BUTTERFLY CLASSIFICATION USING TRANSFER LEARNING

SEMINAR-II PROJECT REPORT

Submitted by

BHARATHWAJ M – RA2011026020065

HARSHIT V – RA2011026020086

MADHESH B – RA2011026020098

Under the guidance of

Mr. T. RAJESH

Dr. T.P. LATCHOUMI

In partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING WITH SPECIALISATION IN ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

of

FACULTY OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
RAMAPURAM CAMPUS, CHENNAI-600089

MAY 2023

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Deemed to be University u/S 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

Certified that the Seminar-II report titled "**BUTTERFLY CLASSIFICATION USING TRANSFER LEARNING**" the bonafide work of **BHARATHWAJ M [RA2011026020065], HARSHIT V [RA2011026020086]** and **MADHESH B [RA2011026020098]** submitted for the course 18CSP106L Seminar –II . This report is a record of successful completion of the specified course evaluated based on literature reviews and the supervisor. No part of the Seminar Report has been submitted for any degree, diploma, title, or recognition before.

SIGNATURE

SUPERVISOR
Mr. T Rajesh, M. Tech.,
Assistant Professor,
Department of CSE,
SRM IST, Ramapuram

SIGNATURE

HEAD OF THE DEPARTMENT
Dr.K.Raja, M.E., Ph.D
Department of CSE,
SRM IST, Ramapuram

Submitted for the Seminar-II Viva Voce Examination held on .... at SRM Institute of Science and Technology, Ramapuram Campus, Chennai-600089.

EXAMINER 1

EXAMINER 2

ABSTRACT

Classification of butterfly species from images is a challenging task due to the high variability in their appearance and the large number of species. This task has important applications in various fields such as conservation biology and wildlife monitoring, where the identification and classification of butterfly species are crucial for understanding their behavior, habitat, and ecological roles.

In this project, we propose a system for butterfly classification using transfer learning, which combines deep learning and machine learning techniques. We used the ResNet50 model as the proposed deep learning model and traditional machine learning models as the existing system. The methodology used for classification is Convolutional Neural Networks (CNN). It is a type of deep learning algorithm commonly used in computer vision tasks, such as image classification and object detection. It is designed to automatically extract and learn hierarchical features from images using convolutional layers, pooling layers, and fully connected layers.The goal of this project is to develop an accurate and efficient system for classifying butterfly species based on their images.

Transfer learning is a machine learning technique that involves using a pre-trained model on a large dataset as a starting point for a new task or domain with a smaller dataset. In transfer learning, the pre-trained model is adapted to the new task by fine-tuning some of its layers or by adding new layers to the model, while keeping the weights of the other layers fixed. The dataset used in this project is a collection of butterfly images belonging to 70 different species. We preprocessed the images by resizing them to a fixed size and augmenting them using various techniques like rotation, zooming, and flipping.

The experimental results show that the proposed deep learning model outperformed the traditional machine learning models in terms of accuracy. We compared the performance of the proposed deep learning model with that of the existing machine learning models. We compared traditional machine learning algorithms, such as VGG19 (90%), Support Vector Machines (75%), Partial Least Squares (76%) and CNN with 4 Convolutinal layers (85%) as the existing system. We achieved an accuracy of 96 -97% on the test set using our proposed model. As we can see that our model easily outperformed the existing models.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION OF THE PROJECT

The Butterflies are one of the most diverse and beautiful insect groups, but identifying and classifying their species is a challenging task due to their wide range of colors, patterns, and shapes. Accurate and efficient classification of butterfly species is crucial for understanding their ecological roles, conservation, and research purposes. However, manual classification is a time-consuming and error-prone process, which calls for an automated solution using advanced machine learning techniques.

In recent years, deep learning algorithms, such as CNNs, have shown remarkable performance in solving complex image classification tasks. Transfer learning is a popular technique in deep learning that involves using pre-trained models as a starting point for new tasks, which has been shown to significantly reduce the amount of data and computational resources required to train new models from scratch.

In this project, we propose a system for butterfly classification using transfer learning, which combines deep learning and machine learning techniques. We used the ResNet50 model as the proposed deep learning model and traditional machine learning models as the existing system. The methodology used for classification is Convolutional Neural Networks (CNN). It is a type of deep learning algorithm commonly used in computer vision tasks, such as image classification and object detection.

It is designed to automatically extract and learn hierarchical features from images using convolutional layers, pooling layers, and fully connected layers. The goal of this project is to develop an accurate and efficient system for classifying butterfly species based on their images.

CHAPTER 2

PROBLEM STATEMENT

The Classification of butterfly species is a challenging task due to the high variability in their appearance and the large number of species. Accurate identification and classification of butterfly species are crucial for various applications such as conservation biology, agriculture, and wildlife monitoring. However, manual classification of butterfly species is time-consuming and prone to errors, which highlights the need for an automated classification system.

The patterns of butterflies are essential features for researchers and people who are interested in categorizing butterflies. However, it is difficult to classify butterflies based on biological patterns, such as

- Shapes,
- Wing colors and
- Veins

Currently, researchers rely mostly on time-consuming processes dealing with manually identifying and classifying by highly trained individuals. Therefore, to tackle, there is a need for modern and automated technologies in dealing with species identification with good accuracy.

Butterflies are a diverse group of insects with over 20,000 different species worldwide, each with unique physical characteristics and patterns. The classification of butterfly species based on their appearance is a challenging task for traditional image classification techniques due to the high variability in the physical characteristics of each species, as well as the subtle differences between similar-looking species.

To overcome these challenges, deep learning techniques such as transfer learning and convolutional neural networks (CNN) can be used. Transfer learning involves using a pre-existing, pre-trained model as a starting point and fine-tuning it on a new dataset. This approach has shown to be effective in achieving high accuracy in image classification tasks, even with small datasets. CNNs are particularly useful for image classification as they are designed to learn and identify features in images, making them well-suited for the task of identifying subtle differences between butterfly species.

CHAPTER 3

SCOPE & OBJECTIVE

3.1 OBJECTIVE

The Objective of our project is to develop a model that can accurately classify different species of butterflies using deep learning techniques. The project utilizes transfer learning with the ResNet50 model and compares its performance with standard machine learning models. The ultimate goal is to create a reliable and efficient system for identifying and classifying butterfly species, which can be used for research and conservation efforts.

3.2 SCOPE OF THE PROJECT

- The Scope of the project is to develop a deep learning-based model for classifying images of different butterfly species using transfer learning with ResNet50.
- Collect and prepare a dataset of images of butterfly species, including image acquisition and labeling.
- Implement image pre-processing techniques, such as resizing, normalization, and augmentation, to improve the performance of the classification model.
- Investigate the use of different hyperparameters, model architectures to optimize the performance of the classification model and compare the performance of the deep learning model with standard machine learning models.
- Explore the potential of the developed model for real-world applications, such as biodiversity research and conservation efforts.
- Ensure effective data management strategies for working with large datasets.
- Collaborate with domain experts in butterfly species classification and conservation to ensure accuracy and relevance of the model.

CHAPTER 4

EXISTING SYSTEM

The existing system refers to the traditional machine learning models that are commonly used for image classification tasks. These models include VGG19, CNN With 4 Convolutional Layers, Partial Least Squares (PLS) and support vector machines (SVMs). These models are typically trained on a dataset of labeled images, where the features of each image are extracted and used to predict the corresponding label.

However, these traditional machine learning models may not perform well on complex image classification tasks such as butterfly classification. This is because they rely on hand-engineered features that may not capture the complex patterns and structures in the images. This is where deep learning comes in.

EXISTING MODELS	ACCURACY
CNN With 4 Convolutional Layer	85%
Support Vector Machine (SVM)	75%
Partial Least Squares (PLS)	76%
CNN With VGG19	90%

Table 4.1: **Existing Models Accuracy**

CHAPTER 5

LITERATURE SURVEY

SI. NO.	JOURNAL NAME	YEAR OF PUBLISH	PAPER TITLE	AUTHOR	DESCRIPTION
1.	IEEE	2020	Butterfly Image Classification Using Convolutional Neural Network (CNN)	Nur Nabila, Nurbaity Sabri, Nur Farahin	This research paper presents a study on butterfly species identification using image processing techniques and Convolutional Neural Network (CNN). The aim of the study is to classify images of butterflies using CNN techniques and evaluate the performance of the classification.
2.	IEEE	2020	Using Partial Least Squares in Butterfly Species Identification	Alexandre Silva, Sincler Meireles, Samira Silva	This research paper proposes a novel approach to recognize butterfly species in images using handcrafted descriptors and the Partial Least Squares (PLS) algorithm. The approach involves extracting features from a butterfly dataset using Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG) descriptors, training PLS models using an one-against-all protocol, and presenting images to all classifiers during the test phase.

3.	IJCONSIST	2021	Classification of color features in butterflies using the Support Vector Machine (SVM)	Dhian Satria, Hendra Maulana	The paper discusses the use of butterflies as a dataset for research in digital images and focuses on the extraction process of color features using the RGB method and the Hue, Saturation, Value (HSV) color space. The proposed classification process uses a Support Vector Machine (SVM) to classify butterfly species.
4.	MDPI	2022	Classification of color features in butterflies using the Support Vector Machine (SVM)	FathimatuR ajeena, Rasha Orban, Malliga	This research work aims to assist science students in correctly recognizing butterflies without harming the insects during their analysis. This paper discusses transfer learning based neural network models to identify butterfly species.

CHAPTER 6

SYSTEM ARCHITECTURE

6.1 SYSTEM ARCHITECTURE DIAGRAM

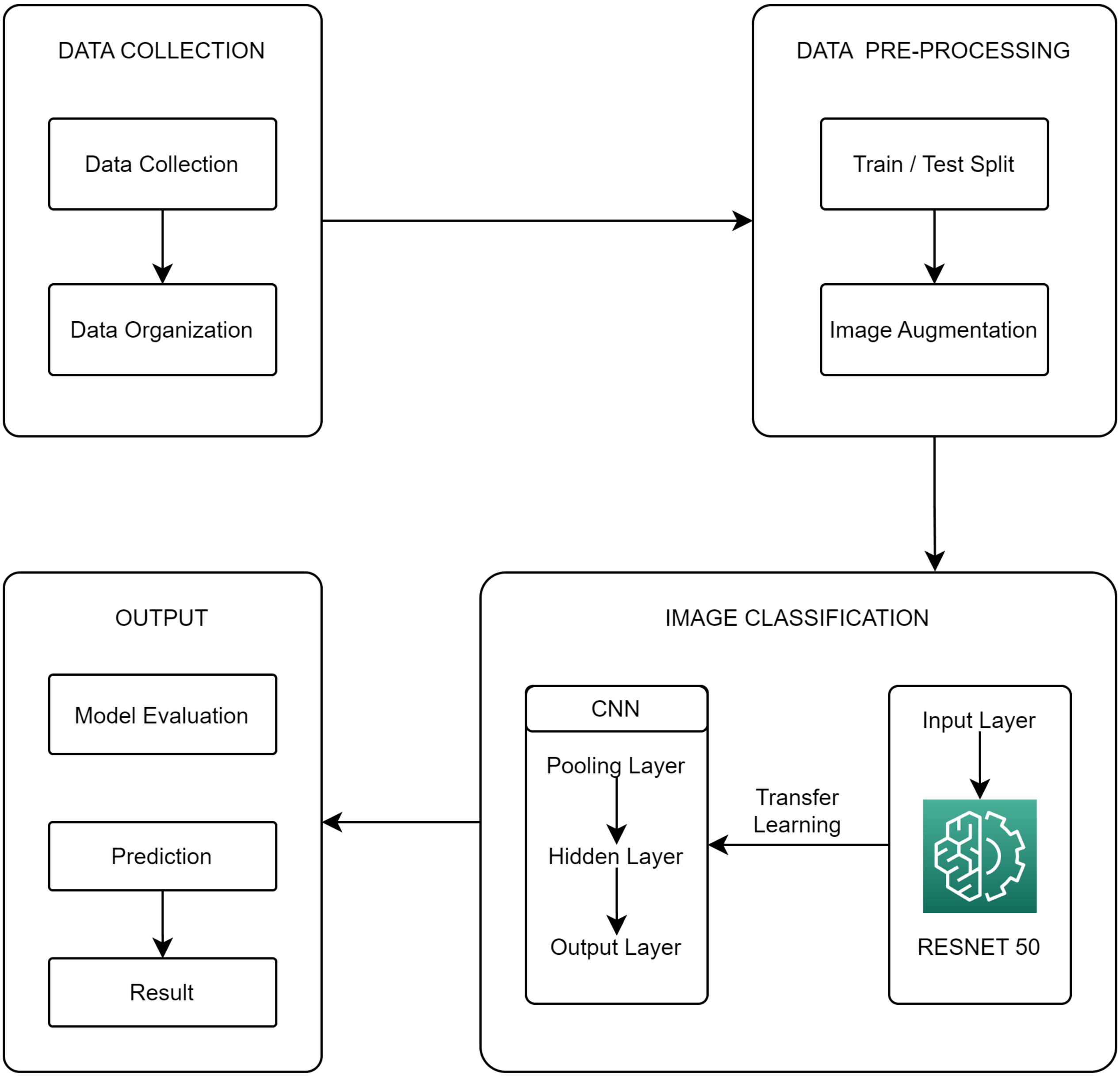


Figure 6.1: **System Architecture Diagram**

6.2 SYSTEM ARCHITECTURE DESCRIPTION

Our entire project is divided into four modules:

Step 1: Data Collection

Step 2: Data Pre-Processing

Step 3: Image Classification

Step 4: Evaluation & Prediction

6.2.1 Data Collection

The butterfly categories dataset available on Kaggle consists of images of 40 different butterfly species collected from various sources, including personal photographs, online databases, and scientific journals. The dataset provides a valuable resource for researchers interested in studying butterfly species, but it only includes images of a small fraction of the total number of butterfly species in the world.

Link to Dataset: https://www.kaggle.com/datasets/gpiosenka/butterfly-images40-species

6.2.2 Data Pre-Processing

In Data Pre-Processing of images, image augmentation is a technique used to increase the size of a dataset by applying various transformations to existing images. For butterfly images, augmentations such as

- Flipping,
- Rotation,
- Scaling,
- Cropping, and
- Adjustments to brightness and contrast

can be used to improve model performance. Care must be taken to apply appropriate augmentations that do not distort the characteristics of the butterfly species.

6.2.3 Building & Training CNN With Resnet-50

Transfer learning and CNN training are two important techniques commonly used in image classification, including butterfly classification.Transfer learning involves using a pre-trained model like ResNet50 to train a new butterfly classification model. The pre-trained model can be fine-tuned by retraining the last layers on the new butterfly image dataset. CNN training involves training a CNN model to classify images by adjusting the weights of the layers to minimize prediction error on a labeled dataset.

6.2.4 Evaluating The Accuracy And Loss

Accuracy evaluation is important in butterfly classification using transfer learning ResNet50 and CNN. It involves measuring the performance of the model using metrics such as accuracy, precision, recall, and F1 score. The test dataset should be split into validation and test subsets. The results can be used to identify areas for improvement in the model.

6.2.5 Predicting the Test Images

Once the butterfly classification model has been trained on a training dataset using transfer learning ResNet 50 and CNN, the next step is to predict the classes of test images. This involves passing the test images through the model and obtaining predictions for each image.The prediction process involves converting the test images into the format required by the model, which is typically a fixed size and color format. This may involve resizing the image, normalizing the pixel values, and converting the color space to match the model requirements. Once the images have been preprocessed, they can be passed through the model to obtain class predictions. The model will output a probability distribution over the different butterfly species, indicating the likelihood that the image belongs to each class.

```
This image most likely belongs to METALMARK
```

Figure 6.2: **Predicting the Test Image**

CHAPTER 7

PROPOSED SYSTEM AND METHODOLOGY

7.1 PROPOSED SYSTEM

In the proposed system, this project is a convolutional neural network (CNN) based on the ResNet50 architecture. Transfer learning is used in this project to leverage the pre-trained ResNet50 model, which was trained on a large dataset of images to identify various objects, including animals and insects.

Convolutional Neural Networks (CNN) is a type of deep learning algorithm commonly used in computer vision tasks, such as image classification and object detection. It is designed to automatically extract and learn hierarchical features from images using convolutional layers, pooling layers, and fully connected layers.The goal of this project is to develop an accurate and efficient system for classifying butterfly species based on their images.

Transfer learning is a machine learning technique that involves using a pre-trained model on a large dataset as a starting point for a new task or domain with a smaller dataset. In transfer learning, the pre-trained model is adapted to the new task by fine-tuning some of its layers or by adding new layers to the model, while keeping the weights of the other layers fixed. The experimental results show that the proposed deep learning model outperformed the traditional machine learning models in terms of accuracy. We achieved an accuracy of 96 -97% on the test set using our proposed model. Also this has been made as a web app using streamlit for easy access.

7.2 METHODOLOGY

The project aims to develop an accurate and efficient system for identifying and classifying images of different butterfly species using transfer learning with ResNet50. The methodology used is convolutional neural networks (CNN) with deep learning techniques. The project involves data collection and preparation, including image acquisition and labeling. The dataset of butterfly images is divided into training, validation, and testing sets, and exploratory data analysis is performed to understand the distribution and characteristics of the dataset.

The next step is data preprocessing, where the images are resized to a fixed size, normalized to ensure the input features have a similar scale, and augmented to increase the variability in the training set and prevent overfitting. Augmentation techniques may include random rotations, flips, and shifts.

Transfer learning is a machine learning technique that involves using a pre-trained model on a large dataset as a starting point for a new task or domain with a smaller dataset. In transfer learning, the pre-trained model is adapted to the new task by fine-tuning some of its layers or by adding new layers to the model, while keeping the weights of the other layers fixed.

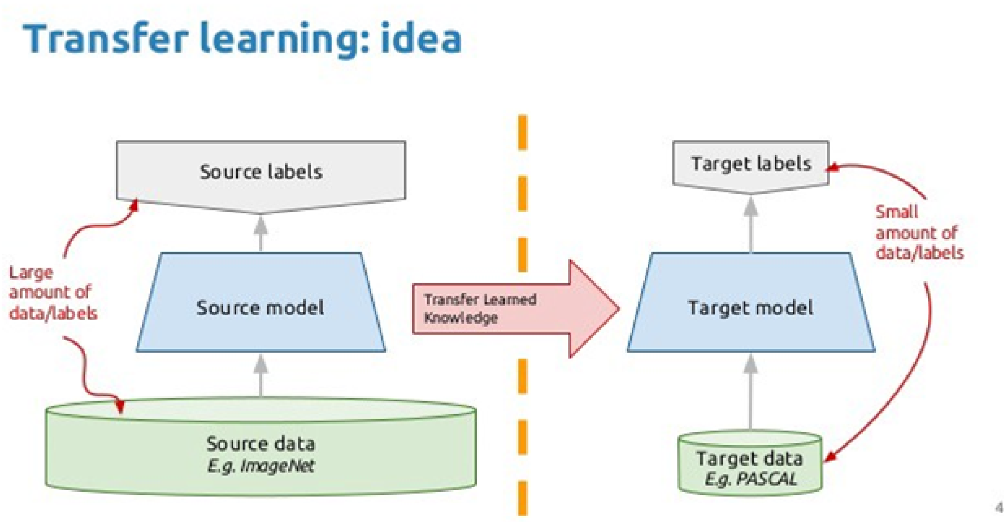


Figure 7.2 (a) : **Architecture of Transfer Learning**

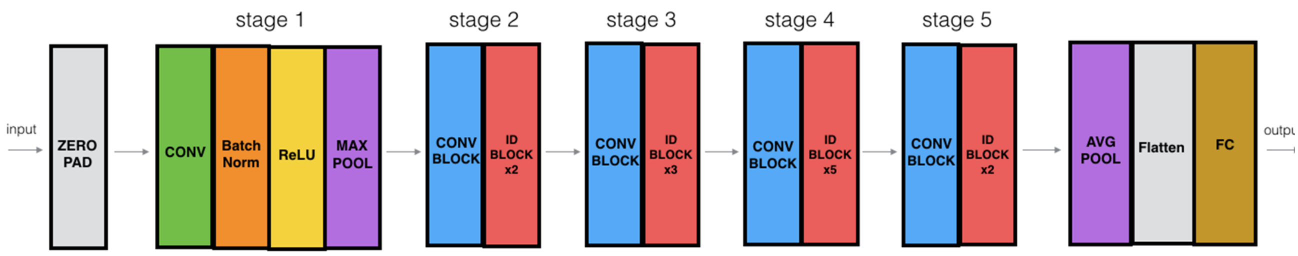


Figure 7.2 (b) : **Architecture of ResNet50**

The model architecture selected for the classification model is ResNet50. It consists of 40 convolutional layers, 1 max pool layer and 1 average pool layer. The initial layers of the model are frozen, and the later layers are fine-tuned for the butterfly classification task. The model is trained on the training set using transfer learning with ResNet50, and its performance is evaluated on the validation set to monitor its performance and prevent overfitting. The hyperparameters, such as learning rate, batch size, and number of epochs, are tuned to optimize the model's performance.

CHAPTER 8

IMPLEMENTATION SOURCE CODE

SOURCE CODE:

```
import numpy as np
import pandas as pd
import os
import tensorflow as tf
import numpy as np
from tensorflow.keras.applications.resnet50 import ResNet50
from tensorflow.keras.models import Model
from keras.preprocessing.image import ImageDataGenerator
from tensorflow.keras.layers import GlobalAveragePooling2D
from tensorflow.keras.layers import MaxPool2D
import matplotlib.pyplot as plt
```

```
dataset= tf.keras.preprocessing.image_dataset_from_directory(
    'train',
    shuffle=True,
    image_size=(224,224),
    batch_size=32
    )
```

```
Found 4955 files belonging to 50 classes.
```

```
class_names = dataset.class_names
class_names
```

```
['adonis',
 'american snoot',
 'an 88',
 'banded peacock',
 'beckers white',
 'black hairstreak',
```

```
img_height = 224
img_width = 224
batch_size = 1

train_ds = tf.keras.preprocessing.image_dataset_from_directory(
    'train',
    seed=123,
    image_size=(img_height, img_width),
    batch_size=batch_size)



test_ds = tf.keras.preprocessing.image_dataset_from_directory(
    'test',
    seed=123,
    image_size=(img_height, img_width),
    batch_size=1)
```

```
Found 4955 files belonging to 50 classes.
Found 250 files belonging to 50 classes.
```

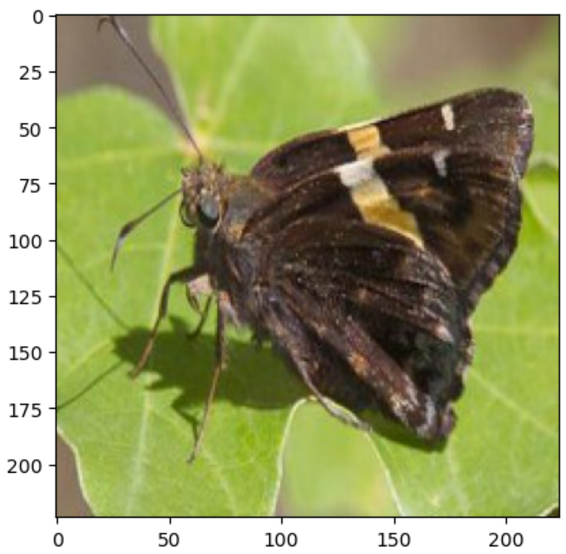
```
resnet_model = ResNet50(include_top=False,weights="imagenet")
x = resnet_model.output
x = GlobalAveragePooling2D()(x)
x = tf.keras.layers.Dense(1024,activation='relu')(x)
x = tf.keras.layers.Dense(512,activation='relu')(x)
res = tf.keras.layers.Dense(50, activation="softmax")(x)

# generate our model

model = Model(inputs=resnet_model.input, outputs=res)

# will not retrain weights

for layer in resnet_model.layers:
    layer.trainable = False

model.compile(
    optimizer=tf.keras.optimizers.SGD(learning_rate=0.001),
    loss=tf.losses.SparseCategoricalCrossentropy(),
    metrics=['accuracy'])

hist = model.fit(
    train_ds,
    validation_data=test_ds,
    epochs=15,
    verbose=1,
).history

model.summary()
```

```
acc= hist['accuracy']
val_acc= hist['val_accuracy']

loss= hist['loss']
val_loss= hist['val_loss']

plt.figure(figsize=(8, 8))
plt.subplot(1, 2, 1)
plt.plot(range(15), acc, label='Training Accuracy')
plt.plot(range(15), val_acc, label='Validation Accuracy')
plt.legend(loc='lower right')
plt.title('Training and Validation Accuracy')

plt.subplot(1, 2, 2)
plt.plot(range(15), loss, label='Training Loss')
plt.plot(range(15), val_loss, label='Validation Loss')
plt.legend(loc='upper right')
plt.title('Training and Validation Loss')
plt.show()
```

OUTPUT:

```
img = tf.keras.preprocessing.image.load_img('C:\\Users\\TESS\\Downloads\\3.jpg',
                                            target_size=(img_height, img_width))
img_array = tf.expand_dims(tf.keras.preprocessing.image.img_to_array(img), 0)

score = tf.nn.softmax(model.predict(img_array))
print("This image most likely belongs to {} ".format(class_names[np.argmax(score)]))
plt.imshow(img)
```

```
1/1 [==============================] - 1s 845ms/step
This image most likely belongs to gold banded
<matplotlib.image.AxesImage at 0x1b7b6a2a820>
```



CHAPTER 9

FUTURE SCOPE

8.2 FUTURE ENHANCEMENTS

There are several potential improvements that can be made to the current model, such as:

- **More data:** Increasing the size of the dataset by collecting more images of butterflies can help to improve the model's accuracy.

- **Fine-tuning:** Fine-tuning the pre-trained model can help to improve the model's accuracy by allowing the model to learn more specific features of the butterfly images.

- **Hyperparameter tuning:** Fine-tuning the hyperparameters of the model can also help to improve its accuracy.

- **Ensemble learning**: Using ensemble learning techniques, such as combining multiple models or training the same model with different initializations, can also help to improve the model's accuracy.

- **Transfer learning with other pre-trained models**: In addition to ResNet-50, there are several other pre-trained models that can be used for transfer learning, such as VGG-16, InceptionV3, and MobileNet. Evaluating the performance of these models on the butterfly classification task and selecting the best one can potentially improve the model's accuracy.

- **User interface:** Developing a user interface that allows users to interact with the model and visualize its predictions can help to make the model more accessible and usable for non-technical users.

- **Adversarial training:** Adversarial training is a technique where the model is trained on adversarial examples, which are intentionally modified images designed to deceive the model. Incorporating adversarial training into the butterfly classification model can potentially improve its robustness and reduce the risk of misclassification on real-world data.

Overall, the butterfly classification model we have implemented is a promising start, and there are several avenues for further improvement

CHAPTER 10

DISADVANTAGES

While using transfer learning with ResNet-50 and CNNs can be a powerful approach for image classification tasks like butterfly classification, there are several potential disadvantages and limitations to be watched.

- **Limited data:** Transfer learning relies on having a pre-trained model that has already been trained on a large dataset of images. However, if the target task has very different characteristics or features than the original dataset, there may not be enough data to fine-tune the pre-trained model effectively. This can result in overfitting or poor generalization to new data.

- **Overfitting:** As with any machine learning model, there is a risk of overfitting the model to the training data. This is especially true when using transfer learning with a pre-trained model, as the model may be very complex and have many parameters that can easily overfit to the training data. Careful regularization and hyperparameter tuning can help mitigate this issue.

- **Limited interpretability:** Deep neural networks like ResNet-50 and CNNs can be difficult to interpret and understand, especially when they are pre-trained on large datasets. This can make it challenging to diagnose errors or understand why the model is making certain predictions. Additionally, transfer learning with pre-trained models can make it even more difficult to interpret the learned features and understand how they are related to the target task.

- **Computing resources:** Training deep neural networks like ResNet-50 and CNNs can be computationally expensive, especially when using large datasets or complex architectures. This can require specialized hardware or cloud computing resources, which can be costly and time-consuming.

- **Bias and fairness:** Pre-trained models may be biased towards certain types of images or classes, which can impact the accuracy and fairness of the model. Additionally, the training data may not be representative of the full range of diversity in the target task, which can lead to biased or unfair predictions.

CHAPTER 11

CONCLUSION

8.1 CONCLUSION

To overcome these challenges, deep learning techniques such as transfer learning and convolutional neural networks (CNN) can be used. Transfer learning involves using a pre-existing, pre-trained model as a starting point and fine-tuning it on a new dataset. This approach has shown to be effective in achieving high accuracy in image classification tasks, even with small datasets. CNNs are particularly useful for image classification as they are designed to learn and identify features in images, making them well-suited for the task of identifying subtle differences between butterfly species.

The transfer learning approach with ResNet-50 helped us to leverage the pre-trained model's knowledge and train only the last layers of the network. This approach resulted in faster training and better accuracy. We have also used data augmentation techniques to increase the size of the training dataset, which further improved the model's performance.

The experimental results show that the proposed deep learning model outperformed the traditional machine learning models in terms of accuracy. We compared the performance of the proposed deep learning model with that of the existing machine learning models. We compared traditional machine learning algorithms, such as VGG19 (90%), Support Vector Machines (75%), Partial Least Squares (76%) and CNN with 4 Convolutinal layers (85%) as the existing system. We achieved an accuracy of 96 -97% on the test set using our proposed model. As we can see that our model easily outperformed the existing models.