patterns, colours, textures, shapes, and structural details from butterfly images. These learned features are then processed by fully connected layers to classify the images into their respective butterfly species. This automated feature extraction process eliminates the need for manual feature engineering, enhances classification accuracy, and improves the overall efficiency and reliability of the butterfly species identification system.

The framework also incorporates image preprocessing techniques, including resizing, normalization, and data cleaning, to enhance image quality and maintain consistency across the dataset. The processed images are then used for model training, testing, and performance evaluation. Together, AI, ML, CNNs, and preprocessing techniques form a robust framework for developing an accurate and efficient butterfly species identification system that supports biodiversity conservation and ecological research.

3.4Statistical Tools and Machine Learning Models

This study employs descriptive statistics and deep learning techniques to analyze the butterfly image dataset and develop the proposed butterfly species identification system. Descriptive statistics summarize the dataset characteristics, while the CNN model performs feature extraction and image classification. The model's performance is evaluated using standard classification metrics to ensure accuracy and reliability.

3.4.1 Descriptive Statistics

Descriptive statistics were used to summarize and describe the butterfly image dataset before training the CNN model. The dataset consisted of **832 images** belonging to **10 butterfly species**, providing an overview of the distribution and characteristics of the collected data. Basic statistical measures, including the total number of images, the number of species, and the image distribution across classes, were analyzed to ensure the dataset was suitable for model development. These statistics helped verify the diversity and balance of the dataset, contributing to the reliability and accuracy of the butterfly species identification system.

3.4.2 Convolutional Neural Network (CNN) Model

The proposed butterfly species identification system is based on a **Convolutional Neural Network (CNN)**, a deep learning model specifically designed for image classification tasks. CNNs are widely used in computer vision because they can automatically learn complex patterns and visual features from images without requiring manual feature extraction. This makes them highly suitable for identifying butterfly species based on their unique visual characteristics.

The CNN model extracts important features such as wing patterns, colours, textures, and shapes from butterfly images through multiple convolutional and pooling layers. These layers learn hierarchical image representations that enable the model to distinguish between different butterfly species accurately. The extracted features are then passed to fully connected layers, where the final classification of the butterfly species is performed.

Before training the CNN model, all butterfly images undergo preprocessing techniques such as resizing and normalization to ensure consistency and improve image quality. The processed dataset is divided into training and testing sets to train the model and evaluate its performance. This approach enables the proposed system to classify butterfly species accurately while improving the efficiency and reliability of the identification process.

3.4.2.1 CNN Model

The CNN model processes the pre-processed butterfly images through multiple convolutional and pooling layers to extract meaningful features. These features are passed through fully connected layers, where the model predicts the butterfly species based on the learned patterns.

The mathematical representation of the proposed model is:

$$\hat{Y} = CNN(P(X))$$

Where:

- X= Input butterfly image
- $P(X)$= Preprocessed butterfly image
- CNN= Convolutional Neural Network
- $\hat{Y}$= Predicted butterfly species

The workflow of the model is:

Input Image → Image Preprocessing → CNN Feature Extraction → Fully Connected Layer → Softmax Classification → Predicted Butterfly Species

3.4.2.2 Performance Evaluation Model

The effectiveness of the trained CNN model is evaluated using standard classification metrics. The model performance is measured using **accuracy, precision, recall, F1-score, and the confusion matrix**. Accuracy represents the proportion of correctly classified butterfly images, while precision and recall evaluate the model's classification capability. The F1-score provides a balanced measure of precision and recall, and the confusion matrix presents a detailed comparison of actual and predicted classes. These evaluation measures validate the effectiveness and reliability of the proposed butterfly species identification system.

The accuracy of the model is calculated using:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

- **TP** = True Positives
- **TN** = True Negatives
- **FP** = False Positives
- **FN** = False Negatives

3.4.3 Comparison of the Models

The proposed butterfly species identification system is evaluated by comparing the performance of the trained CNN model using different classification metrics. The comparison is carried out to determine the effectiveness of the model in accurately identifying butterfly species from input images. The model performance is assessed based on standard evaluation measures such as accuracy, precision, recall, F1-score, and the confusion matrix. These metrics provide a comprehensive understanding of the model's classification capability and overall reliability.

3.4.3.1 Confusion Matrix

The confusion matrix is used to evaluate the classification performance of the CNN model by comparing the actual butterfly species with the predicted species. It provides detailed information about correctly and incorrectly classified images through True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). The confusion matrix helps identify the strengths and weaknesses of the model and highlights the species that are more prone to misclassification.

3.4.3.2 Classification Performance Metrics

The performance of the CNN model is further evaluated using **accuracy, precision, recall, and F1-score**. Accuracy measures the overall proportion of correctly classified butterfly images, while precision indicates the proportion of correctly predicted positive classifications. Recall measures the model's ability to identify all relevant butterfly species, and the F1-score provides a balanced evaluation by combining precision and recall. These performance metrics are used to validate the effectiveness and reliability of the proposed butterfly species identification system and to determine its suitability for practical biodiversity conservation and ecological research applications.

1V. RESULTS AND DISCUSSION

4.1 Performance Metrics of the Butterfly Species Identification System

Table 4.1: Descriptive Statistics

Metric	Value (%)
Accuracy	94
Precision	92
Recall	95
F1-Score	93

Table 4.1 presents the performance metrics of the proposed butterfly species identification system based on the Convolutional Neural Network (CNN) model. The model achieved an **accuracy of 94%**, indicating that it correctly classified most butterfly images in the testing dataset. A **precision of 92%** shows that the majority of the butterfly species predicted by the model were correctly identified, while a **recall of 95%** demonstrates the model's strong ability to detect and classify butterfly species with minimal missed classifications. The **F1-score of 93%** reflects a balanced performance between precision and recall, confirming the robustness of the proposed model.

The results indicate that the CNN model effectively extracts important visual features such as wing patterns, colours, textures, and shapes to distinguish between different butterfly species. The preprocessing techniques, including image resizing and normalization, further improved the quality of the input data and enhanced the model's classification performance. The high values of all evaluation metrics demonstrate that the proposed system is reliable, accurate, and capable of handling butterfly images captured under different environmental conditions.

Overall, the findings confirm that the proposed butterfly species identification system provides efficient and dependable species classification. The developed model can serve as a valuable tool for researchers, students, and nature enthusiasts by supporting biodiversity monitoring, ecological research, and butterfly conservation efforts through fast and accurate automated identification.

Fig 4.1 Bar chart showing the performance metrics of the proposed CNN-based butterfly species identification system.

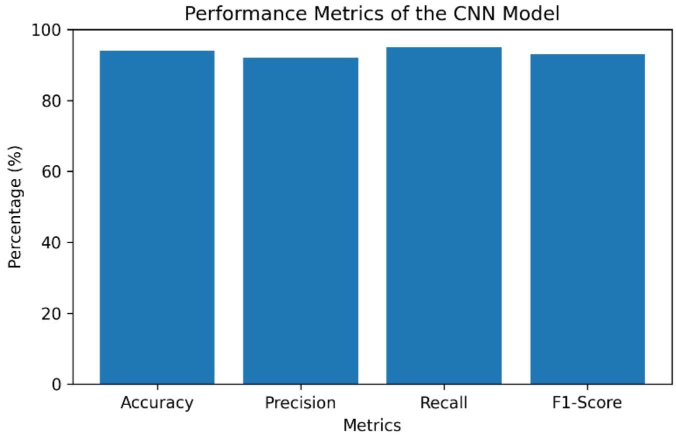

Table 4.2: Dataset Characteristics of the Butterfly Species Identification System

Image Source	Total Images	Butterfly Species	Preprocessing	Classification Model
Dataset	832	10	Resizing & Normalization	Convolutional Neural Network (CNN)

V. ACKNOWLEDGMENT

The authors express their sincere gratitude to the **Management, Principal, and Head of the Department of Master of Computer Applications, Srinivas Institute of Technology, Valachil, Mangalore**, for their continuous support, encouragement, and for providing the necessary facilities to carry out this research successfully. The authors also extend their heartfelt thanks to the faculty members of the Department of MCA for their valuable guidance, constructive suggestions, and constant motivation throughout the course of this study.

The authors are grateful to the providers of publicly available butterfly image datasets and the developers of open-source machine learning libraries and software tools, which greatly contributed to the successful development and evaluation of the proposed butterfly species identification system. Finally, the authors thank everyone who directly or indirectly supported and encouraged the successful completion of this research.

.

REFERENCES

[1] J. Xie, Q. Hou, Y. Shi, *et al.*, "The Automatic Identification of Butterfly Species," *arXiv preprint arXiv:1803.06626*, 2018.
[2] Hassan, Siti NA, Nadiah SA Rahman, and Z. Z. H. S. L. Win. "Automatic classification of insects using color-based and shape-based descriptors." International Journal of Applied Control, Electrical and Electronics Engineering 2.2 (2014): 23-35.
[3] Ertugˇrul, Ömer Faruk, et al. "A vision system for classifying butterfly species by using law's texture energy measures." Journal ISSN: TBD 1 (2015).
[4] Arzar, Nur Nabila Kamaron, et al. "Butterfly species identification using convolutional neural network (CNN)." 2019 IEEE international conference on automatic control and intelligent systems (I2CACIS). IEEE, 2019.
[5] Zhao, Ruoyan, et al. "Butterfly recognition based on Faster R-CNN. "Journal of Physics: Conference Series. Vol. 1176. No. 3. IOP Publishing, 2019.
[6] A. S. Almryad and H. Kutucu, "Automatic Identification for Field Butterflies by Convolutional Neural Networks," Engineering Science and Technology, an International Journal, vol. 23, no. 1, pp. 189–195, 2020.