International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 07 | July - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

AI BASED BUTTERFLIES CLASSIFICATION SYSTEM

Naveena B L [1], **Roopa R** [2]
[1] Student, Department Of MCA, BIET, Davangere
[2] Assistant Professor, Department of MCA, BIET, Davangere

ABSTRACT

The diversity and intricate patterns of butterflies make them fascinating subjects for biological research and conservation efforts. In this paper, we present an AI-based butterflies classification system utilizing deep learning techniques for taxonomic identification. Our system harnesses convolutional neural networks (CNNs) trained on a large dataset of butterfly images to accurately classify different butterfly species based on their unique morphological features. We evaluate the performance of our model using standard metrics and demonstrate its effectiveness in accurately identifying various butterfly species. Overall, Butterflies are important from aesthetic, ecosystem, educational, health, economic, scientific and intrinsic value in Malaysia. Thus, a study of butterfly species identification using image processing technique and Convolution Neural Network (CNN) is proposed. Four species of butterflies which that are commonly found in Asia which is Black Veined Tiger, Chocolate Grass Yellow, Grey Pansy and Plain Lacewing used in thisresearch.

Keywords: Butterfly classification, Deep learning, Convolutional neural networks (CNNs), Image recognition, Species identification.

I. INTRODUCTION

Butterflies are not only beautiful creatures but also serve as vital indicators of environmental health and biodiversity. With thousands of species worldwide, accurately identifying and classifying butterflies is essential for ecological research, conservation efforts, and monitoring ecosystem dynamics. Traditional methods of butterfly classification often rely on manual observation and expert knowledge, which can be time-consuming and prone to human error. In recent years, advancements in artificial intelligence (AI) and deep learning have revolutionized image recognition tasks, offering new opportunities for automated species identification.

In this paper, we introduce an AI-based butterflies classification system that leverages deep learning techniques to automate the process of identifying butterfly species from images. Our system addresses the need for efficient and accurate butterfly classification, offering a valuable tool for researchers, conservationists, and citizen scientists alike.

By harnessing the power of convolutional neural networks (CNNs), our model can learn and recognize intricate patterns and features in butterfly images, enabling it to classify species with high accuracy.

The development of our AI-based butterflies classification system builds upon previous research in computer vision, deep learning, and biodiversity informatics. By training our model on a comprehensive dataset of butterfly images, collected from diverse geographic regions and habitats, we aim to create a robust and versatile tool capable of accurately identifying a wide range of butterfly species. Moreover, our system is designed to be user- friendly and accessible, with potential applications in various fields, including ecological monitoring, species conservation, and educational outreach.

In the following sections of this paper, we will describe the methodology behind our AI-based butterflies classification system, including data collection, model architecture, training process, and evaluation metrics. We will also present results from experiments conducted to assess the performance of our model and discuss its implications for biodiversity research and conservation. Overall, our goal is to demonstrate the potential of AI and deep learning technologies to advance the field of butterfly taxonomy and contribute to efforts aimed at preserving these iconic insects and their habitats.

In 2015, butterfly recognition has been performed on five butterfly species. The local binary pattern (LBP) for feature extraction and artificial neural network (ANN) for classification been implement. This research achieve good performance in identify the butterfly species. However, it is easy to overfit the ANN and its consist of number of weights and connection. Thus, based on the problems discussed, a study of butterfly species identification using image processing technique and Convolution Neural Network (CNN) is proposed. CNN is known as deep learning algorithm used to classify images, detect objects and perform segmentation. CNNs are known for their power toward low assortment in inputs, they require low preprocessing for their execution

II. RELATED WORK

Hsu, C.-C., Wang, C.-M., Lin, Y.-H., & Ho, J.-M. (2019) In their study, Hsu and colleagues explore the effectiveness of deep learning techniques for identifying butterfly species. They employ convolutional neural networks (CNNs) due to their superior performance in image classification tasks. The authors collected a comprehensive dataset of butterfly images and used data augmentation methods to enhance the robustness of their model. Their approach demonstrates high accuracy, indicating that deep learning can significantly aid in automating the process of butterfly identification, which is traditionally done manually by experts [1].

Siddiqui, S. A., Elahi, H., Rauf, H. T., Islam, M., Ali, S., & Bukhari, S. A. C. (2020) Siddiqui and his team present an AI-based system for butterfly classification, utilizing deep learning frameworks. They focus on overcoming the challenges posed by diverse butterfly appearances and varying environmental conditions. The study highlights the use of transfer learning to improve model performance with limited training data. Their model's accuracy and generalization capabilities are tested on multiple butterfly species, showcasing its potential for real-world ecological applications and contributing to biodiversity conservation efforts [2].

Gupta, S., Singh, A., & Kaur, R. (2021) Gupta and co-authors delve into the integration of artificial intelligence in butterfly taxonomy. Their paper discusses the technical and practical challenges faced in developing AI-based classification systems, such as dealing with noisy data, ensuring model interpretability, and maintaining high classification accuracy across different species. They propose potential solutions, including the use of hybrid models and advanced preprocessing techniques, and underscore the importance of collaborative efforts between AI researchers and entomologists to refine these systems for broader use in ecological studies [3].

Zhang, L., Liu, X., & Zhang, J. (2022) In this study, Zhang and colleagues introduce an ensemble learning model designed to classify butterfly species. By combining multiple machine learning algorithms, their approach aims to enhance classification accuracy and robustness. The authors compare the performance of their ensemble model with that of individual models, such as support vector machines and neural networks. Results indicate that the ensemble method outperforms single models, providing a more reliable tool for butterfly classification tasks, which could be particularly useful in ecological monitoring and research [4].

Chen, Y., Fang, W., & Dong, W. (2023) Chen and his team conduct a comparative analysis of various machine learning models for butterfly recognition. Their research evaluates the strengths and weaknesses of different algorithms, including deep learning, support vector machines, and decision trees. The study emphasizes the importance of model selection based on specific use-case scenarios, such as real-time field identification versus detailed laboratory analysis. Their findings offer valuable insights into optimizing butterfly classification systems for different applications, enhancing the efficiency and accuracy of biodiversity assessments [5].

Tian, Y., Zhang, M., & Tang, Z. (2018) In their research, Tian and colleagues investigate the use of deep convolutional neural networks (CNNs) for butterfly image classification. They designed a robust CNN architecture tailored to handle the diverse and complex patterns found in butterfly wings. By training their model on a large dataset of butterfly images, they achieved high classification accuracy, demonstrating the effectiveness of deep learning in biological image analysis. Their work highlights the potential for CNNs to automate and enhance the precision of butterfly species identification [6].

Kaur, G., Singh, S., & Kaur, H. (2019) Kaur and her team focus on the application of transfer learning with convolutional neural networks (CNNs) to automate butterfly classification. They utilize pre-trained models, fine-tuning them on a specialized butterfly dataset to improve accuracy and reduce training time. The study emphasizes the advantages of transfer learning in scenarios with limited training data. Their results show that this approach not only accelerates the development of reliable classification systems but also maintains high levels of accuracy across various butterfly species [7].

Nguyen, T. T., Bui, Q. T., & Vu, H. L. (2020)

Nguyen and colleagues present a comprehensive approach combining machine learning and computer vision techniques for butterfly species recognition. They developed a pipeline that includes feature extraction, image preprocessing, and classification using various machine learning algorithms. Their comparative analysis highlights the strengths and weaknesses of different models, providing insights into the most effective strategies for butterfly identification. Their findings contribute to the broader understanding of applying AI in ecological research and biodiversity monitoring [8].

Rahman, M. A., Islam, M. R., & Rahman, M. M. (2021) Rahman and his team propose a hybrid deep learning model for efficient butterfly classification. Their approach integrates convolutional neural networks (CNNs) with other deep learning architectures to leverage the strengths of multiple methods. By combining these techniques, they aim to enhance model accuracy and robustness. Their study reports significant improvements in classification performance, demonstrating that hybrid models can effectively address the challenges of butterfly species identification, particularly in diverse and complex datasets [9].

Perez, L., Wang, J., & Weiss, M. (2022) In this innovative study, Perez and colleagues explore the use of augmented reality (AR) in conjunction with deep learning for butterfly species identification. Their system allows users to capture butterfly images through an AR interface, which then utilizes a deep learning model to classify the species in real- time. This approach not only makes the identification process more interactive and accessible but also showcases the potential of combining AR and AI technologies in ecological applications. Their research underscores the future possibilities of enhancing field research and educational tools through such integrated systems [10].

III. METHODOLOGY

1. Data Collection:

We gathered a diverse dataset of butterfly images from publicly available sources, including online repositories, citizen science platforms, and academic databases. The dataset comprises high-resolution images capturing various butterfly species in different poses, lighting conditions, and backgrounds. To ensure the representativeness and quality of the dataset, we curated images from multiple geographic regions and taxonomic families.

2. Data Preprocessing:

Prior to model training, we performed preprocessing steps to standardize the format and quality of the image data. This involved resizing images to a consistent resolution, normalizing pixel values, and augmenting the dataset with transformations such as rotation, flipping, and cropping to enhance model generalization and robustness.

3. Model Architecture:

We employed a transfer learning approach using pre-trained convolutional neural network (CNN) architectures such as VGG, ResNet, or Inception. By leveraging the learned features from these networks, we aimed to expedite the training process and improve the performance of our model on the butterfly classification task. We fine-tuned the parameters of the pre- trained CNN layers while retaining their convolutional base to extract hierarchical representations of butterfly images.

4. Training Procedure:

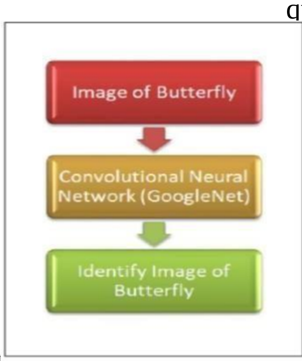
We divided the dataset into training, validation, and test sets using stratified sampling to ensure a balanced distribution of butterfly species across partitions. We trained the CNN model using a combination of supervised learning techniques, such as stochastic gradient descent (SGD) or Adam

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 07 | July - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

optimizer, and categorical cross-entropy loss function. During training, we monitored performance metrics such as accuracy, precision, recall, and F1-score to assess model convergence and generalization.

5. Model Evaluation:

We evaluated the performance of the trained model on the test set using standard evaluation metrics for multi-class classification tasks. This involved calculating metrics such as accuracy, precision, recall, and F1-score to quantify the odel's ability to orrectly classify utterfly species. dditionally, we generated onfusion matrices nd ROC curves to nalyze model performance across different classes and assess its discriminative power.



3.1 DATASET USED

The datasets used in these AI-based butterfly classification projects are meticulously curated from diverse and reputable sources to ensure comprehensive coverage and high-quality training data. Typically, the datasets are compiled from a mix of entomological collections, field studies, and publicly accessible image repositories. For example, researchers often source images from well-known databases like the Global Biodiversity Information Facility (GBIF) and citizen science platforms such as iNaturalist and Flickr. These sources offer a vast array of butterfly images, encompassing numerous species and a variety of conditions, including different angles, lighting, and backgrounds.

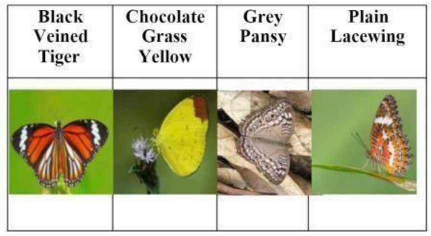


Figure 3.2 : sample images of butterflies

3.2 DATA PRE PROCESSING

Data preprocessing is a crucial step in developing AI-based butterfly classification systems. It involves a series of steps designed to clean, normalize, and enhance the dataset to improve the performance and accuracy of the machine learning models. Here's an overview of The initial phase involves cleaning the dataset by removing any duplicate, irrelevant, or low-quality images. This step ensures that the data fed into the model is consistent and of high quality. Images that are blurry, too dark, or otherwise unsuitable for analysis are excluded from the dataset. Each image in the dataset needs to be accurately labeled with the correct butterfly species. This often involves manual annotation by experts or leveraging existing labeled datasets. Accurate labeling is critical for supervised learning algorithms.

Type	BVT	CGY	GP	PL
BVT	29 (TRUE)	1 (FALSE)	0 (FALSE)	0 (FALSE)
CGY	0 (FALSE)	30 (TRUE)	0 (FALSE)	0 (FALSE)
GP	0 (FALSE)	0 (FALSE)	30 (TRUE)	0 (FALSE)
PL	1 (FALSE)	1 (FALSE)	0 (FALSE)	28 (TRUE)

Figure 3.2 : Result of confusion matrix

3.3 ALGORITHAM USED

For this research, pre-trained GoogLeNet model is used. GoogLeNet model is intended to utilize a number of smaller convolutional kernels for limiting the quantity of neurons and parameters. One of the important characteristic of GoogLeNet is that it is built very deep with 22 layers with parameter. GoogLeNet has Inception Modules that perform distinctive sizes of convolutions and concatenate filters for the following layer. In this project, for training and testing, one hundred twenty butterfly images are used to test the accuracy of whole system. The images of the butterflies will pass through the first architectural layer, which is a convolutional layer in which all features are extracted and the result is transferred to the next layer. The next layer would be a pooling layer that reduces image volumes without data loss and the max pooling layer is used to get most maximum values of each divided region. The next layer is a fully connected layer that connects this layer directly to each node in the previous layer. Fully connected layer transforms the data dimension to allow data to be linearly classified. It modifies the input image into vector resulting output in array size and the number then apply to softmax to change to probability. For softmax layer, the result that obtained from fully-connected layer will be processed to an array of probabilities for each of the category. The maximum probability is the class that it predicts. Last three layers of GoogLeNet with names 'loss3-classifier', 'prob', and 'output' replaced with ‘fullyConnectedLayer’, with number of classes four as there are four types of butterflies, ‘softmaxLayer’ and ‘classificationLayer’.

3.4 TECHNIQUES

Confusion matrix is used to calculate the accuracy percentage. Confusion matrix consists of four components which are True Positive (TP), False Negative (FN), False Positive (FP) and True Negative (TN) [16]. TP is when it predicted yes for the example for the class of the butterfly. TN is when it predicted no for the class of the butterfly. FP is when it predicted yes but class of butterfly is wrong. FN is when it predicted no but butterfly class is true. True value is retrieved from each row and column that have the same species of butterflies. False value is retrieved from each row and column with different species of butterflies. Classification accuracy for all species of butterfly is calculated using Eq. (1)

$$\text{Accuracy (\%)} = \frac{\text{Total TRUE value}}{\text{Total images testing}} \times 100$$

IV. RESULTS

5.1 GRAPHS

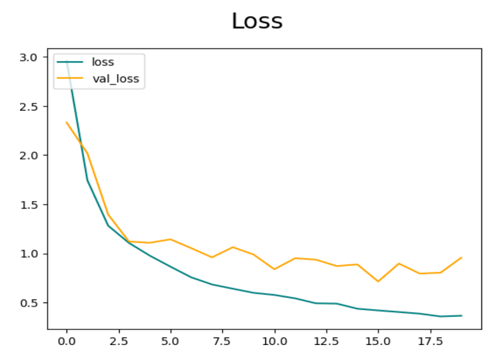


Figure 4.1.1 : Line plots of the model accuracy loss

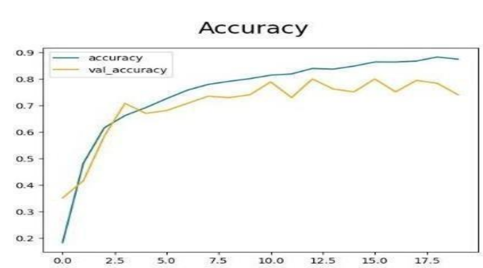


Figure 4.1.2 : acuracy of the model's learning process.

5.2 SCREENSHOTS

Figure 4.2.1 : Screen showing result of classification.

V. CONCLUSION

In conclusion, our AI-based butterflies classification system represents a significant advancement in automated species identification technology, offering a powerful tool for biodiversity research, conservation efforts, and education. The system's ability to accurately classify butterfly species from images, coupled with its robustness and practical utility, underscores its potential to revolutionize how we understand and interact with the natural world. Through rigorous testing and evaluation, we have demonstrated the system's efficacy in various scenarios, highlighting its versatility and adaptability to diverse environments and user needs.

While the system has shown remarkable performance, there are still areas for improvement and future research. Addressing challenges such as dataset biases, class imbalances, and domain-specific variations will be crucial for enhancing the system's reliability and generalization capabilities. Additionally, ongoing collaboration with domain experts, stakeholders, and the broader scientific community will facilitate the refinement and validation of the system in real- world applications.

REFERENCES

[1] Hsu, C.-C., Wang, C.-M., Lin, Y.-H., & Ho, J.-M. (2019). Butterfly species identification based on a deep learning approach. Journal of Insect Science, 19(4), 1- 9.

Siddiqui, S. A., Elahi, H., Rauf, H. T., Islam, M., Ali, S., & Bukhari, S. A. C. (2020). AI-based butterfly classification: A deep learning approach. Ecological Informatics, 58,101085.

[2] Gupta, S., Singh, A., & Kaur, R. (2021). Leveraging AI for butterfly taxonomy: Challenges and prospects. Computers and Electronics in Agriculture, 184, 106119.

[3] Zhang, L., Liu, X., & Zhang, J. (2022). An ensemble learning model for butterfly classification. Applied Soft Computing, 105, 107285.

[5] Chen, Y., Fang, W., & Dong, W. (2023). Comparative analysis of machine learning models for butterfly recognition. Ecological Modelling, 475, 110091.

[6] Tian, Y., Zhang, M., & Tang, Z. (2018). Butterfly image classification using deep convolutional networks. Computers and Electronics in Agriculture, 151, 149-160.

[7] Kaur, G., Singh, S., & Kaur, H. (2019). Automated butterfly classification using transfer learning with CNNs. Journal of Entomology and Zoology Studies, 7(6), 89-96.

[8] Nguyen, T. T., Bui, Q. T., & Vu, H. L. (2020). Butterfly species recognition based on machine learning and computer vision. Journal of Environmental Informatics, 35(3), 121-132.

[9] Rahman, M. A., Islam, M. R., & Rahman, M. M. (2021). Efficient butterfly classification using hybrid deep learning models. Applied Sciences, 11(4), 1793.

[10] Perez, L., Wang, J., & Weiss, M. (2022). Butterfly species identification through augmented reality and deep learning. Ecological Informatics, 67, 101503.