Leopold-Franzens Universität Innsbruck
Faculty of Mathematiks, Computer Science and Physics

Department of Mathematics

ULG Data Science

Master Thesis

Image-based butterfly species identification with convolutional neural networks

of

Friederike Barkmann, MSc
(Matr.-Nr.: 11731745)

Submission Date: March 13, 2025
Supervisors: Stephan Antholzer, PhD and Dr. Johannes Rüdisser

Abstract

Automated images-based species identification has a high potential to support biodiversity data collection, especially as the number of collected records increased considerably in recent years. In this study, a pre-trained convolutional neural network (ResNet-152) was tuned to classify butterfly and moth species based on images and its performance was assessed. The dataset used for training was collected by citizen scientists with the application "Schmetterlinge Österreichs". It has over 500,000 images of 162 species and is considerably larger than datasets in similar studies.There were large differences on the number of images per species, so I employed and evaluated methods to handle this class imbalance. A high accuracy of 97.1 % on testdata was achieved when not correcting for class imbalance. Models trained with methods that better represent small classes reached slightly lower accuracies but had higher mean recall per species (93.2 % for oversampling of classes with only few species and 90.9 % for the application of a weighted loss function). Species with many training images could be predicted with high recall and precision, while images from small classes displayed lower mean values for both recall and precision, as well as higher variance. Among the species groups that could be identified accurately by the models are the family Papilionidae and the tribe Nymphalini which contain many species with characteristic wing patterns. Identification was more difficult for species of the families Lycanidae and Hesperiidae and the genus *Erebia.* The achieved accuracy of the classification was higher than in other studies. It demonstrates the applicability of deep learning models for species identification which can reduce human effort and provide reliable feedback to citizen scientist that collect biodiversity data.

Contents

1 Acknowledgments

I want to thank my supervisors Stephan Antholzer and Johannes Rüdisser for their advice and especially Johannes for his support that made it possible for me to write this thesis next to my work as a PhD student in the Viel-Falter butterfly monitoring.

I want to thank Ronald Würflinger from the Billa Foundation ”Blühendes Österreich” for the provision of the large image dataset of the application ”Schmetterlinge Österreichs” and for the valuable exchange during this project. I also want to thank all the volunteers that were involved in recording butterflies and moths all over Austria and Helmut Höttinger who is responsible for verification of the correct identification of species.

I am very grateful that my work was supported by the project EuroCC Austria and want to especially thank Andreas Lindner for his support with adapting this project for high performance computing, the data parallelization that considerably reduced training time and all the answers to my many questions.

The computational results presented here have been achieved in part using the LEO HPC infrastructure of the University of Innsbruck.

I acknowledge the EuroHPC Joint Undertaking for awarding this project access to the EuroHPC supercomputer LEONARDO, hosted by CINECA (Italy) and the LEONARDO consortium through an EuroHPC Development Access call.

2 Funding

This project was supported by the **Viel-Falter Butterfly Monitoring** which has received funding from the Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK) and by the project **EuroCC Austria** which has received funding from the European High Performance Computing Joint Undertaking (JU) and Germany, Bulgaria, Austria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, Greece, Hungary, Ireland, Italy, Lithuania, Latvia, Poland, Portugal, Romania, Slovenia, Spain, Sweden, France, Netherlands, Belgium, Luxembourg, Slovakia, Norway, Türkiye, Republic of North Macedonia, Iceland, Montenegro, Serbia under grant agreement No 101101903.

3 Zusammenfassung

Citizen Science Programme zur Beobachtung von Biodiversität liefern weltweite Daten von Millionen Teilnehmer*innen und leisten einen wichtigen Beitrag zur Biodiversitätsforschung. Dabei werden oft große Mengen an Bildern von zu beobachtenden Organismen gesammelt. Die Sicherstellung der korrekten Artbestimmung anhand der Bilder kann einen hohen Arbeitsaufwand bedeuten. Automatisierte Bestimmungsmethoden können diesen Aufwand verringern und mit Artvorschlägen zudem zu Verbessung der Artkenntnisse der Nutzer*innen beitragen.

Für die Klassifizierung von Bildern sind ”Convolutional Neural Networks” (CNN) besonders gut geeignet. In dieser Arbeit wurde das CNN Modell ResNet-152 mit einem Datensatz der App ”Schmetterlinge Österreichs” der Billa-Stiftung Blühendes Österreich[1] auf die Bestimmung von Schmetterlingen trainiert. Der verwendete Datensatz umfasst über 500.000 Bilder von Tag- und Nachtfaltern und ist damit deutlich größer als in ähnlichen Studien bisher verwendete Datensätze [10, 38, 48]. Wie es bei solchen Datensätzen häufig der Fall ist, unterscheidet sich die Anzahl an Bildern pro Arten stark. Es wurden daher verschiedene Methoden angewandt, um diese Verteilung beim Trainieren des Modells auszugleichen. Es wurde untersucht, wie genau einzelne Arten und Artgruppen mit den unterschiedlichen Methoden bestimmt werden konnten. Um die mögliche Arbeitseinsparung abschätzen zu können, wurde analysiert, welcher Anteil der Bilder vom Modell bestimmt werden kann, wenn eine Genauigkeit von mindestens 99,5 % erreicht werden soll.

[1] https://www.bluehendesoesterreich.at/

Es konnten 97,1 % der Bilder korrekt bestimmt werden. Diese hohe Genauigkeit wurde mit einem Modell ohne Korrektur der ungleichen Verteilung der Daten erreicht. Diese Modell war daher besonders stark auf die Arten mit vielen Bildern optimiert und hatte die beste Leistung auf dem gesamten Testdatensatz. Ein Ansatz, bei dem Bilder von kleinen Klassen (Arten mit wenig Bildern) häufiger im Training verwendet werden, zeigte ebenfalls eine hohe Genauigkeit mit 96,3 %. Arten mit wenig Bildern werden durch diese Methods besser vom Modell repräsentiert. Dies zeigt sich im höheren Mittelwert der Sensitivität (Recall) über alle Arten, der mit der Korrektur (93,2 %) höher ist als ohne (90,1 %). Der Wert gibt an, welcher Anteil der Bilder einer Art vom Modell auch dieser Art zugeordnet wird.

Die Modelle funktionieren für Arten mit vielen Bildern sehr gut, während die Ergebnisse für Arten mit wenigen Bildern gemischt sind. Die teilweise hohe Fehlerrate bei Arten mit wenig Bildern könnte zum einen direkt an der geringen Anzahl an verfügbaren Bildern liegen. Zum anderen kann es sein, dass gerade unauffällige und schwierig zu bestimmende Arten seltener fotografiert werden. Arten der Familie Papilionidae (Ritterfalter) und des Tribus Nymphalini (mit z.B. *Aglais urticae* (Kleiner Fuchs) und *Vanessa atalanta* (Admiral)) konnten besonders gut vom Modell bestimmt werden. Beide Gruppen enthalten viele auffällige, gut anhand des Flügelmuster bestimmbare Arten. Arten der Familien Lycaenidae (Bläulinge) und Hesperiidae (Dickkopffalter) und der Gattung *Erebia* wurden vom Modell hingegen weniger oft korrekt bestimmt.

Welche Genauigkeit der Bestimmungen ausreichend ist, hängt von der Anwendung ab. Die Genauigkeit kann erhöht werden, wenn Bestimmungen mit geringer Sicherheit verworfen werden. Für eine korrekte Bestimmung von 99.5 % der Bilder, können ca. 90 % der Bilder vom Modell verarbeitet werden. Die restlichen 10 % bedürfen einer zusätzlichen Kontrolle durch Expert*innen. Ist ein Schmetterlinge nur klein abgebildet oder verdeckt, wird die automatische Bestimmung erschwert. Zudem ist wichtig, dass Bestimmungsmerkmale erkennbar sind. Entpsrechende Richtlinien für Nutzer*innen von Apps können die Genauigkeit der Bestimmung zusätzlich erhöhen.

4 Introduction

The number of image-based species records has increased considerably in recent years due to the use of automated cameras and the participation of citizen scientists [5]. While automated cameras provide information about species occurrences at selected locations, citizen science projects can have a broad spatial coverage with images of species that are collected by millions of participants from all over the world [13, 8]. Such data are a valuable source of information on species occurrence and substantially contribute to biodiversity monitoring and research [9]. Ensuring high accuracy of species identification from provided images is one of the challenges of working with biodiversity data sourced from citizen science apps [49]. Due to the high number of images, the confirmation of species identifications by experts can be time-consuming and hence expensive. Automated machine learning methods for species identification from images have potential to reduce this effort and increase benefits for users of applications for biodiversity data collection. By providing direct feedback, automated machine learning methods increase the educational value of such applications and motivate users to add more observations [33].

Machine learning techniques open up new possibilities to extract information from images and to make use of large image datasets in biodiversity monitoring in a time- and cost-efficient way [4, 53]. Convolutional Neural Networks (CNNs) are especially suited for visual recognition tasks such as object detection, semantic segmentation, and image classification because of their strong feature learning ability [11]. They have already been used for the identification of different taxa, such as vertebrate species in terrestrial and marine habitats [43, 15, 17], insects [19, 37, 48] and plants [35, 55]. CNNs are implemented in apps for species identification and recording, such as Flora incognita [35] or iNaturalist [1]. Although neural networks and other machine learning methods are already being used in different biological applications, their potential to facilitate biodiversity data collection is not yet fully explored and many challenges remain for their successful application [5, 17].

An effective application of CNNs and other machine learning models requires a large amount of training data [5]. The labeling of a sufficient number of images with the correct species can require a high initial investment before classification can be automated [4]. The number of required images depends on the required accuracy of model predictions and the variability within and between classes. Images of species in their natural environment are especially challenging because they often have high variability within a class. This inter-class variability can be the result of images being taken with diverse backgrounds and lighting situations, from variable distances and angles and of different parts, developmental stages, or behaviors of an organism. The between-class variability differs between different taxa and not all are suited for identification based on images. in particular insects can be challenging or impossible to identify from morphological features even for experts [34].

Another common challenge with datasets of species records is a high class imbalance: while some species are common and can be observed and photographed easily, others are rare, life hidden or are difficult to photograph [29]. Even if a high total number of images is collected, there might be only few images from minority classes (species with few images) compared to the majority classes (species with many images). The low absolute number of images from minority classes is a problem, when there are too few images for generalization to other images from that class. Data augmentation can reduce overfitting and improve model performance by increasing the size of the training dataset. This is achieved by applying different transformations on the existing images such as geometric transformations (e.g. flipping, rotations, warping), color space transformations or noise injection [41, 45]. Transfer learning, using a model that was pre-trained on a dataset with similar features, was also shown to reduce the amount of required training data [54] and to be especially beneficial for small datasets [56]. In addition, the low relative number of images from minority classes can also be problematic. Due to the class imbalance, minority classes are less well represented during training which negatively affects model performance for those classes [7]. Different techniques were proposed to overcome the problems that are associated with imbalanced datasets. A more balanced distribution of the data can be achieved by oversampling the minority classes or by undersampling the majority classes [7, 30]. Another approach is to apply a weighted loss function, that assigns a higher penalty to misclassifications from minority classes [44].

Butterflies are a well-suited insect group for the application of automated species identification from images. In contrast to many other groups of insects, most butterflies and many moth species can be distinguished without the assessment of microscopic characteristics and, in many cases, pictures contain the necessary information for species identification. However, challenges of using butterflies still include highly similar looking species or species with high intra-specific variability as well as species that can only be distinguished when a certain side or part of the wing is visible [10]. Butterflies have a positive image [42] and their characteristic wing patterns make them popular photo motives. In addition, they are a well-studied group that is often used as a biodiversity indicator due to many favorable characteristics. Butterflies react sensitively to changes in environmental conditions [6, 52], inhabit a wide range of terrestrial habitats [52] and are representative for many (but not all) groups of terrestrial insects [50, 14, 2]. Considering the important role of butterflies as biodiversity indicators, the development and assessment of methods for their automated species identification is highly relevant in times of increasing pressure on and a consequential decline of biodiversity [25]. In addition, such methods can increase public knowledge about butterflies and moths and raise awareness for insect diversity through educational applications.

Various studies have already explored the possibilities of an automatic image-based identification of butterflies using neural networks. Within these studies, different features including characteristics of wing shape, color and texture were used in combination with artificial neural networks to identify butterfly species [26, 27, 31]. While being successful in the identification of some species, all these studies have the drawback that they did not use images from butterflies in their natural environments but instead standardized pictures of collected and pinned specimen. Besides this, they worked on relatively small datasets with only few species and therefore, the transferability and practical use of the proposed methods is limited. CNNs and the availability of larger datasets has opened new possibilities in the field. CNNs could correctly predict over 98 % of images in a study that included 10 species and 832 images [36]. For larger datasets with more species, accuracies were lower. Chang et al. [10] reached an accuracy of 0.71 with ResNet18 on a dataset with 14,270 images of 636 species. Nie et al. [38] reached an accuracy of over 0.95 with a ResNet model on a dataset with 10,881 images of 82 species and Theivaprakasham et al. [48] predicted butterfly species with an accuracy of nearly 0.95 based on a dataset with 34,024 images of 315 species using ResNet-152.

The aim of this thesis was to assess the performance of the convolutional neural network ResNet-152 for the classification of Austrian butterflies and day-active moths. The model was trained on a dataset collected by Citizen Scientists with the application "Schmetterlinge Österreichs"[1] of the Billa-Foundation *Blühendes Österreich*. With over 500,000 images of 166 butterfly and 32 moth species the dataset is considerably larger than the ones used in other studies. Correct identification of the images in the dataset was verified by an experienced entomologist. The dataset has a high class imbalance with over 30,000 images for the most common and only a few images for rare species.

The following research questions were addressed:

- How high are the top-1, top-3 and top-5 accuracy of the ResNet-152 model

[1] www.schmetterlingsapp.at

that is trained on the above mentioned dataset? (For the top-1 accuracy the highest probability classification has to be correct, while for the top-3 and top-5 accuracy the correct class can also be in the three or five classes with the highest predicted probability.)

- How does overall and per class accuracy differ between a model trained on the original imbalanced dataset, one trained with over-sampling of minority classes and one trained with a weighted loss function?
- Which species and species groups can be predicted with high accuracy and which are difficult to predict?
- With what confidence threshold can an accuracy of 99.5 % be provided and how high is the proportion of images that are predicted with that confidence?

5 Methods

5.1 Dataset

The original dataset from the butterfly app of the foundation “Blühendes Österreich” contains 541,677 images of 185 butterfly and moth species. Images were taken by users of the app between 2016 and 2023 with an unstructured scheme, meaning that users had no guidelines on where to record species but could freely choose locations. When recording a species, users uploaded one or multiple images of that species with the location. All users of the app could suggest a classification (i.e. species identification) for an image. The correct classification of all images was verified by a supervising expert from the foundation ”Blühendes Österreich”.

The dataset contains images that show different life stages of butterflies. For all species except a few moth species, adult life stages were photographed most often. As the identification of adult butterflies and moths is also the focus of this study, 11,273 images that showed eggs, larvae or pupae and images with more than one species were excluded. The dataset was reduced to species with at least 50 images to ensure that enough images remain in the test dataset to evaluate performance for individual species. The dataset that was finally used in this study contains 529,835 images of 162 species. It contains 131 of the 210 butterfly species [21] and 31 of the about 4000 moth species that occur in Austria [23]. In the cleaned dataset, the largest class contains 29,612 images, hence the class imbalance ratio is about 1:150. 108 species had less than 1000 images and 39 species less than 100 images (fig. 5.1).

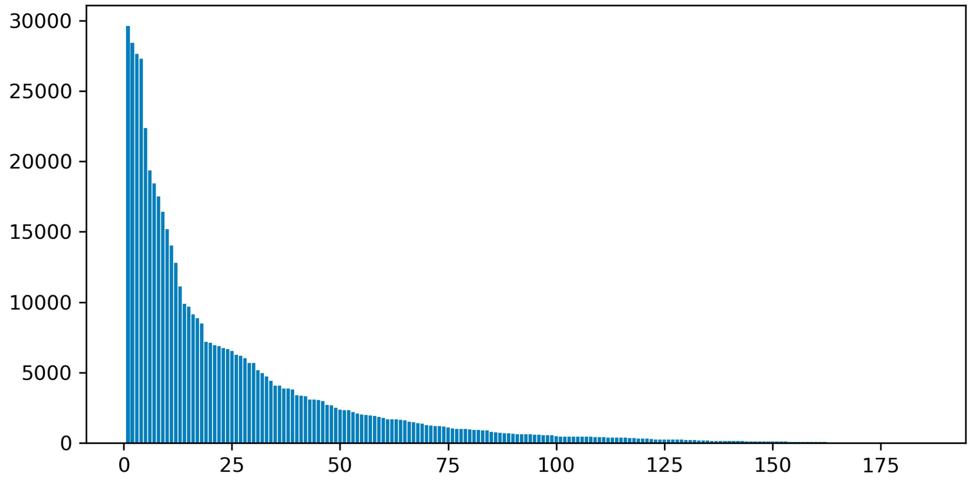

Figure 5.1: Distribution of the number of images per species for the dataset with >500,000 images of butterflies and moths that were collected with the application ”Schmetterlinge Österreichs” of the Billa Foundation ”Blühendes Österreich” between 2016 and 2023

10 % of the dataset were randomly selected for testing the models using a stratified sampling to ensure that the same proportion of images is taken from each class. The remaining 90 % were divided in 80 % training and 20 % validation data in each training run, also using a stratified sampling.

5.2 Data augmentation

Data augmentation was applied during training to enhance variability of training data. Training images were cropped to an area between half and the full area of the original image, the aspect ratio was randomly changed to a value between 0.8 and 1.2, and the images was resized to 224 x 224 pixels. Images were flipped horizontally and vertically with a probability of 0.3 for each, distorted with a scale of 0.2 with a probability of 0.4 and randomly rotated between -50° and 50°. The RGB colour channels were normalized based on ImageNet with the means 0.485, 0.456, 0.306 and standard deviations 0.229, 0.224, 0.225. Images in the validation and test datasets were resized to 224 pixel at the shorter site, randomly cropped to 224 x 224 pixel and normalized in the same way as the training data. See fig. 8.5 and fig. 8.6 supplements for randomly selected examples of original and augmented images.

5.3 Model

A ResNet-152 model [20] that was pre-trained on ImageNet [12] was used. The last layer of the model was adapted to the number of classes in this study. Due to the large amount of the available training data and their difference to the ImageNet dataset a full fine-tuning of the pre-trained ResNet-152 model, with all parameters rendered trainable, was conducted to get the best performance out of the full model. ResNet is a residual learning network that solves some of the problems connected to high network depth. Increasing the depth of a neural network can improve network performance but with too many layers performance was observed to decrease again. This degradation in accuracy is caused by difficulties with the optimization of large networks [11]. Residual learning networks overcome this limitation with residual connections (or shortcut connections) that connect activations of a layer with another layer by skipping layers in between [20]. ResNet models performed best in different studies that compared model performance for butterfly identification [10, 48, 38]. Of these studies, Theivaprakasham et al. [48] were the only ones to include multiple variants of ResNet and achieved the best performance with the ResNet-152 based on a dataset with >34,0000 images of 315 species. Among the compared models were different variants of ResNet [20], DenseNet [22] and VGG [46]. As the classification tasks are similar, ResNet-152 was also used in this study.

5.4 Hyper-paramters

The model was trained over 50 epochs (iterations over the whole training dataset) with a batch size of 128. The batch size defines how many images are processed before the model parameters are updated and is an important hyper-parameter that influences model accuracy and efficiency.

Categorical cross entropy loss and the Adam algorithm [28] were used for model optimization. Adam is a computationally efficient optimizer that is well-suited for machine learning problems with large datasets and many model parameters [28]. Theivaprakasham et al. [48] compared different optimization algorithms for butterfly identification with CNNs and found that Adam converged faster than other algorithms.

The learning rate was set to 0.0004 at the beginning of training and multiplied with 0.5 if validation accuracy did not improve for two epochs. The learning rate controls how much the model parameters are changed based on the loss function. It strongly affects model performance and can be challenging to choose. Too small a learning rate can result in slow training while too large a learning rate can lead to overshooting and lack of convergence [16] (Subchapter 8.3). The learning rate is reduced during training to allow for faster training in the beginning and better convergence in later epochs.

5.5 Handling class imbalance

To obtain a balanced representation of the classes during training, two approaches were applied. In the first approach, minority classes were oversampled and majority classes undersampled when loading the images for training the model. Each image was given a weight that is the inverse of the number of images in the class an image belongs to. In each epoch of training, images are selected with replacement with a probability that is proportional to their weight and data augmentation is applied to each of them prior to using them for training the model. Images from minority classes are therefore selected more often compared to those from majority classes. The overall number of selected images in each epoch was the same as the size of the dataset for comparability with the other approaches. In the second approach, a weighted loss function was used that applies a higher penalty to misclassifications of images from minority classes. Weights were again the inverse of the number of images in the class an image belongs to. The model was also trained using the original dataset without correction for class imbalance.

5.6 Model evaluation

The performance of the models that were trained with the above mentioned approaches was assessed based on the test dataset. For each model, the top-1, top-3 and top-5 performance over all images were assessed. The top-1 performance evaluates how often the correct class is the one predicted by the model with the highest probability. For the top-n accuracy it is evaluated how often the correct class is one of the n classes with the highest probability. Each predicted image has the same impact on the accuracy, hence, classes with many images have a stronger influence on this metric. In addition, precision and recall were calculated for each species and mean values over all species calculated to assess model performance with a stronger emphasis on species with few images. Precision assesses how many images that are predicted to belong to a certain species really belong to that species (true positives /(true positives + false positives)). Recall which is also called sensitivity assesses how many of the images that truly belong to a class are predicted to be in that class (true positives / (true positives + false negatives)).

Precision and recall per species were compared for groups of species to assess whether some groups can be determined more easily than others. While some butterfly families contain only few species in Austria (e.g. Riodinidae with only one species and Papilionidae with seven species) the largest family (Nymphalidae) contains over

100 species. Groups were therefore built based on different taxonomic levels from genus to family with the aim to group species with similar characteristics. The species *Hamaeris lucina* is the only species in the family Riodinidae and was not assigned to any group. The number of species per group ranged from 6 to 35 (table 5.1). See figures 8.1 to 8.3 for images of some species of each group.

Table 5.1: Number of species in each of the butterfly groups that were used to assess model performance

Group	**Taxonomic Level**	**Number of Species**
Lycaenidae	family	35
Moth	no taxon	31
Satyrinae (without Erebia)	subfamily	19
Argynnini	tribe	14
Pieridae	family	13
Hesperiidae	family	13
Nymphalini	tribe	9
Melitaeini	tribe	8
Limenitidinae and Apaturinae	subfamily	7
Erebia	genus	6
Papilionidae	family	6

It was assessed at what confidence level of the predictions an accuracy of 99.5 % can be reached and what proportion of the images are covered with that threshold.

5.7 Software and high-performance computing

Training on such a large dataset as the one used in this study can take a long time on one graphics processing unit (GPU), e.g. a consumer GPU on a laptop. Therefore, high-performance computing (HPC) resources were used. The training of neural networks is highly parallelizable and even a single GPU is a massively parallel processor, consisting of hundreds or thousands of compute cores that share the workload. An HPC system takes this further and several GPUs can be connected to collaborate simultaneously on solving a problem. In order to accelerate the training of the model described in this study, a technique called data parallelism was used. With data parallelism a model is replicated on every used GPU, while every GPU trains the model on a subset of the data. Every epoch the data gets shuffled. The initialization of the model on each device is identical and sophisticated communication patterns between the GPUs ensure that the parameter updates of the model are constantly synchronized. The PyTorch library was used [40], which natively comes with a framework called Distributed Data Parallel (DDP) [32]. It enables data parallel training, with the data distributed among compute devices such as GPUs. A further layer of abstraction is provided by Hugging Face with its Accelerate library [18] that makes using PyTorch DDP more straightforward.

A number of proof of concepts of the scaling with data parallelization using DDP with Accelerate were conducted on LEO5 [57]. LEO5, the latest HPC system in the LEO series at the University of Innsbruck, contains 54 Nvidia enterprise GPUs of different kinds. Nvidia A30 GPUs were used – the most common ones on LEO5 – and showed a significant speedup. The time for an epoch could be reduced from an initial two hours to twelve minutes going from one to four GPUs. This has been achieved with additionally increasing the bandwidth of loading data onto the GPUs by utilizing more compute cores for this task. The speedup can thus not only be attributed to the higher number of very capable GPUs, but also to the powerful CPUs (central processing units) that steer and connect the GPUs.

LEO5, however, is not a dedicated GPU, or AI, system, and resources are shared on a fine-grained level. This can lead to either long waiting times in the queue or tempt the user to request non-optimal resource allocations. Therefore the EuroHPC supercomputer LEONARDO [51] was used for training the model on the full dataset. The supercomputer LEONARDO hosted by CINECA (Italy) and the LEONARDO consortium and the number 9 in the TOP500 list [47] at the time of writing, has a

Booster GPU partition equipped with more than 12,000 custom Nvidia A100 GPUs. It stands out with its high availability and therefore short queuing times and a fast interconnection of GPUs. For training on LEONARDO 8 GPUs were employed, with a batch size of 16 processed by each. Since the GPUs work on individual batches synchronously, the effective total batch size increased to 128. A large batch size reduces the randomness, or noise, that is crucial to discover lower minima in the loss landscape via new paths. Choosing too low a batch sizes per GPU, on the other hand, increases computation times as GPUs have to synchronize more frequently.

5.8 Code and Data Availability

The scripts that were used to run the models on the supercomputer LEONARDO are available on github:

https://github.com/FriederikeBarkmann/CNN_butterfly_identification.

The trained models are available on Hugging Face:

https://huggingface.co/RikeB/CNN_butterfly_identification.

The images that were used for training can be viewed on the homepage of the butterfly app of Blühendes Österreich: https://schmetterlingsapp.at/.

There are plans to publish the dataset in a more widely usable format.

6 Results

6.1 Performance of the approaches

The accuracy on test data was highest for the model with no correction of class imbalance with 0.9711. The model trained with oversampling of minority and undersampling of majority classes had an accuracy of 0.9631 and the one with a weighted loss function an accuracy of 0.9484. Top-3 performance was >0.98 for all models and reached >0.99 for the best performing model. Top-5 accuracy was >0.99 for all models (table 6.1).

See 8.4 in the supplements for examples of wrongly predicted images.

Table 6.1: Overall and mean per species accuracy of ResNet-152 with different methods to handle class imbalance during training; none: no correction of class imbalance, oversampling: oversampling of minority classes and undersampling of majority classes, weighted loss: weighted loss function

Handling class imbalance	**Top-1**	**Top-3**	**Top-5**
None	0.9711	0.9923	0.9953
Oversampling	0.9631	0.9893	0.9932
Weighted loss	0.9484	0.9844	0.9907

During training, validation loss and accuracy did not improve further after <10 epochs for all models (fig. 6.1 and 6.2).

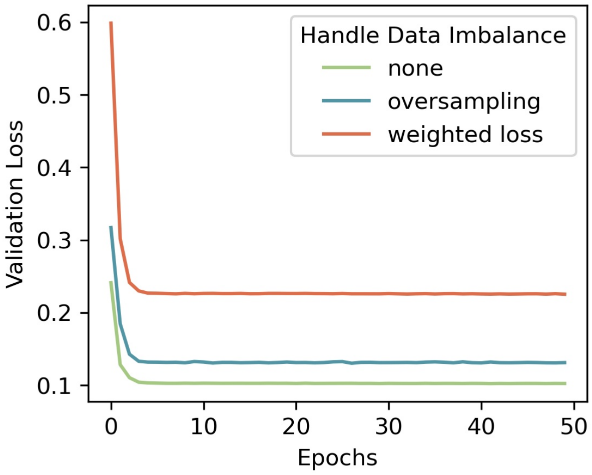

Figure 6.1: Validation loss during training of ResNet-152 with different methods to handle class imbalance; none: no correction of class imbalance, oversampling: oversampling of minority classes and undersampling of majority classes, weighted loss: weighted loss function

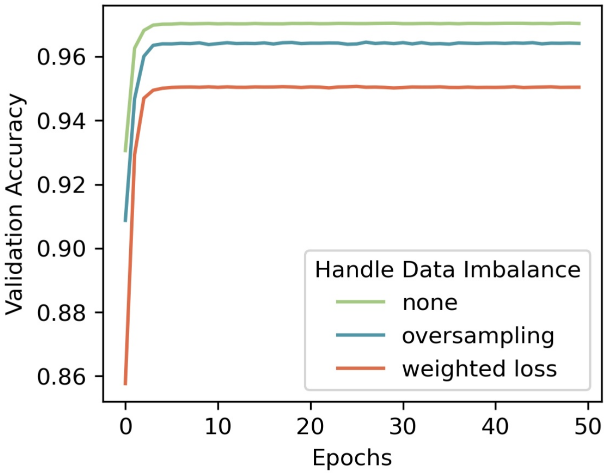

Figure 6.2: Validation accuracy during trainingof ResNet-152 with different methods to handle class imbalance; none: no correction of class imbalance, oversampling: oversampling of minority classes and undersampling of majority classes, weighted loss: weighted loss function

6.2 Performance by species

The mean recall over all species in the dataset was highest for the model with oversampling with 0.9321. The model with weighted loss reached a mean recall of 0.9085 and the one without correction of class imbalance a mean recall of 0.9011. Mean precision was highest for the model without correction for class imbalance with 0.9404. For the other models it was 0.9178 (oversampling) and 0.8698 (weighted loss) (table 6.2). Below, only the detailed results of the two models with the highest mean recall and precision are presented. For species with <100 images in the dataset, mean recall is lower than for all species (0.7044 for the model without correction of class imbalance and 0.8206 for the model with oversampling). Mean precision for species with <100 images is higher for the model without correction of class imbalance (0.9192) compared to the model with oversampling (0.8803). For species with >5000 images in the datset, recall was 0.9809 and precision 0.9760 for the model without correction of class imbalance and 0.9602 and 0.9726 respectively for the model with oversampling.

Table 6.2: Mean recall and precision over all classes using ResNet-152 with different methods to handle class imbalance during training; none: no correction of class imbalance, oversampling: oversampling of minority classes and undersampling of majority classes, weighted loss: weighted loss function

Handling class imbalance	**Mean Recall**	**Mean Precision**
None	0.9011	0.9404
Oversampling	0.9321	0.9178
Weighted loss	0.9085	0.8698

For both models, the variance in precision and recall per species was higher for species with few images. While some species with few images had high recall and precision, others had low values. For the majority classes, precision and recall were high in both models (fig. 6.3 and 6.4).

For the model without correction for class imbalance, precision was >0.6 for all species while recall was low for some species (6.3). The species with a recall <0.6 are *Boloria aquilonaris* (0.2222) *Agriades orbitulus* (0.2857), *Parnassius phoebus* (0.4444), *Polyommatus thersites* (0.5238), *Agriades optilete* (0.5556) and *Pheangaris alcon* (0.5556). There were 207 images of *Polyommatus thersites* in the dataset and the other mentioned species had <100 images (see table 8.1 in the supplements).

For the model with oversampling of minority classes, all but one species had a precision >0.6 (fig. 6.4). For *Polyommatus thersites* precision was 0.3721. Recall was <0.6 for *Agriades orbitulus* (0.2857), *Parnassius phoebus* (0.5556) and *Phengaris alcon* (0.5556). All of these species are represented by <100 images in the dataset (see table 8.1 in the supplements).

The model without correction for class imbalance predicted the species of minority classes generally less often, than they appear in the dataset while this was not the case for the model with oversampling (fig. 6.5).

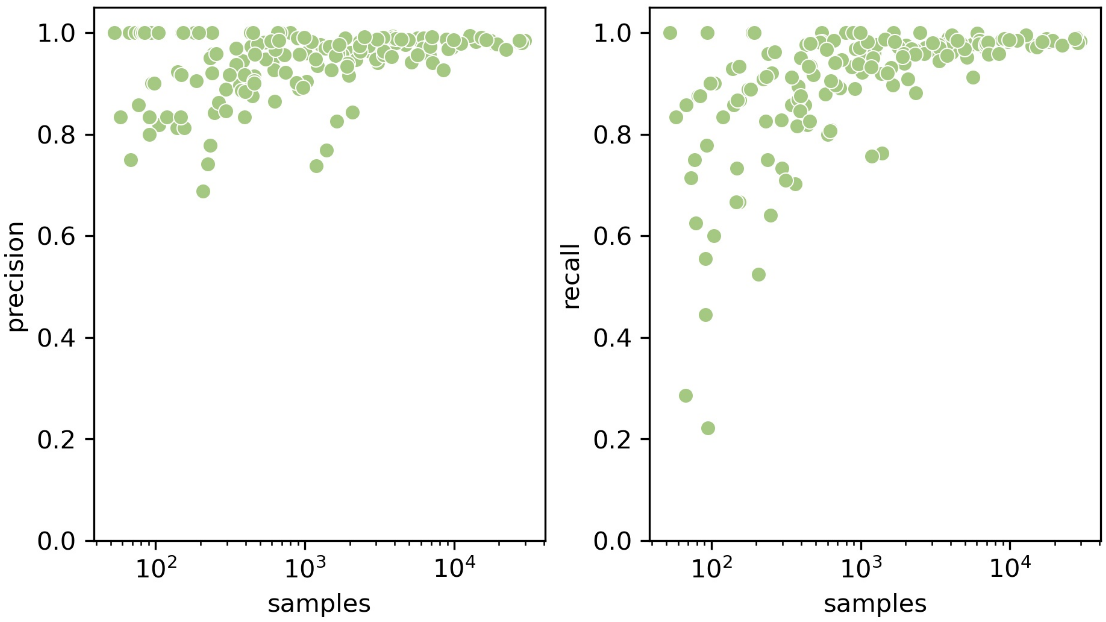

Figure 6.3: Recall and precision for individual species against number of images per species for the model without correction of class imbalance

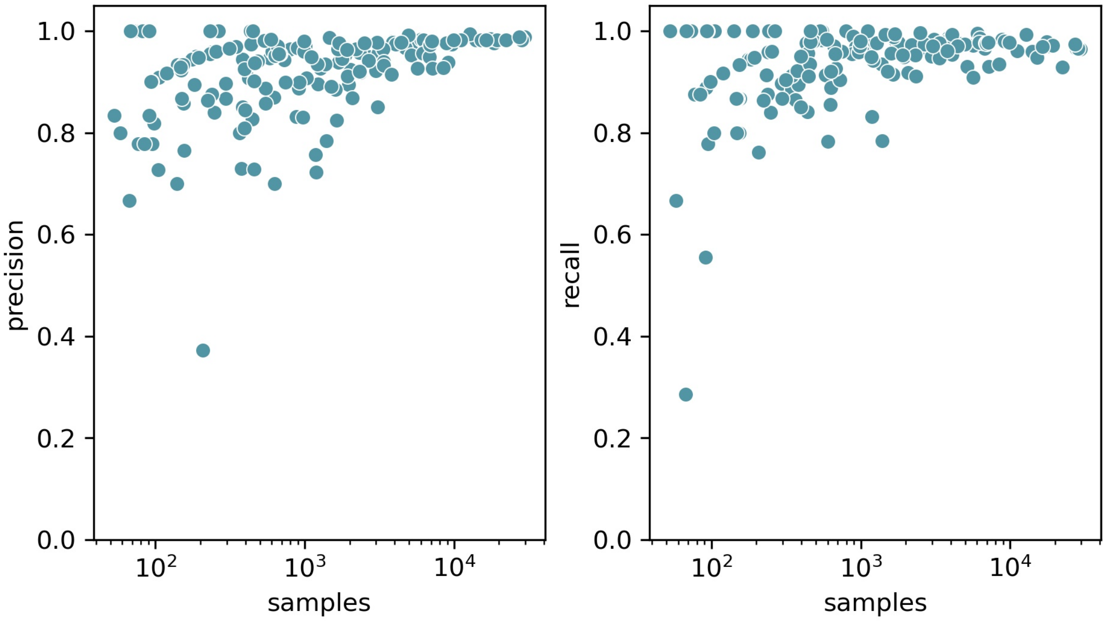

Figure 6.4: Recall and precision for individual species against number of images per species for the model with correction of class imbalance by oversampling minority classes during training

6.3 Performance by species groups

For moth species, recall and precision was high with little variance compared to other species groups for both models. Of the groups of butterflies, the Papilioniadae, Nymphalini and - for the model without correction of imbalance - also the group with Limentidinae and Apaturinae had high median values for recall and precision with little variance. Variance in the two metrics was especially high for the groups Hesperiidae and Lycaenidae. The genus *Erebia* had comparably low precision for both models and low recall for the model without correct for class imbalance (fig. 6.6 and fig. 6.7).

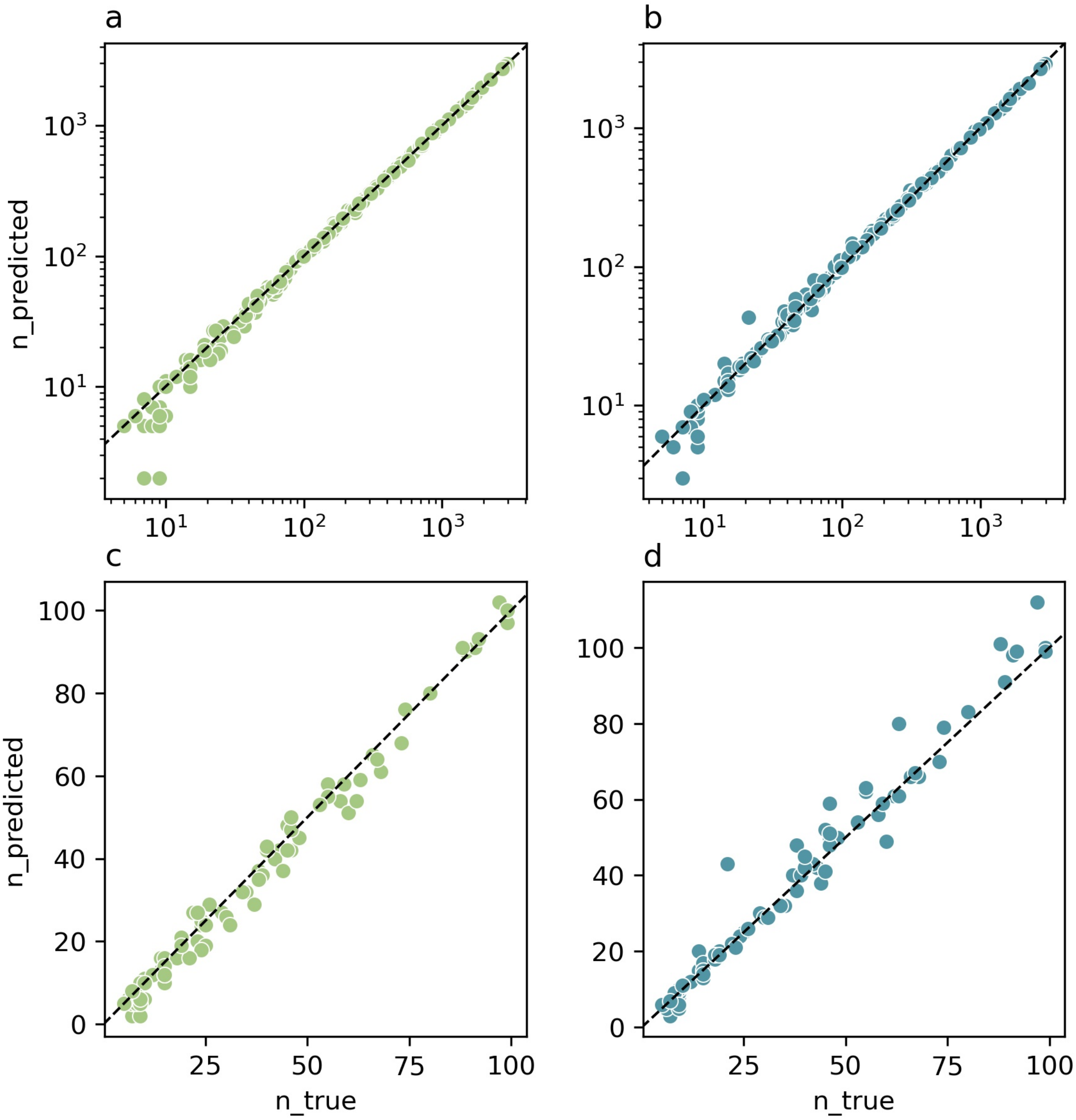

Figure 6.5: Number of images per species in the test dataset (n_true) and number of times it was predicted (n_predicted) with (a, c) model trained without correction of class imbalance and (b, d) model trained with oversampling of minority classes, the lower plots show only the species with up to 100 images

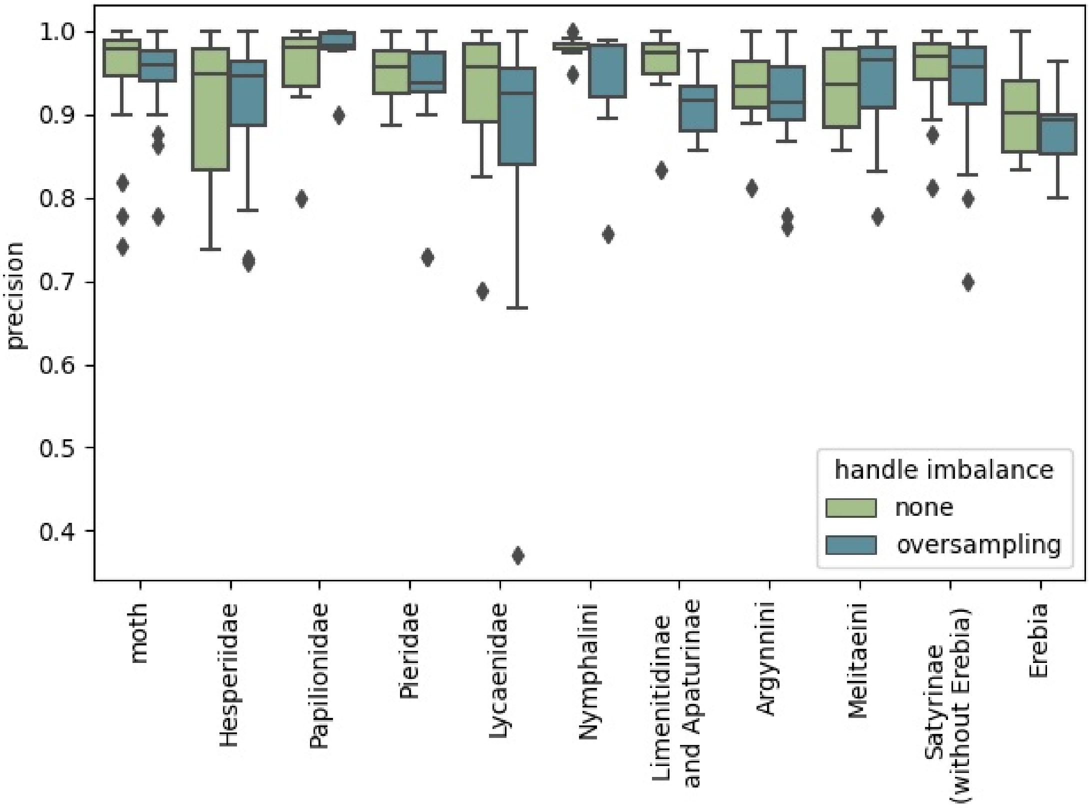

Figure 6.6: Precision for individual species by group; none: model trained without correction of class imbalance; oversampling: model trained with oversampling of minority classes and undersampling of majority classes to correct for class imbalance

6.4 Confidence threshold

For a correct prediction of 99.5 % of the test images, a confidence threshold at 0.9343 would be necessary for the model without correction for class imbalance. When only predictions with a higher confidence are accepted, 92.51 % of the images are covered. For the model with oversampling the confidence threshould is 0.9646 including 89.13 % of the images (6.8). For the model with the weighted loss function, the confidence threshold is 0.9394 at which 85.05 % of the images are included.

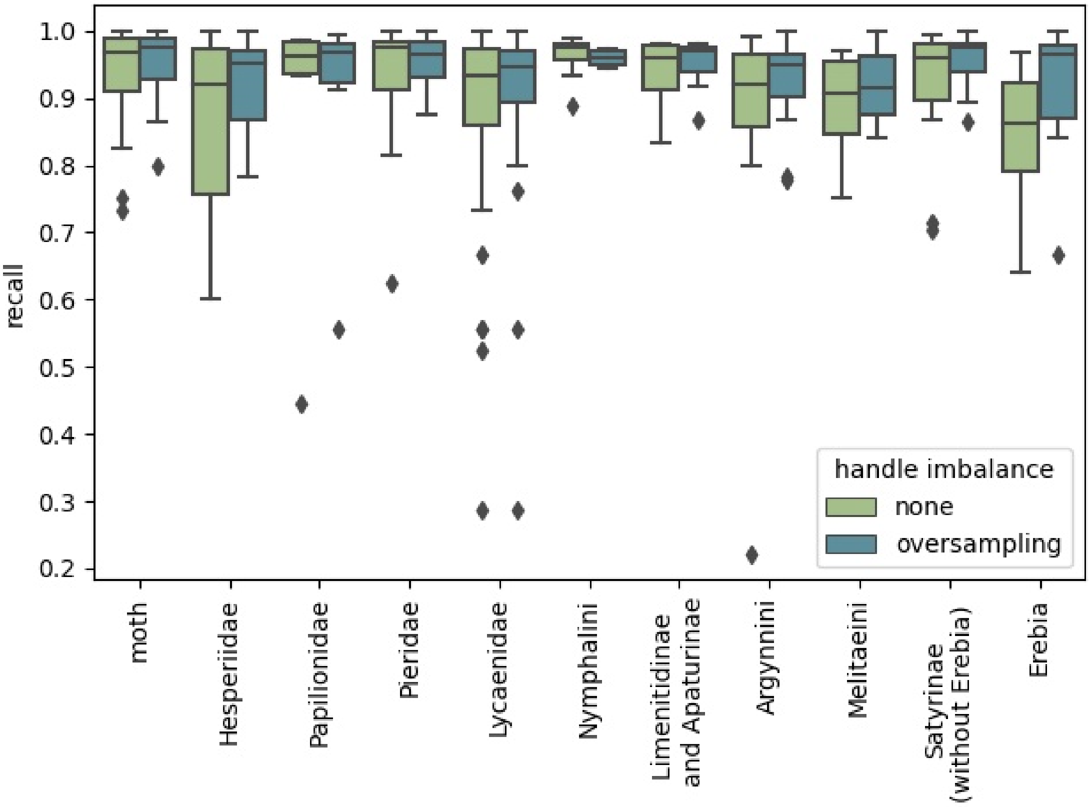

Figure 6.7: Recall for individual species by group; none: model trained without correction of class imbalance; oversampling: model trained with oversampling of minority classes and undersampling of majority classes to correct for class imbalance

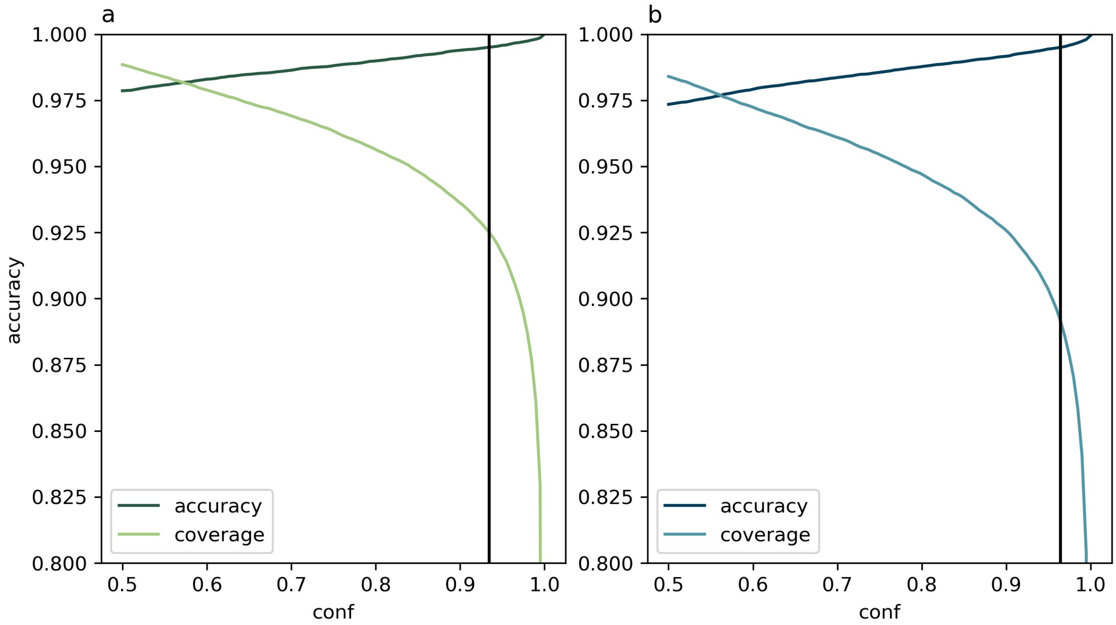

Figure 6.8: Confidence of model predictions and accuracy and coverage that are reached when only accepting predictions above that confidence, the vertical black lines marks the confidence threshold that is needed for an accuracy >0.995; (a) the model trained without correction of class imbalance and (b) the model trained with oversampling of minority classes

7 Discussion

7.1 Model performance

The butterflies and moth species in the dataset could be classified with a high accuracy of >0.97. The accuracy achieved in this study is higher than in other studies on the use of convolutional neural networks for butterfly species identification [48, 38, 10]. The dataset used here by far exceeds the size of datasets used in other studies and offeres unique opportunities to train a convolutional neural network on the fine-grain classification task of identifying butterfly and moth species from images.

7.2 Comparison of approaches

All of the three compared approaches had high top-1 accuracies, with the model with no correction for oversampling performing best. A model that is trained without correction on a highly imbalanced dataset is more strongly optimized for the majority classes [7]. Those species also appear most often in the test dataset which represents the imbalanced distribution of the classes in the original dataset. Therefore, a higher accuracy on the whole dataset was expected. Mean precision over all species was also higher for that approach. When comparing the mean recall over all species though, the model with oversampling of minority classes showed the best performance with a recall >0.93. Differences in recall between the two models were especially pronounced for the minority classes. Due to better representation of the minority classes during training, they were predicted more often by the model with oversampling compared to the model without correction and better represented their number of occurrences in the dataset. When minority classes are better represented in the dataset, there is higher recall (more species that are truly in that class are also predicted to be in it). Whereas species with fewer images still show lower performance, oversampling of minority classes could partly overcome the negative effects of class imbalance with only a small reduction in accuracy on the whole dataset.

7.3 Performance for individual species and groups

Precision and recall were high for species with many observations while performance varied strongly for species with only few images even when accounting for class imbalance by oversampling of minority classes. The generally lower performance for species with only few images could be caused by a lack of sufficient images for learning relevant features for generalization to other images from that species [30]. The high variance in recall and precision for species with only few images shows that the number of images needed for correct identification strongly differs between species. In this context, the collection of images by citizen scientists plays an important role. How often species are photographed not only depends on how commonly they occur, but also on morphological features such as size and wing patterns due to detectability and recognizability bias. The detectability bias describes the well known phenomenon in biodiversity monitoring that some species are detected more easily e.g. due to their larger size, conspicuous colors or active behavior. This affects surveys by experts in systematic monitoring schemes [24] and records made by citizen scientists [3]. The detectability bias is especially important in citizen science monitoring. Koch et al. [29] showed that in datasets collected by citizen scientists, species that are difficult to identify by both humans and machine learning models are less well represented. Therefore, the effect of the number of training images and of morphological features of species are difficult to disentangle.

Model performance differed between species groups. See figures 8.1 to 8.3 for images of some species from each group. The butterfly groups with the highest recall and precision for both models are the Papilionidae and the Nymphalini. The Papillionidae are a small family with only six species in Austria. Most of them such as *Papilio machaon* (Old-World Swallowtail) or *Iphiclides podalirius* (sail swallowtail) can be easily distinguished by their wing patterns. The one species in this group with low recall is *Parnassius phoebus* (Small apollo) which is highly similar to *Parnassius apollo* (Mountain Apollo) but appeares less frequently in the dataset. The Nymphalini are a tribe of the largest family of butterflies, the Nymphalidae. The Nymphalini contain many conspicuous species that have characteristic and unique wing patterns such as *Aglais io* (European peacock), *Vanessa atalanta* (Red Admiral) or *Aglais urticae* (Small tortoiseshell) that make them easy to distinguish and popular for taking pictures.

Butterfly groups with a comparably low recall and high variance in it are the Lycaenidae, Hesperiidae and the *Erebia*. The family Lycaenidae (with the subfamilies Polyommatinae (blues), Lycaeninae (coppers) and Theclinae (hairstreaks) in Austria) contains many species that can be challenging to identify even by an expert. In addition, many species of these group can only be distinguished by patterns on the underside of the wing, that are not always visible in images. In the family Hesperiidae (skippers) especially the genus *Pyrgus* is challenging to identify from images. The same is true for the genus *Erebia* that contains many similar species.

Due to their low number, all moth species were analyzed as one group. Only 31 of the 162 species in the dataset are moth species, even though the majority of the about 4000 species in the order Lepidoptera that occur in Austria [23] are moth species and only about 210 belong to the butterflies (Papilionoidea) [21]. This imbalance can be explained by the selection of only few, mostly day-active or otherwise noticeable moth species for monitoring in the citizen science application "Schmetterlinge Österreichs". The high precision and recall that was reached for moth species is therefore not necessarily representative for this highly diverse group that contains many similar species. Some of the species in the dataset for this study have closely related, similar species that are not in the dataset. An automated detection of all Austrian moth species is therefore very likely a lot more challenging than for the selected species in this study.

The coverage of butterfly species in the dataset is a lot higher than for moth species, but still not all species are represented. This is partly caused by the complete lack of observations for some species and also by the selection of only species with >50 images for this study. Especially for the difficult to distinguish genuses *Erebia* and *Pyrgus* species are missing from the dataset. As for the moths, species identification is probably more difficult for these groups when all species are included. The models usability is limited when an identification of all species that are known to occur in Austria is required due to the absence of some species in the training dataset. Depending on the context, the identification of species that are very rare can be less relevant. A targeted search for images of these species could overcome this limitation in the dataset.

7.4 Practical application

The model trained in this study can identify butterfly species from images with high accuracy. The level of accuracy considered sufficient depends on the application. For educational purposes, errors in the predictions can be less problematic when uncertainties are reported and multiple species with high probabilities are shown. The high top-5 accuracies of >99 % that is achieved with all models are especially relevant in this context.

For use in biodiversity monitoring or research, higher accuracies may be required. Confidence thresholding can increase accuracy by rejecting images with uncertainties of the prediction [56, 39]. To obtain an accuracy of over 99.5 % in this stury only about 10 % of predictions had to be rejected. Automated species identification of butterflies can thus reduce human effort considerably by leaving only difficult cases for human identification. The assessment of precision and recall per species can give insights into quality of prediction for applications on individual species such as habitat suitability models.

For most images in the dataset, a butterfly is clearly visible in the center of the image. Among the images that could not be identified, there are many on which the butterfly can hardly be detected, either because it is relatively small in the image, taken from an unfortunate angle or obscured by objects in the foreground. Missindentification was also caused by important features not being visible in the image and by images of larvae or with more than one species that were not detected during preparation of the dataset. These cases show that guidelines on how to best take images for users of citizen science applications could further improve accuracy of predictions.

Based on the dataset and the methods that were used in this thesis, further studies on automated butterfly species identification are conducted and will be submitted for publication in scientific journals.

Bibliography

[1] Altrudi, S. [2021], 'Connecting to nature through tech? the case of the inaturalist app', *Convergence: The International Journal of Research into New Media Technologies* **27**(1), 124–141.

[2] Anderle, M., Brambilla, M., Angelini, L., Guariento, E., Paniccia, C., Plunger, J., Seeber, J., Stifter, S., Tappeiner, U., Tasser, E. and Hilpold, A. [2024], 'Efficiency of birds as bioindicators for other taxa in mountain farmlands', *Ecological Indicators* **158**, 111569.

[3] Arazy, O. and Malkinson, D. [2021], 'A framework of observer-based biases in citizen science biodiversity monitoring: Semi-structuring unstructured biodiversity monitoring protocols', *Frontiers in Ecology and Evolution* **9**.
URL: *https://www.frontiersin.org/journals/ecology-and-evolution/articles/10.3389/fevo.2021.693602*

[4] Barta, Z. [2023], 'Deep learning in terrestrial conservation biology', *Biologia futura* **74**(4), 359–367.

[5] Besson, M., Alison, J., Bjerge, K., Gorochowski, T. E., Høye, T. T., Jucker, T., Mann, H. M. R. and Clements, C. F. [2022], 'Towards the fully automated monitoring of ecological communities', *Ecology letters* **25**(12), 2753–2775.

[6] Brereton, T., Roy, D. B., Middlebrook, I., Botham, M. and Warren, M. [2011], 'The development of butterfly indicators in the united kingdom and assessments

in 2010', *Journal of Insect Conservation* **15**(1-2), 139–151.
URL: *http://dx.doi.org/10.1007/s10841-010-9333-z*

[7] Buda, M., Maki, A. and Mazurowski, M. A. [2018], 'A systematic study of the class imbalance problem in convolutional neural networks', *Neural networks : the official journal of the International Neural Network Society* **106**, 249–259.

[8] Campbell, C. J., Barve, V., Belitz, M. W., Doby, J. R., White, E., Seltzer, C., Di Cecco, G., Hurlbert, A. H. and Guralnick, R. [2023], 'Identifying the identifiers: How inaturalist facilitates collaborative, research-relevant data generation and why it matters for biodiversity science', *BioScience* **73**(7), 533–541.
URL: *https://doi.org/10.1093/biosci/biad051*

[9] Chandler, M., See, L., Copas, K., Bonde, A. M., López, B. C., Danielsen, F., Legind, J. K., Masinde, S., Miller-Rushing, A. J., Newman, G., Rosemartin, A. and Turak, E. [2017], 'Contribution of citizen science towards international biodiversity monitoring', *Biological Conservation* **213**, 280–294. SI:Measures of biodiversity.
URL: *https://www.sciencedirect.com/science/article/pii/S0006320716303639*

[10] Chang, Q., Qu, H., Wu, P. and Yi, J. [2017], Fine-grained butterfly and moth classification using deep convolutional neural networks.
URL: *https://api.semanticscholar.org/CorpusID:46604360*

[11] Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S. and Miao, Y. [2021], 'Review of image classification algorithms based on convolutional neural networks', *Remote Sensing* **13**(22), 4712.

[12] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L. [2009], Imagenet: A large-scale hierarchical image database, *in* '2009 IEEE Conference on Computer Vision and Pattern Recognition', IEEE, pp. 248–255.

[13] Di Cecco, G. J., Barve, V., Belitz, M. W., Stucky, B. J., Guralnick, R. P. and Hurlbert, A. H. [2021], 'Observing the observers: How participants con-

tribute data to inaturalist and implications for biodiversity science', *BioScience* **71**(11), 1179–1188.
URL: *https://doi.org/10.1093/biosci/biab093*

[14] Gerlach, J., Samways, M. and Pryke, J. [2013], 'Terrestrial invertebrates as bioindicators: an overview of available taxonomic groups', *Journal of Insect Conservation* **17**(4), 831–850.

[15] Gomez Villa, A., Salazar, A. and Vargas, F. [2017], 'Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks', *Ecological Informatics* **41**, 24–32.

[16] Goodfellow, I., Bengio, Y. and Courville, A. [2016], *Deep Learning*, MIT Press. `http://www.deeplearningbook.org`.

[17] Goodwin, M., Halvorsen, K. T., Jiao, L., Knausgård, K. M., Martin, A. H., Moyano, M., Oomen, R. A., Rasmussen, J. H., Sørdalen, T. K. and Thorbjørnsen, S. H. [2022], 'Unlocking the potential of deep learning for marine ecology: overview, applications, and outlook', *ICES Journal of Marine Science* **79**(2), 319–336.

[18] Gugger, S., Debut, L., Wolf, T., Schmid, P., Mueller, Z., Mangrulkar, S., Sun, M. and Bossan, B. [2022], 'Accelerate: Training and inference at scale made simple, efficient and adaptable.', `https://github.com/huggingface/accelerate`.

[19] Hansen, O. L. P., Svenning, J.-C., Olsen, K., Dupont, S., Garner, B. H., Iosifidis, A., Price, B. W. and Høye, T. T. [2020], 'Species-level image classification with convolutional neural network enables insect identification from habitus images', *Ecology and evolution* **10**(2), 737–747.

[20] He, K., Zhang, X., Ren, S. and Sun, J. [2015], 'Deep residual learning for image recognition'.
URL: *https://arxiv.org/abs/1512.03385*

[21] Höttinger, H. and Pennerstorfer, J. [2005], *Rote Liste der Tagschmetterlinge Österreichs (Lepidoptera: Papilionoidea & Hesperioidea)*, na.

[22] Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K. Q. [2017], Densely connected convolutional networks, *in* '2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)', pp. 2261–2269.

[23] Huemer, P. [2013], *Die Schmetterlinge Österreichs (Lepidoptera): systematische und faunistische Checkliste*, Tiroler Landesmuseen Innsbruck.

[24] Isaac, N. J., Cruickshanks, K. L., Weddle, A. M., Marcus Rowcliffe, J., Brereton, T. M., Dennis, R. L., Shuker, D. M. and Thomas, C. D. [2011], 'Distance sampling and the challenge of monitoring butterfly populations', *Methods in Ecology and Evolution* **2**(6), 585–594.

[25] Johnson, C. N., Balmford, A., Brook, B. W., Buettel, J. C., Galetti, M., Guangchun, L. and Wilmshurst, J. M. [2017], 'Biodiversity losses and conservation responses in the anthropocene', *Science* **356**(6335), 270–275.

[26] Kang, S.-H., Song, S.-H. and Lee, S.-H. [2012], 'Identification of butterfly species with a single neural network system', *Journal of Asia-Pacific Entomology* **15**(3), 431–435.

[27] Kaya, Y., Kayci, L. and Uyar, M. [2015], 'Automatic identification of butterfly species based on local binary patterns and artificial neural network', *Applied Soft Computing* **28**, 132–137.

[28] Kingma, D. P. and Ba, J. [2017], 'Adam: A method for stochastic optimization'.
URL: *https://arxiv.org/abs/1412.6980*

[29] Koch, W., Hogeweg, L., Nilsen, E. B., O'Hara, R. B. and Finstad, A. G. [2023], 'Recognizability bias in citizen science photographs', *Royal Society Open Science* **10**(2), 221063.

[30] Leevy, J. L., Khoshgoftaar, T. M., Bauder, R. A. and Seliya, N. [2018], 'A

survey on addressing high-class imbalance in big data', *Journal of Big Data* **5**(1).

[31] Li, F. and Xiong, Y. [2018], 'Automatic identification of butterfly species based on homsc and glcmoib', *The Visual Computer* **34**(11), 1525–1533.

[32] Li, S., Zhao, Y., Varma, R., Salpekar, O., Noordhuis, P., Li, T., Paszke, A., Smith, J., Vaughan, B., Damania, P. and Chintala, S. [2020], 'Pytorch distributed: Experiences on accelerating data parallel training'.
URL: *https://arxiv.org/abs/2006.15704*

[33] Lotfian, M., Ingensand, J. and Brovelli, M. A. [2021], 'The partnership of citizen science and machine learning: Benefits, risks, and future challenges for engagement, data collection, and data quality', *Sustainability* **13**(14), 8087.

[34] Lukhtanov, V. [2019], 'Species delimitation and analysis of cryptic species diversity in the xxi century', *Entomological Review* **99**, 463–472.

[35] Mäder, P., Boho, D., Rzanny, M., Seeland, M., Wittich, H. C., Deggelmann, A. and Wäldchen, J. [2021], 'The flora incognita app – interactive plant species identification', *Methods in Ecology and Evolution* **12**(7), 1335–1342.

[36] Mattins, R. F., Sarobin, M. V. R., Aziz, A. A. and Srivarshan, S. [2023], 'Object detection and classification of butterflies using efficient cnn and pre-trained deep convolutional neural networks', *Multimedia Tools and Applications* .

[37] Nam, N. T. and Hung, P. D. [2018], Pest detection on traps using deep convolutional neural networks, *in* Unknown, ed., 'Proceedings of the 2018 International Conference on Control and Computer Vision - ICCCV '18', ACM Press, New York, New York, USA, pp. 33–38.

[38] Nie, L., Wang, K., Fan, X. and Gao, Y. [2017], Fine-grained butterfly recognition with deep residual networks: A new baseline and benchmark, *in* '2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)', IEEE, pp. 1–7.

[39] Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C. and Clune, J. [2018], 'Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning', *Proceedings of the National Academy of Sciences of the United States of America* **115**(25), E5716–E5725.

[40] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J. and Chintala, S. [2019], Pytorch: An imperative style, high-performance deep learning library, *in* 'Advances in Neural Information Processing Systems 32', Curran Associates, Inc., pp. 8024–8035.
URL: *http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf*

[41] Perez, L. and Wang, J. [n.d.], 'The effectiveness of data augmentation in image classification using deep learning'.
URL: *http://arxiv.org/pdf/1712.04621v1*

[42] Schlegel, J. and Rupf, R. [2010], 'Attitudes towards potential animal flagship species in nature conservation: A survey among students of different educational institutions', *Journal for Nature Conservation* **18**(4), 278–290.

[43] Schneider, S., Greenberg, S., Taylor, G. W. and Kremer, S. C. [2020], 'Three critical factors affecting automated image species recognition performance for camera traps', *Ecology and evolution* **10**(7), 3503–3517.

[44] Sharma, S. and Gosain, A. [2025], 'Addressing class imbalance in remote sensing using deep learning approaches: a systematic literature review', *Evolutionary Intelligence* **18**(1), 1–28.

[45] Shorten, C. and Khoshgoftaar, T. M. [2019], 'A survey on image data augmentation for deep learning', *Journal of Big Data* **6**(1).

[46] Simonyan, K. and Zisserman, A. [2015], 'Very deep convolutional networks for large-scale image recognition'.
URL: *https://arxiv.org/abs/1409.1556*

[47] Strohmaier, E., Dongarra, J., Simon, H. and Meuer, M. [n.d.], 'Top 500 - the list'.
URL: *https://www.top500.org/*

[48] Theivaprakasham, H. [2021], 'Identification of indian butterflies using deep convolutional neural network', *Journal of Asia-Pacific Entomology* **24**(1), 329–340.

[49] Theobald, E. J., Ettinger, A. K., Burgess, H. K., DeBey, L. B., Schmidt, N. R., Froehlich, H. E., Wagner, C., HilleRisLambers, J., Tewksbury, J., Harsch, M. A. et al. [2015], 'Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research', *Biological Conservation* **181**, 236–244.

[50] Thomas, J. A. [2005], 'Monitoring change in the abundance and distribution of insects using butterflies and other indicator groups', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **360**(1454), 339–357.

[51] Turisini, M., Cestari, M. and Amati, G. [2024], 'Leonardo', *Journal of large-scale research facilities JLSRF* **9**(1).

[52] van Swaay, C., Warren, M. and Loïs, G. [2006], 'Biotope use and trends of european butterflies', *Journal of Insect Conservation* **10**(2), 189–209.
URL: *https://link.springer.com/article/10.1007/s10841-006-6293-4*

[53] Wäldchen, J. and Mäder, P. [2018], 'Machine learning for image based species identification', *Methods in Ecology and Evolution* **9**(11), 2216–2225.

[54] Wang, W. and Yang, Y. [2019], 'Development of convolutional neural network and its application in image classification: a survey', *Optical Engineering* **58**(04), 1.

[55] Wang, Z., Cui, J. and Zhu, Y. [2023], ‘Review of plant leaf recognition’, *Artificial Intelligence Review* **56**(5), 4217–4253.

[56] Willi, M., Pitman, R. T., Cardoso, A. W., Locke, C., Swanson, A., Boyer, A., Veldthuis, M. and Fortson, L. [2019], ‘Identifying animal species in camera trap images using deep learning and citizen science’, *Methods in Ecology and Evolution* **10**(1), 80–91.

[57] Zentraler Informatikdienst der Universität Innsbruck [n.d.], ‘Leo5 - introduction and overview’.
URL: *https://www.uibk.ac.at/zid/systeme/hpc-systeme/leo5/leo5-intro.html*

8 Supplements

The supplements contain

- Figure 8.1 to 8.3: Images of some of the species of each of the groups that were used to assess model performance
- Figure 8.4: Random selection of wrong predictions from the model with oversampling.
- Figure 8.5: Random selection of original images from the dataset
- Figrue 8.6: Augmented versions of the randomly selected images shown in 8.2
- Table 8.1: List with all butterfly and moth species that were used for training the model with the number of images and model performance on test data for each species

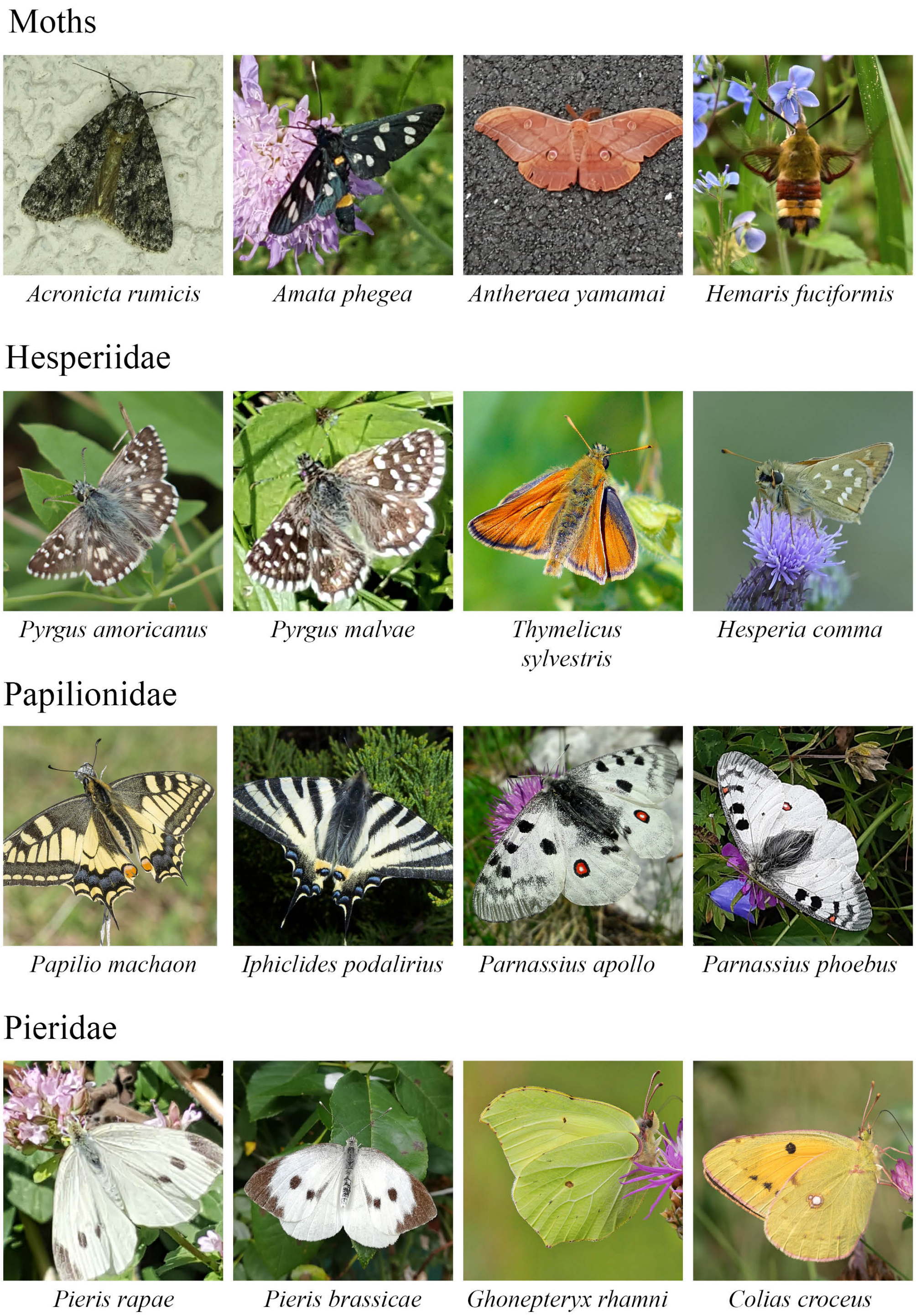

Figure 8.1: Images of some of the species of the groups moths, Hesperiidae, Papilionidae and Pieridae

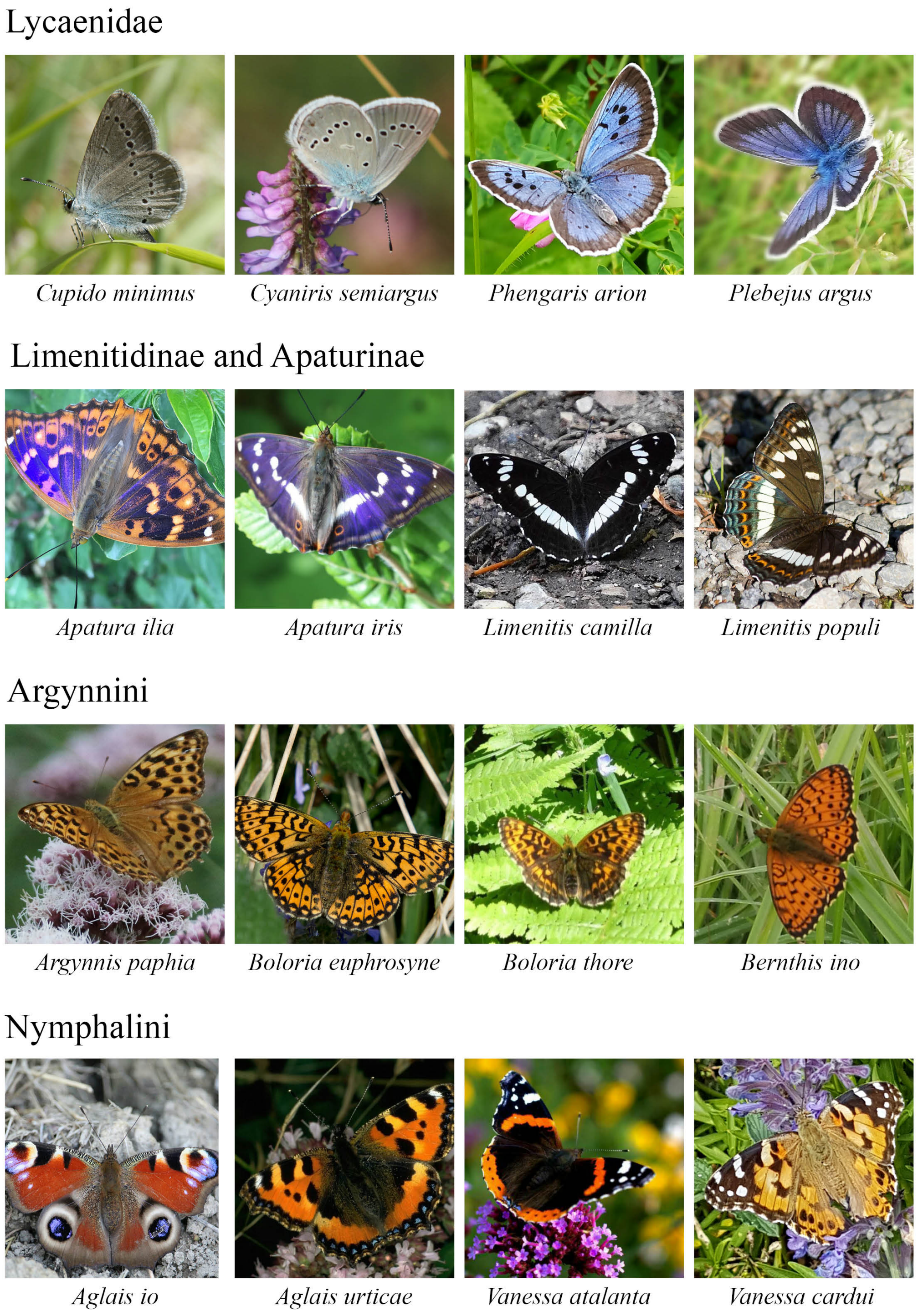

Figure 8.2: Images of some of the species of the groups Lycaenidae, Limenitidinae and Apturinae, Argynnini and Nymphalini

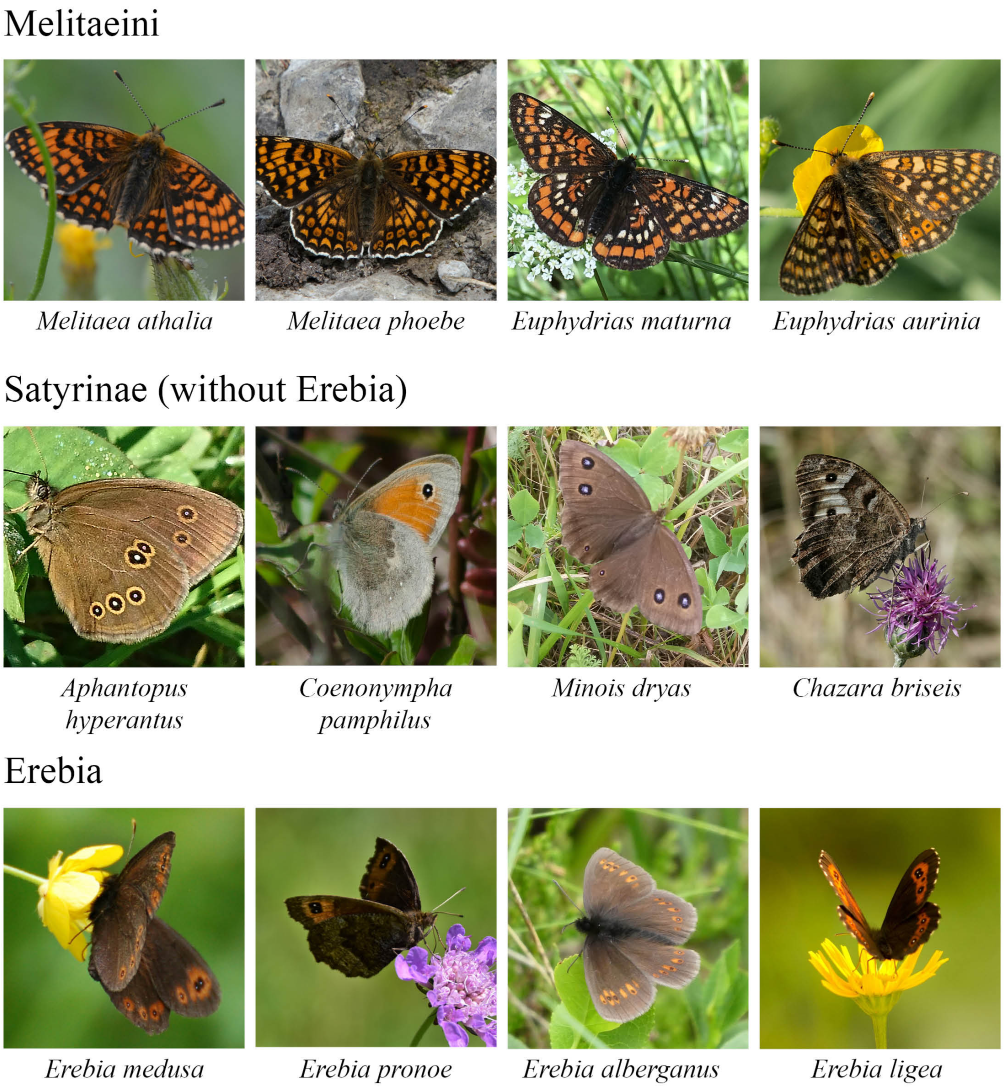

Figure 8.3: Images of some of the species of the groups Melitaeini, Satyrinae (without *Erebia*) and *Erebia*

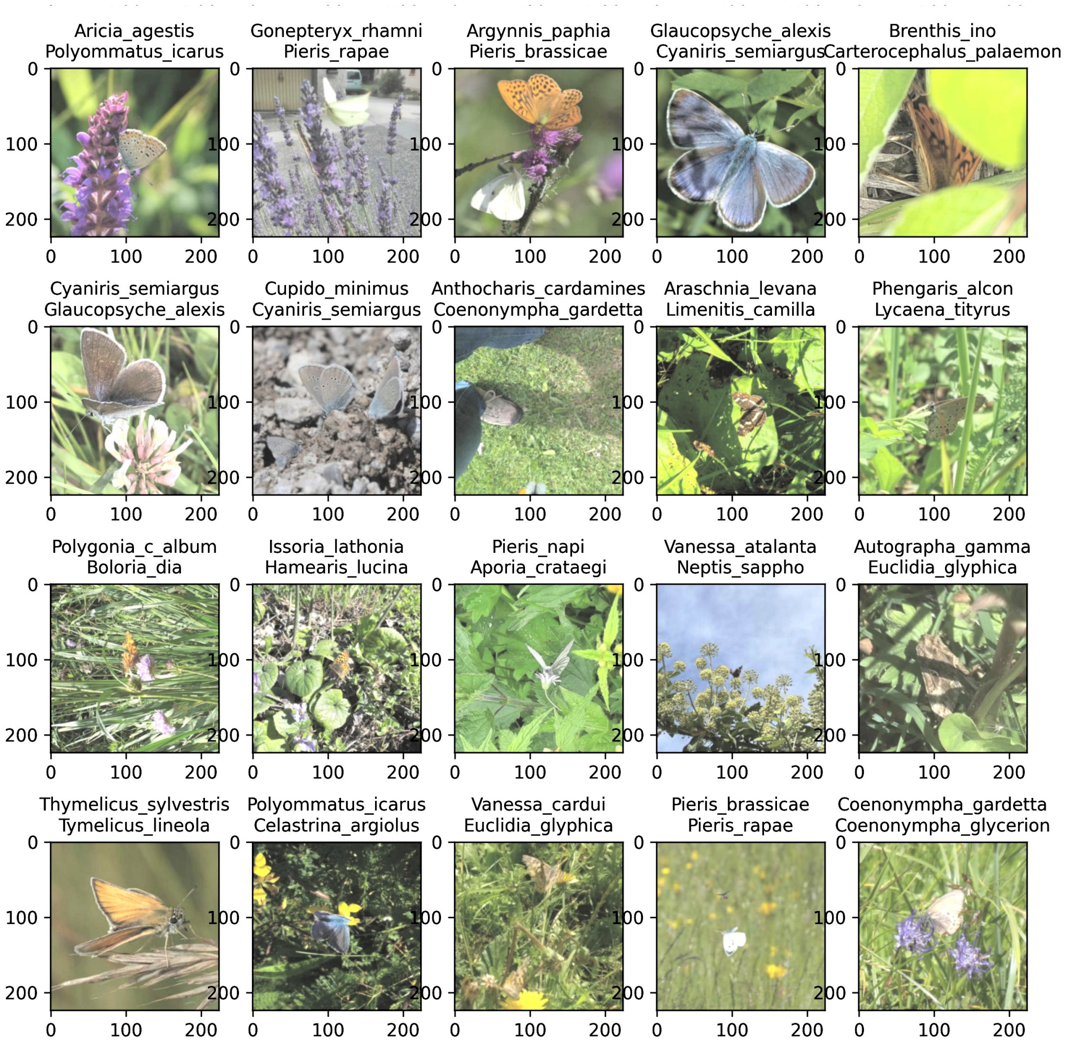

Figure 8.4: Random selection of wrong predictions from the model with oversampling. Upper text: correct species, lower text: predicted species

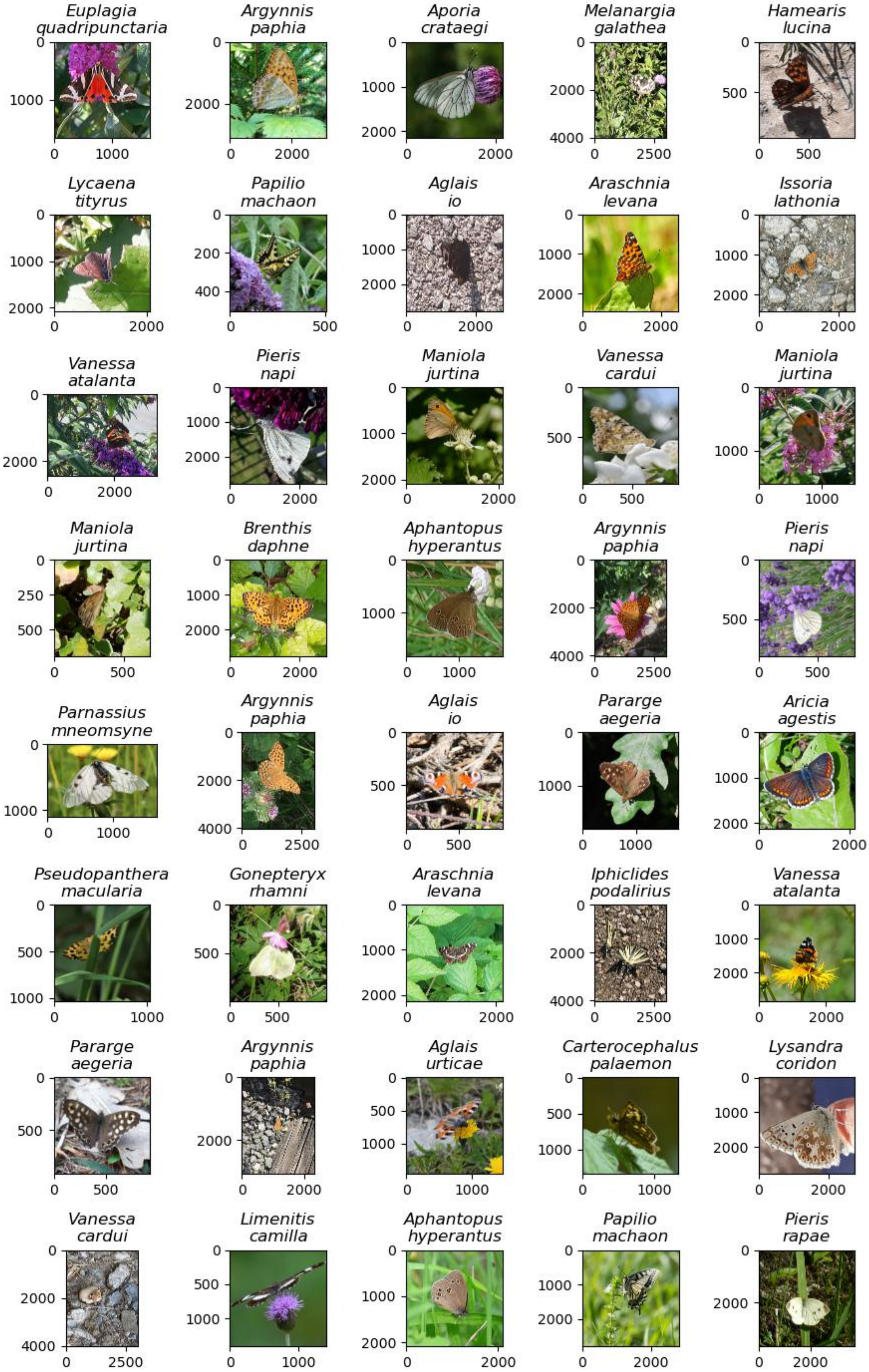

Figure 8.5: Random selection of original images from the dataset

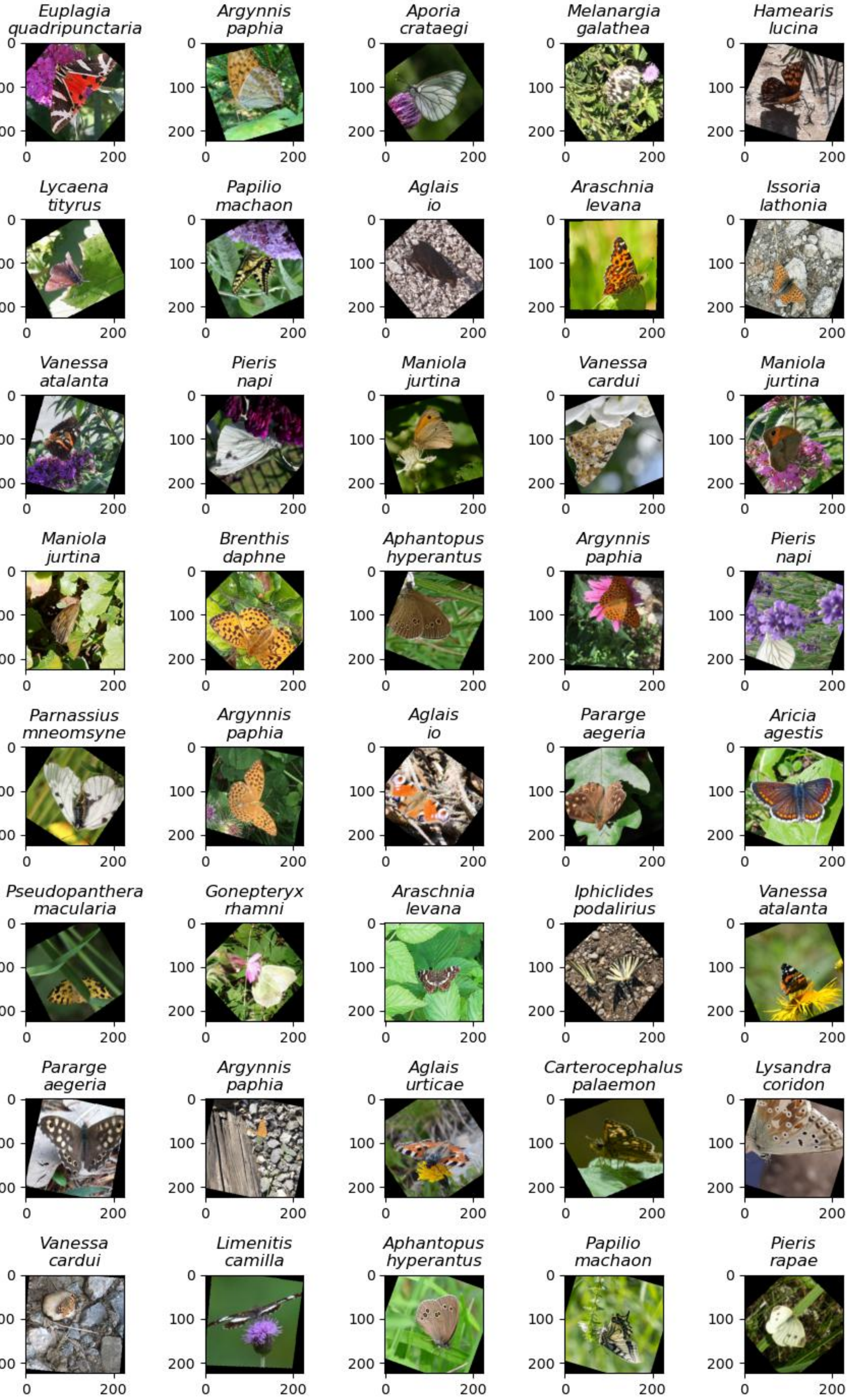

Figure 8.6: Augmented versions of the randomly selected images shown in 8.5

Table 8.1: LList with all butterfly and moth species that were used for training the model with the number of images and model performance on test data for each species, n: number of images of each species (10 % of the images in each class were used for testing the models), None: model trained without correction for class imbalance, Overs.: Model trained with oversampling of minority classes, Recall and Precision were assessed on test data

Species	**n**	**Group**	**None Recall**	**None precision**	**Overs. Recall**	**Overs. Precision**
Aglais io	29612	Nymphalini	0.9831	0.9841	0.9649	0.9882
Maniola jurtina	28406	Satyrinae (without Erebia)	0.9803	0.9803	0.9659	0.9821
Argynnis paphia	27625	Argynnini	0.9906	0.9796	0.9656	0.9871
Vanessa atalanta	27268	Nymphalini	0.9875	0.985	0.9736	0.9885
Polyommatus icarus	22356	Lycaenidae	0.9754	0.9672	0.9289	0.9816
Gonepteryx rhamni	19344	Pieridae	0.984	0.9779	0.971	0.9817
Aglais urticae	18432	Nymphalini	0.9821	0.9789	0.9707	0.976
Coenonympha pamphilus	17522	Satyrinae (without Erebia)	0.9874	0.9846	0.9789	0.9845
Vanessa cardui	16423	Nymphalini	0.9817	0.9853	0.9695	0.9815
Polygonia c-album	15197	Nymphalini	0.9711	0.9906	0.9474	0.9823
Araschnia levana	14029	Nymphalini	0.9743	0.982	0.9594	0.9825
Melanargia galathea	12795	Satyrinae (without Erebia)	0.9953	0.9938	0.9922	0.9937
Ochlodes sylvanus	11125	Hesperiidae	0.9847	0.9786	0.9605	0.9825

Species	**n**	**Group**	**None Recall**	**None precision**	**Overs. Recall**	**Overs. Precision**
Pararge aegeria	9882	Satyrinae (without Erebia)	0.9858	0.9858	0.9777	0.9817
Aphantopus hyperantus	9673	Satyrinae (without Erebia)	0.9824	0.9704	0.9762	0.9732
Macroglossum stellatarum	9132	moth	0.988	0.9678	0.9803	0.9382
Iphiclides podalirius	8865	Papilionidae	0.9853	0.9865	0.9831	0.9765
Pieris rapae	8471	Pieridae	0.9587	0.9259	0.9351	0.9274
Pieris napi	7179	Pieridae	0.9568	0.9411	0.9304	0.9265
Papilio machaon	7129	Papilionidae	0.9804	0.9915	0.9776	0.9803
Lycaena plaeas	6933	Lycaenidae	0.9784	0.9727	0.9755	0.9727
Leptidea sinapis/juvernica	6865	Pieridae	0.9753	0.9558	0.9767	0.9491
Anthocharis cardamines	6733	Pieridae	0.9822	0.9721	0.9658	0.9701
Issoria lathonia	6657	Argynnini	0.9715	0.9729	0.976	0.9615
Euclidia glyphica	6525	moth	0.9832	0.9802	0.977	0.9623
Euplagia quadripunctaria	6246	moth	0.984	0.9887	0.9856	0.9825
Lasiommata megera	6188	Satyrinae (without Erebia)	0.9774	0.9696	0.9774	0.9558
Chiasmia clathrata	6021	moth	0.9983	0.9804	0.995	0.9836
Autographa gamma	5671	moth	0.9771	0.9893	0.9718	0.9752
Pieris brassicae	5666	Pieridae	0.9118	0.9556	0.9083	0.9263
Cupido argiades	5152	Lycaenidae	0.9515	0.9423	0.9301	0.9523
Ematurga atomaria	4954	moth	0.9677	0.9856	0.9737	0.9918

Species	**n**	**Group**	**None Recall**	**None precision**	**Overs. Recall**	**Overs. Precision**
Brintesia circe	4709	Satyrinae (without Erebia)	0.9788	0.9767	0.9745	0.9663
Lycaena dispar	4423	Lycaenidae	0.9842	0.9864	0.9706	0.9772
Celastrina argiolus	4074	Lycaenidae	0.9607	0.9799	0.9533	0.9773
Colia croceus	4068	Pieridae	0.9951	0.9878	0.9926	0.9735
Amata phegea	3873	moth	0.9897	0.9922	0.9871	0.9795
Coenonympha glycerion	3870	Satyrinae (without Erebia)	0.9793	0.9896	0.9819	0.9769
Minois dryas	3795	Satyrinae (without Erebia)	0.9553	0.9528	0.9605	0.9148
Lysandra coridon	3374	Lycaenidae	0.9436	0.9636	0.9466	0.9327
Lycaena tityrus	3364	Lycaenidae	0.9702	0.956	0.9762	0.9398
Erynnis tages	3331	Hesperiidae	0.976	0.9909	0.97	0.9642
Coenonympha arcania	3089	Satyrinae (without Erebia)	0.9806	0.9743	0.9838	0.956
Aricia agestis	3087	Lycaenidae	0.9741	0.9406	0.9773	0.8507
Neptis rivularis	3038	Limenitidinae and Apaturinae	0.9803	0.9868	0.9704	0.9768
Fabriciana adippe	2976	Argynnini	0.9664	0.9474	0.9497	0.9218
Boloria dia	2716	Argynnini	0.9706	0.9635	0.9632	0.9493
Limenitis camilla	2664	Limenitidinae and Apaturinae	0.9586	0.9733	0.9737	0.9418
Pseudopanthera macularia	2506	moth	1	0.9921	0.996	0.9766
Lysandra bellargus	2349	Lycaenidae	0.8809	0.9628	0.9106	0.9185
Carterocephalus palaemon	2322	Hesperiidae	0.9655	0.9825	0.9698	0.9574

Species	**n**	**Group**	**None Recall**	**None precision**	**Overs. Recall**	**Overs. Precision**
Nymphalis antiopa	2317	Nymphalini	0.9569	0.9737	0.9526	0.9208
Erebia aethiops	2197	Erebia	0.9682	0.9467	0.9727	0.964
Cyaniris semiargus	2075	Lycaenidae	0.9087	0.8438	0.9183	0.8682
Apatura ilia	2021	Limenitidinae and Apaturinae	0.9752	0.961	0.9604	0.9238
Boloria euphrosyne	1970	Argynnini	0.9391	0.9158	0.9492	0.8947
Lasiommata maera	1940	Satyrinae (without Erebia)	0.9588	0.9394	0.9485	0.9109
Pyrgus malvae	1906	Hesperiidae	0.9529	0.9333	0.9529	0.963
Callophrys rubi	1853	Lycaenidae	0.9622	0.9889	0.973	0.9524
Carcharodus alceae	1782	Hesperiidae	0.9719	0.9558	0.9775	0.9457
Pontia edusa	1691	Pieridae	0.9822	0.9765	0.9941	0.9767
Aporia crataegi	1687	Pieridae	0.9822	0.9765	0.9822	0.9379
Cydalima perspectalis	1662	moth	1	0.9595	0.988	0.9647
Cupido minimus	1637	Lycaenidae	0.8963	0.8258	0.9146	0.8242
Plebejus argus	1603	Lycaenidae	0.9312	0.9551	0.9562	0.8844
Speyeria aglaja	1506	Argynnini	0.9205	0.9267	0.9205	0.891
Parnassius mneomsyne	1459	Papilionidae	0.9863	0.973	0.9932	0.9864
Thymelicus sylvestris	1390	Hesperiidae	0.7626	0.7681	0.7842	0.7842
Hesperia comma	1373	Hesperiidae	0.9197	0.9692	0.9343	0.9343
Hamearis lucina	1276	none	0.9766	0.9766	0.9766	0.9259

Species	**n**	**Group**	**None Recall**	**None precision**	**Overs. Recall**	**Overs. Precision**
Apatura iris	1229	Limenitidinae and Apaturinae	0.9593	0.9365	0.9756	0.8955
Melitaea diamina	1211	Melitaeini	0.9504	0.935	0.9421	0.9344
Tymelicus lineola	1190	Hesperiidae	0.7563	0.7377	0.8319	0.7226
Nymphalis polychloros	1177	Nymphalini	0.9322	0.9483	0.9492	0.7568
Siona lineata	1110	moth	0.982	0.982	1	0.9487
Brenthis daphne	1024	Argynnini	0.9216	0.9038	0.9608	0.9074
Antheraea yamamai	1011	moth	0.9505	0.96	0.9703	0.9703
Triodia sylvina	991	moth	1	0.99	0.9798	0.9798
Melitaea didyma	986	Melitaeini	0.9697	0.9897	0.9697	0.96
Melitaea athalia	971	Melitaeini	0.9381	0.8922	0.9588	0.8304
Hemaris fuciformis	921	moth	0.9674	0.957	0.9674	0.899
Erebia ligea	911	Erebia	0.8901	0.8901	0.956	0.8878
Diacrisia sannio	889	moth	1	0.9889	0.9888	0.967
Lycaena virgaureae	877	Lycaenidae	0.9318	0.9011	0.9545	0.8317
Callimorpha dominula	798	moth	1	1	1	0.9639
Parnassius apollo	743	Papilionidae	0.9459	0.9211	0.9595	0.8987
Hipparchia semele	726	Satyrinae (without Erebia)	0.8904	0.9559	0.9041	0.9429
Lopinga achine	682	Satyrinae (without Erebia)	0.8971	1	0.9265	0.9545
Thecla betulae	671	Lycaenidae	0.9403	0.9844	0.9552	0.9552

Species	n	Group	None Recall	None precision	Overs. Recall	Overs. Precision
Scolitanides orion	660	Lycaenidae	0.9848	1	0.9697	0.9697
Lycaena hippothoe	633	Lycaenidae	0.9048	0.9661	0.9206	0.9508
Cupido decolorata	629	Lycaenidae	0.8095	0.8644	0.8889	0.7
Glaucopsyche alexis	623	Lycaenidae	0.8065	0.9259	0.8548	0.8689
Brenthis ino	604	Argynnini	0.8	0.9412	0.7833	0.9592
Sytyrium spini	592	Lycaenidae	0.9661	0.9828	0.9831	0.9831
Agrius convolvuli	580	moth	0.8793	0.9444	0.9138	0.9464
Heteropterus morpheus	548	Hesperiidae	1	0.9483	1	0.8871
Neptis sappho	547	Limenitidinae and Apaturinae	0.9818	0.9818	0.9818	0.8571
Phengaris nausithous	534	Lycaenidae	0.9811	0.9811	1	0.9815
Saturnia pyri	482	moth	0.9167	0.9778	0.9792	0.94
Satyrium w-album	460	Lycaenidae	0.9783	0.9574	0.9783	0.9375
Erbia medusa	459	Erebia	0.9348	0.9149	0.9783	0.9
Sphinx pinastri	459	moth	0.9783	0.9	1	0.902
Pieris manii	456	Pieridae	0.8261	0.9048	0.9348	0.7288
Zerynthia polyxena	447	Papilionidae	0.9333	1	0.9111	1
Hipparchia alcyone	445	Satyrinae (without Erebia)	0.9333	0.875	0.9556	0.8269
Melitaea phoebe	438	Melitaeini	0.8182	0.973	0.8409	0.9737
Arctia caja	432	moth	0.9767	1	0.9767	1

Species	**n**	**Group**	**None Recall**	**None precision**	**Overs. Recall**	**Overs. Precision**
Boloria selene	422	Argynnini	0.8571	0.9	0.9286	0.907
Polyommatus daphnis	397	Lycaenidae	0.95	0.8837	0.95	0.8444
Phengaris teleius	396	Lycaenidae	0.875	0.8333	0.85	0.8095
Phengaris arion	393	Lycaenidae	0.8462	0.9167	0.9487	0.925
Coenonympha gardetta	383	Satyrinae (without Erebia)	0.8947	0.9444	0.8947	0.9444
Hipparchia fagi	382	Satyrinae (without Erebia)	0.8684	0.8919	0.8947	0.85
Pieris bryoniae	375	Pieridae	0.8158	0.8857	0.9211	0.7292
Lasiommata petropolitana	366	Satyrinae (without Erebia)	0.7027	0.8966	0.8649	0.8
Mimas tiliae	345	moth	0.9118	0.9688	0.9118	0.9688
Euphydrias aurinia	345	Melitaeini	0.8571	0.9375	0.8857	0.9688
Pyrgus amoricanus	313	Hesperiidae	0.7097	0.9167	0.9032	0.9655
Polyommatus dorylas	296	Lycaenidae	0.7333	0.8462	0.8667	0.8966
Boloria titania	295	Argynnini	0.8276	0.8889	0.8966	0.8667
Melitaea cinxia	265	Melitaeini	0.9615	0.8621	1	1
Deilephila elpenor	254	moth	0.92	0.9583	0.96	0.96
Erebia euryale	248	Erebia	0.64	0.8421	0.84	0.84
Boloria eunomia	243	Argynnini	0.9583	0.9583	1	1
Calliteara pudibunda	239	moth	0.9583	0.92	0.9583	0.9583
Laothoe populi	238	moth	0.75	1	0.875	0.875

Species	**n**	**Group**	**None Recall**	**None precision**	**Overs. Recall**	**Overs. Precision**
Sphinx ligustri	233	moth	0.913	0.7778	0.913	1
Cossus cossus	231	moth	0.8261	0.95	0.913	0.9545
Smerinthus ocellata	223	moth	0.9091	0.7407	0.8636	0.8636
Polyommatus thersites	207	Lycaenidae	0.5238	0.6875	0.7619	0.3721
Satyrium pruni	194	Lycaenidae	1	1	0.9474	0.9474
Acherontia atropos	188	moth	1	0.9048	1	0.95
Libythea celtis	183	Nymphalini	0.8889	1	0.9444	0.8947
Favonius quercus	176	moth	0.8889	1	0.9444	0.9444
Fabriciana niobe	155	Argynnini	0.8667	0.8125	0.8667	0.7647
Eumedonia eumedon	154	Lycaenidae	0.6667	1	0.8	0.8571
Satyrium acaciae	153	Lycaenidae	0.9333	1	0.9333	0.9333
Limenitis populi	150	Limenitidinae and Apaturinae	0.8667	1	0.8667	0.8667
Phalera bucephala	148	moth	0.7333	0.9167	0.8	0.9231
Spialia sertorius	147	Hesperiidae	0.6667	0.8333	0.8667	0.9286
Boloria thore	141	Argynnini	0.8571	0.9231	1	0.9333
Arethusana arethusa	139	Satyrinae (without Erebia)	0.9286	0.8125	1	0.7
Limenitis reducta	119	Limenitidinae and Apaturinae	0.8333	0.8333	0.9167	0.9167
Acronicta rumicis	105	moth	0.9	0.8182	1	0.9091
Muschampia floccifera	104	Hesperiidae	0.6	1	0.8	0.7273

Species	**n**	**Group**	**None Recall**	**None precision**	**Overs. Recall**	**Overs. Precision**
Lycaena helle	98	Lycaenidae	0.9	0.9	0.9	0.8182
Boloria aquilonaris	95	Argynnini	0.2222	1	0.7778	0.7778
Colias phicomone	94	Pieridae	1	0.9	1	0.9
Erebia manto	93	Erebia	0.7778	1	1	0.9
Agriades optilete	92	Lycaenidae	0.5556	0.8333	0.8889	1
Parnassius phoebus	91	Papilionidae	0.4444	0.8	0.5556	1
Phengaris alcon	91	Lycaenidae	0.5556	0.8333	0.5556	0.8333
Macrothylacia rubi	84	moth	0.875	1	0.875	0.7778
Euphydrias maturna	82	Melitaeini	0.875	1	0.875	1
Colias palaeno	79	Pieridae	0.625	1	0.875	1
Euphydryas cynthia	77	Melitaeini	0.75	0.8571	0.875	0.7778
Chazara briseis	73	Satyrinae (without Erebia)	0.7143	1	1	1
Pyrgus carthami	68	Hesperiidae	0.8571	0.75	1	1
Agriades orbitulus	67	Lycaenidae	0.2857	1	0.2857	0.6667
Erebia pronoe	58	Erebia	0.8333	0.8333	0.6667	0.8
Polyommatus damon	53	Lycaenidae	1	1	1	0.8333