A Novel use of Spatiotemporal Metadata to Bolster Wildlife Classification

Travis Dawson
Department of Advanced Computing Sciences
Faculty of Science and Engineering
Maastricht University
Maastricht, The Netherlands

***Abstract*—Wildlife image classification operating on the combination of global social media coverage, citizen-science, and camera-trap networks could potentially provide an unparalleled global real-time form of wildlife observation, enabling conservation efforts and ecological studies. The current state of wildlife classification is required to deal with issues such as dataset long-tail distributions, and variable data quality due to the harsh and remote conditions inhabited by wildlife.**

A novel wildlife dataset is created to facilitate the study. The dataset encompasses observations of the *Felidae* and *Elephantidae* taxonomic families, as well as extensive metadata forming a spatiotemporal snapshot of the environment describing each observation. The resulting dataset acts as a generic representative encompassing the issues underlying wildlife classification.

We investigate the performance of image and metadata classifiers within the taxonomic structure. The study of metadata classification within the taxonomic structure is considered to be a novel contribution of the study. It determines an increasing classification performance with decreasing taxonomic levels, achieving a 90% balanced accuracy at the subspecies level using an XGBoost classifier. The performance of the image classifier using the EfficientNet-B6 Convolutional Network (CNN) architecture within the taxonomic structure, determines an opposing trend of decreasing classification performance with decreasing taxonomic levels due to the high inter-species visual similarity and limited training data at lower taxonomic levels. These findings form the foundation of the proposed novel cascading ensemble (*CE*) classifier, which displays an accuracy improvement of up to 10 times that of traditional image classification on the proposed dataset.

***Index Terms*—wildlife classification, convolutional neural network, CNN, metadata, spatiotemporal, automated species identification, taxonomy, images, XGBoost, cascading classifier, ensemble classifier**

I. Introduction

Wildlife classification has become an essential monitoring, observation, and research tool, allowing us an unparalleled glimpse into the real-time state of global wildlife populations. It is utilized across a variety of domains, ranging from conservation management as a tool to monitor population sizes and behaviour [2], [15], [24], [27], [32] through to ecological studies pursuing knowledge of the natural world [16].

This thesis was prepared in partial fulfilment of the requirements for the Degree of Bachelor of Science in Data Science and Artificial Intelligence, Maastricht University. Supervisor(s): Christof Sieler, Mirela Popa

The primary use of wildlife classification is the removal of the bottleneck surrounding image labelling, saving researchers and organizations time and funding [2], [15]. The presence of social media content, camera-trap networks, remote monitoring, and citizen-science platforms generate a global network of monitoring and observation potential. The use of an accurate wildlife classification system could potentially enable a cost-effective near real-time form of global wildlife observation, reaching to the most remote corners of the globe.

Multiple prominent issues impede the progress of robust wildlife classification, specifically varied image data quality, unbalanced data resulting in a long-tail distribution, and high inter-species visual similarity. A robust classifier should be capable of utilizing all available information in order to accurately predict the wildlife class despite any present issues.

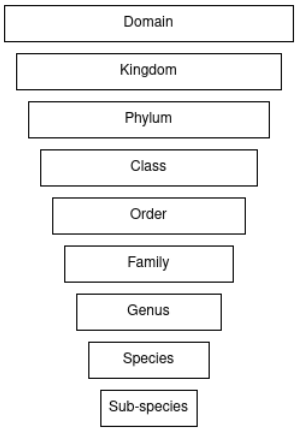


Fig. 1: Taxonomic Classification

The term observation is used to describe each individual sighting of wildlife and the associated information. Modern cameras capturing observations enable the collection of metadata associated with the image. Common metadata includes: date, time, and location. Using the common metadata to extract spatiotemporal features from external sources produces a snapshot of the immediate environment of the observation. This snapshot can be leveraged to inform the image classification process, assuming an underlying relation between wildlife and their environment. Existing studies have

successfully investigated the advantage of using metadata to inform classification [29], [30].

This study, further investigates the use of the taxonomic structure (Figure 1) underlying wildlife, as a means of structuring the classification process as a cascading tree classifier. This structure generates successive levels of abstraction aligned with each taxonomic level. An existing study has demonstrated an improvement when using the taxonomy tree in classification [6].

Specifically, this study investigates the performance of both image and metadata classifiers within the taxonomic tree structure, to leverage their potential strengths and mitigate their potential weaknesses to form a novel cascading ensemble classification method. This method is capable of classifying observations to the subspecies taxonomic level. This can be encapsulated in the formulation of three research questions:

1) How does taxonomic level influence the performance of metadata classification?
2) How does taxonomic level influence the performance of image classification?
3) How does the proposed novel classification method improve upon baseline classifiers?

The novel contributions of this study to the field of wildlife classification include:

1) The collection and generation of a new dataset containing both labelled images and accurate metadata.
2) A study of metadata based models through the taxonomic levels.
3) A novel cascading ensemble wildlife classification model.

II. Related Works

A. Datasets

Wildlife images are currently sourced primarily from citizen-science platforms and camera-trap networks. Citizen-science platforms enable members of the public to upload captured observations of wildlife, where through group consensus the observation is labelled. Two prominent platforms, iNaturalist [10] and Zooniverse [34], respectively containing 139 million and 250 million observations encompassing all domains of life. The dataset within this study made use of a subset of the observations from iNaturalist.

A selection of iNaturalist images form the iNaturalist challenge [31], providing a baseline dataset to push the boundaries of nature image classification.

Camera-trap networks are a set of fixed-location, strategically placed cameras, that capture a burst of images when a motion sensor is triggered by wildlife. The largest camera-trap network to date, is the Snapshot-Safari network [17]. It stretches across 25 distinct networks around the globe and contains 9 million observations.

The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [22], is an annual image classification challenge based on a dataset encompassing 1000 common classes. Classes include wildlife such as, elephants, snakes, and crocodiles amongst others. This challenge serves as a benchmark against which state-of-the-art image classifiers can be compared.

B. Image Classification

The results of the ILSVRC challenge provide a historical record of image classification performance, including current state-of-the-art image classifiers and their benchmark performance. Convolutional Neural Network (CNN) models provided the initial success in image classification. AlexNet [11] was created for the 2012 ILSVRC challenge achieving a 39.7% top-1 error-rate. In 2013 the VGG-16 [26] model achieved a 24.7% top-1 error-rate surpassing AlexNet. The capability of CNN models continued to grow with the creation of ResNet [8] and EfficientNet [28] of which the latter achieved a 88.4% top-1 accuracy. The performance of CNN models had now surpassed human recall levels [32]. Deviations from the CNN architecture into the realm of vision transformer models such as ViT-G/14 [33], has achieved accuracy as high as 90% top-1 accuracy.

Gomez, Salazar, & Vargas, 2017 [7] present a study comparing the performance of multiple CNN architectures on the Snapshot Serengeti dataset (a camera-trap network of Snapshot-safari). The study determined a 88.8% top-1 accuracy using a residual network (ResNet).

C. Image Classification Using Metadata

Tang, Paluri, Fei-fei, Fergus, & Bourdev, 2015 [29] explored the use of location metadata as a means of improving existing geospatial image classification. The study determined an almost 7% mean accuracy improvement.

Ellen, Graff, & Ohman, 2019 [5] further investigated the use of metadata to support image classification within the aquatic domain. The metadata included geotemporal (depth, location, time, etc.) and hydrographic (temperature, salinity, chlorophyll α, etc.) features. Multiple methods of metadata inclusion were investigated, including simple concatenation, metadata interaction, and more metadata interaction models. The study noted a boosted classification performance, with the top-model achieving 93% accuracy within a 27 class dataset.

Terry, Roy, & August, 2019 [30] study the effect of metadata on the classification of ladybirds (*Coccinellidae*) within the United Kingdom. The study made use of weather records, species distribution maps, habitat information, and phenology records amongst others. They made a distinction between two types of metadata: primary metadata, which contains the information included in the original sighting, and derived (secondary) metadata, which contains information further extracted using the primary metadata. The study determined an improvement of 9.1% top-1 accuracy. Notably, the study utilized select model architectures introduced by Ellen et al., 2019 [5], concluding that both the ensemble (metadata

interaction) and combined (simple concatenation) architectures demonstrate improved performance over standalone image classification, with the combined architecture demonstrating the most significant improvement.

D. Classification Using Taxonomic Structure

Carlos & Silla, 2011 [25] conducted a survey of hierarchical classification methods across domains. They provide a framework of hierarchical classification approaches, and a discussion of each methodology. Hierarchical classification is modelled typically as a tree structure or a Directed Acyclic Graph (DAG). The study describes the flat-classifier approach, the local classifier per node approach, the local classifier per parent node approach, the local classifier per level approach, and the global classifier approach. Elaborating further on the flat-classifier, local classifier per parent node, and the global approach due to their relevance in this study. The flat-classifier approach is the standard classification approach classifying the leaf nodes of the tree. The local classifier per parent node approach (selective classifier), assigns a classifier to each parent node (including the root) in order to classify/ select each child node at each respective depth, forming a cascade of classification. The global classifier approach aims to classify the entire tree across all depths and nodes.

Gomez-donoso, Escalona, Pérez-estev, Cazorla, 2021 [6] explored the capability of a modified global hierarchical approach to classify the iNaturalist [31] dataset. They achieved a taxonomic multi-level classification (global approach) by utilizing an EfficientNet [28] convolutional backbone and modifying the output into 7 parallel, fully connected layers, each tasked with classifying a select taxonomic level. The model architecture resulted in a 62% top-1 accuracy. Notably Gomez-donoso et al., 2021 state they expect an accuracy boost if a cascade classification structure is utilized, however that the computation costs involved make this infeasible for any real application use.

Rezende, Xavier, Ascher, Fernandes, & Pirez, 2022 [20] explored the framework provided by Carlos & Silla, 2011 applied to biological databases, specifically those following a taxonomic structure. Notably they specified the issues related to hierarchical classification, namely balancing partial or full depth prediction, deep tree levels/ classification, unbalanced classes, and a high number of classes. A parallel can be drawn to the issues underlying wildlife classification and the issues face by Rezende et al., 2022 [20].

III. Dataset

The dataset is comprised of observations sourced from the citizen-science platform iNaturalist [10]. The observations contain wildlife images and primary metadata. Secondary metadata in the form of spatiotemporal data is derived from Open-Meteo [18], an open source weather Application Programming Interface (API). The combination of the wildlife observations and spatiotemporal snapshots describing each observation forms a novel dataset [4]

A. Observations

The available images on the platform originate from a combination of citizen-science and camera-traps. Due to the time and resource limitations of the study, a subset of the available iNaturalist data is retrieved. This serves as a generic representative of the issues encountered within wildlife classification. Specifically, the taxonomic families of the *Animalia* kingdom, *Elephantidae* and *Felidae* were selected to form the data subset. This subset encompasses a diverse global population inhabiting both accessible and remote regions of the globe, as seen in Figure 2. Additionally, it is characterized by a long-tail population distribution, fulfilling the role of a generic wildlife dataset. The resulting set of observations spans 2 taxonomic families, 16 taxonomic genera, 48 taxonomic species, and 67 taxonomic subspecies. For a comprehensive breakdown of each taxonomic level please review Tables II, III, IV, and V in the Appendix (section IX).

The primary metadata of the observation includes the date and time, geographic location, positional accuracy, and the relevant taxonomic labels for the family, genus, and species. Note, that for the subspecies taxonomic label, iNaturalist [10] has the capability to provide a subspecies label in the data download. However within this dataset the subspecies label is derived from the provided common name, as the potential for subspecies classification became apparent late in the study. As a result, the subspecies labels are not present for all observations.

Additionally, common house hold cats—*Felis catus*—are present within the observations. All observations from this class are removed from the dataset to maintain wildlife only classes.

The location of endangered wildlife species is obscured from poachers by iNaturalist. A random location within a 22 km^2 area surrounding the original location is assigned as a replacement. The publicly available positional accuracy metadata value is effected by the obfuscation. All observations with positional accuracy's below 40 kilometers are accepted, under the assumption that relevant critical geographic and environmental conditions are captured due to the scale of weather states and geographic features.

The dataset contains mislabelled and erroneous images that requiring pre-processing before use within a classification model. Section IV-B1 provides a detailed description of the pre-processing steps required to generate the final observation images within the dataset. The final images show variable quality and resolution as a result of the pre-processing, providing a challenging dataset for traditional image classification.

B. Environmental Data

Secondary metadata in the form of environmental descriptors are derived from the observation's primary metadata, similarly to the study conducted by Terry et at., 2019 [30].

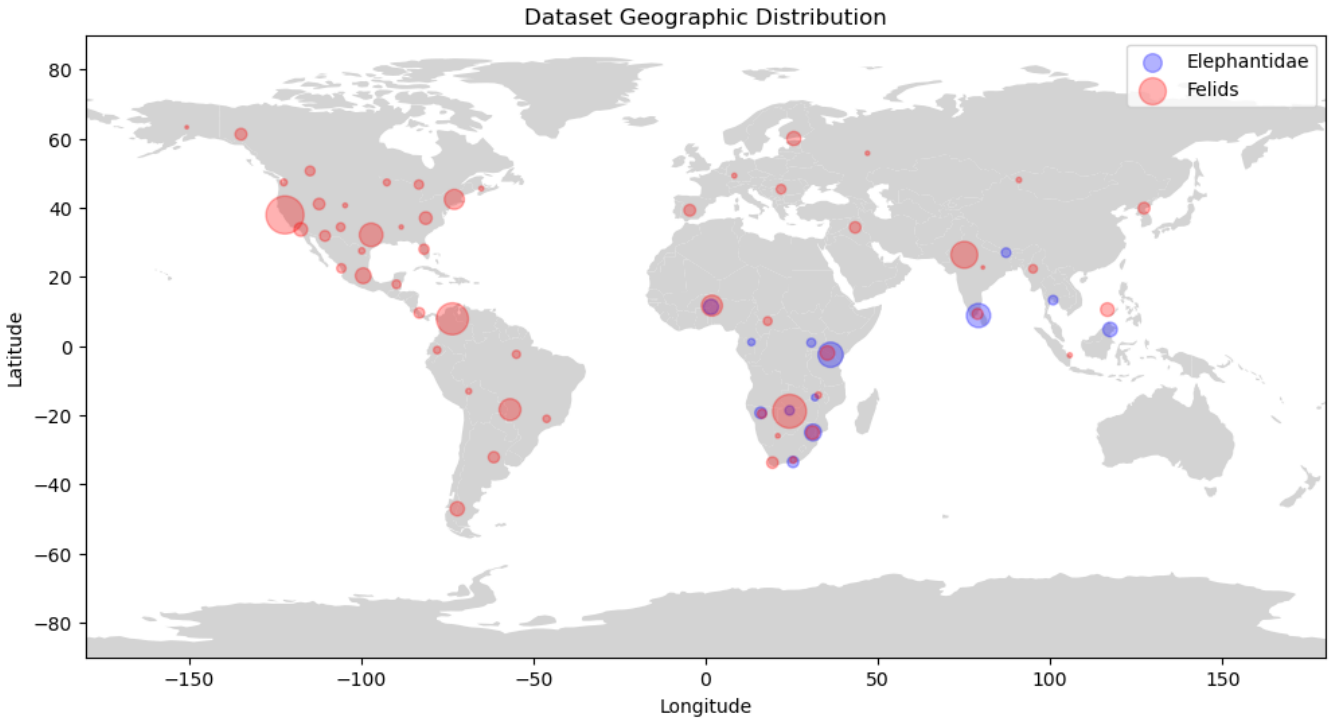


Fig. 2: The Geographic Distribution of *Elephantidae* and *Felidae* Showcasing the Global Observation Hotspots and Quantities

The secondary metadata forms a snapshot of the spatiotemporal conditions of the observation. This is accomplished through the use of Open-Meteo [18] open-source weather API. Open-Meteo sources comprehensive weather conditions directly from Copernicus satellite data [9]. The API allows for the collection of daily aggregate and hourly weather information, at a resolution of 11 km^2 around the specified location. The weather information produces features such as elevation, precipitation, cloud-cover, apparent temperature, and maximum temperature, amongst others.

From the derived secondary metadata, further extrapolation is applied to extract additional features. In total, a set of 53 environmental features form the spatiotemporal snapshot of each observation. This set of features is referred to as metadata going forward. For a comprehensive breakdown of all metadata (primary and secondary) features, review Table VI.

Missing values are interpolated using the annual aggregate profile per species. Where interpolation could not be applied to the missing values, the observation record is dropped from the dataset.

IV. Methods

The methods employed to generate the novel CE classifier, are the byproduct of the performance of metadata and image classifiers within the taxonomic tree. Within the framework of Carlos & Silla, 2011 [25] the image and metadata classifiers form symmetric local classifiers per parent node, using the taxonomic tree as the underlying structure. Despite the valid consideration of massive computational expense by Gomez-donosos et al., 2021 [6] the cascading methodology is further explored to determine the magnitude of the potential performance improvement on both metadata and image classifiers. This section elaborates on the pre-processing, methods, and structure of the metadata and image models, before detailing their combined collaboration within the CE classifier.

A. Metadata Classification

Metadata classifiers attempt to capture the relationship between the geographic and environmental features, and the observed wildlife. The metadata model has a feature-space encompassing 53 unique metadata values describing each observation. To determine a robust classifier that captures the potential relation between environment and wildlife, five classification models are explored: XGBoost [3], decision tree, neural network, random forest, and Adaboost [23].

1) Pre-processing: The metadata is pre-processed within a pipeline to generate a usable format of data that can serve as a set of input features with corresponding taxonomic labels. All models perform identical steps of pre-processing. Critical components include the limitation of taxonomic classes to include at least five unique observations, in order to serve as a class in the training data; and the bolstering of minority classes through the use of random oversampling within the imbalanced-learn library [13] in order to account for the data imbalance.

An additional critical element to the pipeline is the encoding of coordinate features. Coordinates in the form of

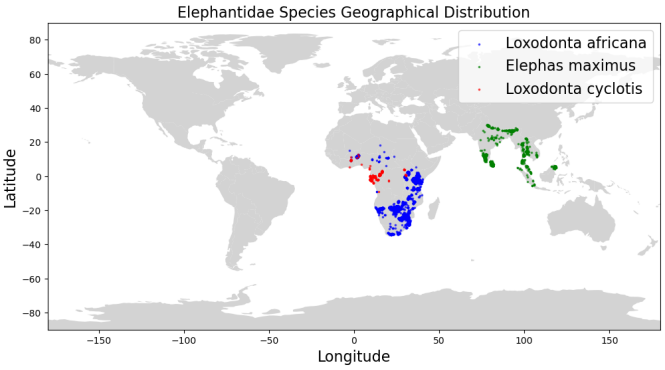
latitude and longitude generate an enormous feature space, which may present a challenge to the models to learn the relevance of the location. In order to encode the coordinates, a K-means clustering algorithm [14] is trained on the latitude and longitude features to generate a set of observation hot-spots, reducing the feature space. However, due to the cascading classification process a unique model is required at each taxonomic parent node of the taxonomic tree. This requires automated selection of the centroid hyper-parameter in order to capture the relevant geographic information contained within the taxonomic children, which vastly differs per taxonomic level. Automated centroid selection at each parent node requires a set of K-means models trained on a range of centroid values from $[4-60]$. The mean Silhouette score [21] of each model is calculated with the global maximum indicating the optimal number of centroids. A visualized example of the encoding process can be seen in Figure 3.

The above pre-processing steps are performed to generate the input features and labels for the metadata models, but also serve to enforce the taxonomic parent node restriction on the dataset. Such that only observations of the taxonomic child nodes are present in the feature vector and labels of each parent node model.

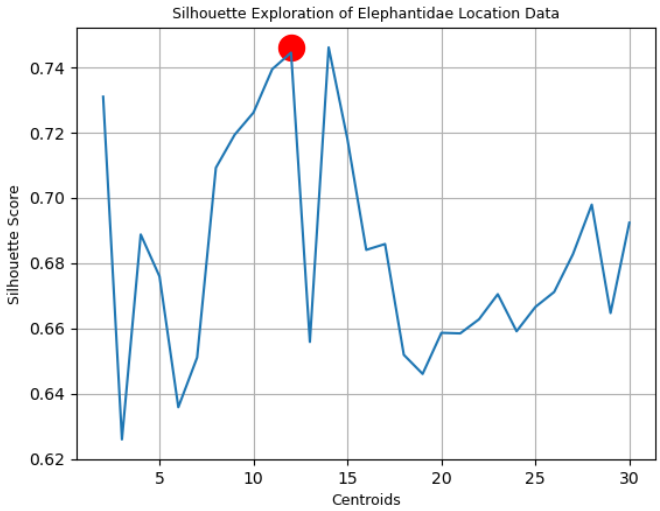
2) Models: The set of the proposed metadata classification algorithms is trained on the same input features generated by the pre-processing per taxonomic parent node, with the exceptions of necessary algorithm dependent modifications such as One-Hot-Encoding (OHE). Prior to each model's training, the balanced class-weightings are calculated to effectively weigh the model learning process to reduce the impact of the imbalanced classes. Each proposed algorithm performs hyper-parameter tuning and enforces a best-model save policy based on the cross-validation balanced accuracy evaluation metric.

The XGBoost model [3] is an optimized gradient descent algorithm. The prominent hyper-parameter requiring tuning is the maximum allowed depth within the decision tree ensemble classifiers. The maximum tree depth is tuned using 5-fold cross-validation to generate a mean model balanced accuracy, per maximum depth hyper-parameter. The set of maximum depths is within the range of depth 1 to the number of input features. The model producing the optimal accuracy within the hyper-parameter tuning process is assumed to be the best model and saved according to the best-model policy. This process is repeated for each local parent node classifier.

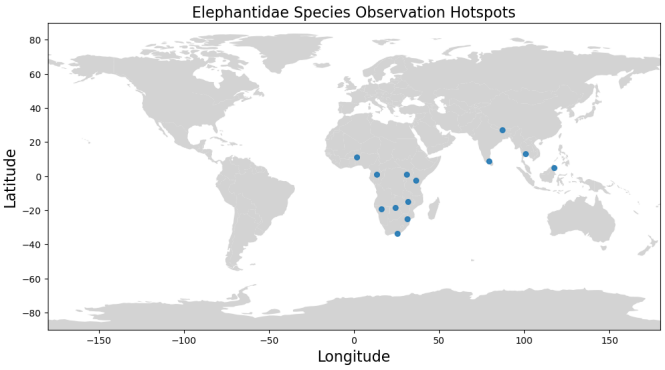
The decision tree model contains set hyper-parameters, specifically 2 as the minimum number of samples to generate a split in the tree, and the Gini split evaluation metric. The hyper-parameter tuning optimizes the maximum tree depth, performing an identical optimization process as described for the XGBoost model. The top-performing model is saved due to the best-model policy.



(a) *Elephantidae* Observation Geographic Distribution



(b) Silhouette Score for Centroid Range with Global Maximum Score



(c) *Elephantidae* Observation Locations Reduced to Hotspots

Fig. 3: Coordinate Encoding Process

The neural network model architecture makes use of a variably sized input layer dependent on the number of OHE features present. The hidden layers comprise of two densely connected layers, with 80 and 60 neurons in each respective layer. A softmax output layer generates class prediction probabilities. The variably sized input layer is critical due to the unknown quantity of OHE location hotspots. The top-

performing model is saved due to the best-model policy. The architecture is visualized in Figure 4.

Hyper-parameter tuning focused on finding the optimal learning rate due to the varied level of abstraction within separate taxonomic levels. The set of learning rates optimized over include rates $\in \{0.1, 0.01, 0.001, 0.0001\}$. An Adam optimizer is used within the model training. Training epochs are evaluated using categorical cross-entropy as the measure of loss with categorical accuracy as the resultant metric. The top-performing model is saved due to the best-model policy.

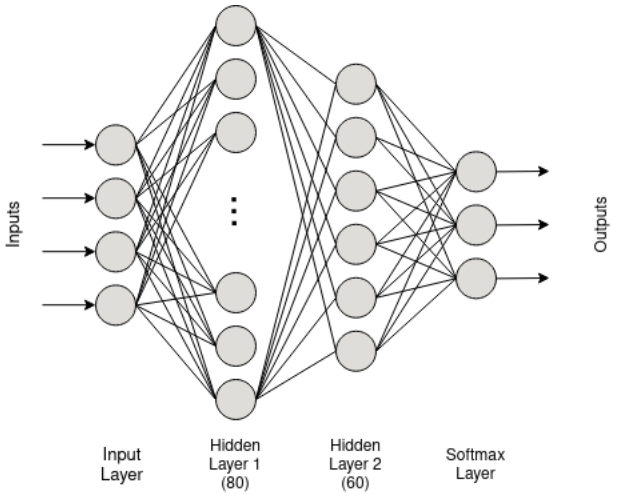


Fig. 4: Neural Network Metadata Model Architecture

The random forest model's fixed hyper-parameters include a minimum sample split of two observations, as well as 100 trees within the forest. Identically to the proposed hyper-parameter tuning of both XGBoost and the decision tree, the random forest model optimizes over the ensemble maximum tree depth using the previously specified range. The top-performing model is saved due to the best-model policy.

The Adaboost model [23] is an linear ensemble method utilizing multiple weak models in order to rectify errors of the previous model. The main hyper-parameter optimized over is the quantity of weak models utilized in ensemble. The Adaboost models makes use of 5-fold cross-validation across a range of estimators from a single estimator through to 200 estimators. Model performance is evaluated using balanced accuracy. The top-performing model is saved due to the best-model policy.

B. Image Classification

Image classification attempts to determine the wildlife present exclusively from information contained within the observation image. Based on the current state of image classification models, the state-of-the-art model is considered to be an image transformer such as ViT-G/14 [33]. However, due to the quantity of models to be trained, minimizing the computational and spatial costs is paramount. Hence, the EfficientNet-B6 model is selected to serve as the image classifier due to its reduced size and efficient training [28] while maintaining near state-of-the-art performance (based on ImageNet [22] dataset results).

1) Pre-processing Using Mega-detector: The raw images of the observations within the dataset contain erroneous and mislabelled data. Errors consist of empty frames and images of foot-prints (spoor) amongst others. To generate the set of usable images present within the dataset, Mega-detector is utilized. Mega-detector [1] is a trained You Only Look Once (YOLO) [19] object-detection model, capable of detecting, placing a bounding box, and labelling three classes (animal, human, and vehicle).

The raw observation images are processed by Mega-detector using a NVIDIA GeForce 3060 GPU unit. Each prediction enforced a strict 75% certainty cut-off to reduce misdetections. The resulting bounding-boxes and associated labels are use to crop the original images, producing a single or multiple cropped and centered images per raw observation image. To maintain the image resolution, despite cropping, a combination of Lanczos interpolation and edge enhancement kernels are used. Each cropped image is assumed to maintain the provided wildlife labels from the observation. Figure 5 showcases the above process.

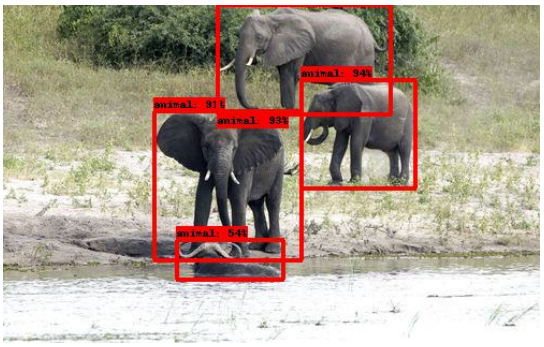
2) Model: The EfficientNet model [28], specifically model variant B6, is a CNN. It contains a unique compound scaling architecture to achieve increased accuracy and efficiency over alternative CNN architectures. This model variant contains 43.3 million parameters. The input dimension requires a $(528, 528, 3)$ image, with pixel values within the $[0, 255]$ value range. The capabilities of transfer learning expedite the process of model training [30] and reduce the quantity of training data required. Hence, an EfficientNet-B6 model pre-trained on the ImageNet [12] dataset has been used.

The model is augmented to suit the dataset. The 1000 class output layer is replaced by a two-dimensional global average pooling layer to flatten and average the previous convolutional layer, followed by a densely connected softmax output layer tailored to the required number of class predictions. During training, all layers of the pre-trained network are frozen, with the exception of the global pooling layer and softmax output layer.

Each local parent node model training makes use of an NVIDIA GeForce 3060 GPU unit. The training process makes use of batch sizes ranging from $[4 - 32]$ due to the long-tailed distribution. Additionally, each model is trained over 25 epochs, with a best-model only policy. Prior to training, class weightings are calculated to achieve a balanced rate of learning across all classes. The model's hyper-parameters and optimization specifications are as follows: an Adam optimizer with learning rate 0.001, categorical cross-entropy loss, and balanced accuracy evaluation metric.

Note, due to the computational and time demands of the study, no hyper-parameter tuning is conducted on the image

(a) Raw Observation Image



(b) Mega-detector Animal Detections

(c) Sub-image 1

(d) Sub-image 2

(e) Sub-image 3

Fig. 5: Image Pre-processing Using Mega-detector: Figure 5a shows the raw observation image, Figure 5b shows the Mega-detector animal detections highlighted by bounding-boxes, and Figures 5c, 5d, 5e show the extracted sub-images.

classification models. Each image model is required to be selected and trained manually due the excessive memory requirements of successive model training overloading the GPU unit.

C. Cascading Ensemble Classification

The novel CE classifier capitalizes upon the taxonomic performance trends within the metadata and image classification components. By exploiting the taxonomic strengths, and mitigating the taxonomic weaknesses, a robust classifier is created. This classifier is capable of classifying wildlife taxonomic levels as low as the subspecies taxonomy. Based on the results explored within sections V-A and V-B, the metadata classifier's observed performance increases with decreasing taxonomic depth, whereby the image classifier's performance demonstrates the opposite behaviour, decreasing performance with decreasing taxonomic depth. The top-performing metadata classification model to be utilized within the cascading ensemble classifier is XGBoost.

A novel architecture is proposed to capitalize upon the observed taxonomic performance trends, termed a cascading ensemble classifier. Within the framework provided by Carlos & Silla, 2011 [25] the architecture is comprised of dual selective classifiers (local classifier per parent node approach) utilizing metadata and image models respectively (metadata and image tree). The selective classifiers jointly decide the predicted child node at each parent node within the top-down cascading classification. The joint-decision effectively exploits the opposing taxonomic trends of metadata and image classifiers by weighting, combining, and restricting the result to a normal distribution to maintain a softmax output. The weightings are specified within Table I.

Due to the Mega-detector [1] pre-processing, the image classification models must be capable of dealing with multiple images per observation. To account for this, all images per observation are predicted, averaged, and restricted to a normal distribution to enforce a softmax output per observation image classification.

Taxonomic Level	Metadata Weight	Image Weight
Family	0.1	0.9
Genus	0.2	0.8
Species	0.5	0.5
Subspecies	0.9	0.1

TABLE I: Metadata and Image Taxonomic Level Weights

The concept of the CE classifier is summarized within Figure 6. The communications protocol in the figure represents the above described joint-classification process. The cascading ensemble classifier makes use of over 60 models to effectively classify images within the proposed dataset, fulfilling the roles of location encoding, metadata classification, and image classification.

Note, it is essential that the output of at each parent node within both the image and metadata trees contain the same size dimensions.

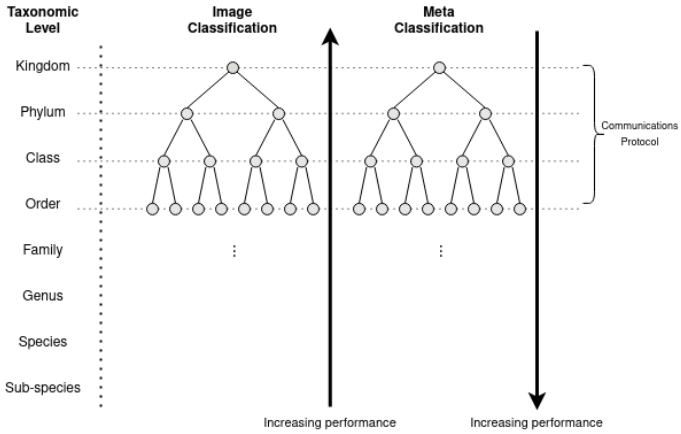


Fig. 6: Cascade Ensemble Classifier Concept

V. Experiments

The set of experiments conducted within this study revolve around determining the performance of metadata and image classification models within the taxonomic structure, to determine taxonomic performance trends. Additionally, to evaluate and compare the resulting CE classification within the taxonomic structure and against baseline classification performances.

A. Taxonomic Metadata Classifier Comparison and Taxonomic Trend

The aim of this experiment is to evaluate multiple metadata classification model's performance within the taxonomic structure, to determine a potential trend and identify the most robust metadata classification model.

The pre-processing pipeline utilized to generate the set of input features and labels, creates a disjoint validation set encompassing 20% percent of the data.

Each of the proposed models, XGBoost [3], decision tree, neural network, random forest, and Adaboost [23] is trained on the same feature vectors and labels per taxonomic parent node. Each model is trained per the specifications in Section IV-A2 and further evaluated using the disjoint validation dataset. The models performance is analyzed using balanced accuracy and f1-score as the chosen performance metrics due to the long-tailed distribution. Specifically, analysing taxonomic performance to determine trends, and further analysis at the individual level.

B. Taxonomic Image Classification Trend

The aim of this experiment is to evaluate the image classification model's performance within the taxonomic structure, to determine potential taxonomic trends.

The image classifier is constructed with architecture and hyper-parameters detailed within Section IV-B2. Similarly to the metadata pipeline, the image data is restricted to the taxonomic parent node, enforcing classes limited to the children of the parent node. Each image dataset is evaluated using a disjoint validation set, containing 20% of the available images to that node. The sum of images contained within all validation sets is approximately 5000 images.

To provide comparable metrics to the metadata classification experiment, the image classification models are evaluated using balanced accuracy and f1-score. Specifically, analysing taxonomic level performance to determine potential trends, and further analysis at the individual level.

C. Cascading Ensemble Classifier Joint Performance

The aim of this experiment is to evaluate the joint-classification of the CE classifier against the individual image and metadata classification component's performance within the cascading structure. This is to determine the effectiveness of image and metadata model weighting within the novel classifier. Within the experiment, the novel CE classifier is referred to as joint-classification to represent the combining of the metadata and image model predictions.

The cascading ensemble classifier uses the structure described in Section IV-C. Immediate comparison of the predicted label at each taxonomic level within the cascade, enforces early stopping where misclassification occurs. The experiment records the predicted label of each individual metadata and image model, the combined prediction label, and the true label at each symmetric node within the dual trees during the cascading prediction process.

The disjoint validation set evaluating the models performance is the same disjoint set of images as used within Section V-B, with each linked observation providing the metadata. Note, the validation metadata forms a disjoint set from the metadata training and test sets used in Section V-A.

The metric of comparison similarly focuses on the balanced accuracy at each taxonomic level, as well as the mean differences in recall, precision, and f1-score at each taxonomic level between the joint and image classifiers. Note, the metrics are considered different from the prior two experiments, as only those successfully classified at the parent taxonomy level are further classified within the cascading structure.

D. Cascading Ensemble Classifier Baseline Comparison

The aim of this experiment is to compare the performance of the novel CE classifier against baseline metadata and image classification models. The baseline models use the standard flat-classification approach to classify 45 species of wildlife. Each of the classification models follows the model structure specified for the above metadata and image classifier experiments. The validation dataset is the same as for the cascading ensemble joint-performance experiment, comprising of approximately 5000 unseen observations (images and metadata) on which each model is evaluated. The metric of comparison utilized is accuracy, to compare the overall performance of the models on the validation set.

VI. Results

This section presents the results and further elaborates on the findings across all four experiments.

A. Taxonomic Metadata Classifier Comparison and Taxonomic Trend

Figure 7 visualizes the mean balanced accuracy performance of each model type, per taxonomic level. The resulting figure contains a red-vertical dividing line placed to indicate the taxonomic family results are not representative of the expected results. This is as only 2 taxonomic families *Felidae* and *Elephantidae* are present in the dataset, providing an incomplete result. The figure captures an almost linear increasing trend in all metadata model's performance as the taxonomic level decreases from genus to subspecies. Notably the XGBoost model [3] outperformed all others at all taxonomic levels in terms of balanced accuracy.

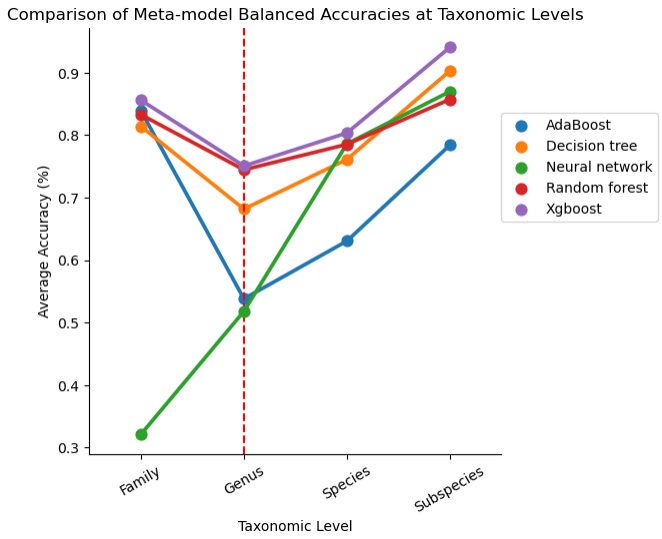


Fig. 7: Average Balanced Accuracy Metadata Model Performance per Taxonomic Level

Figure 8 shows a similar trend within the f1-score metric. The XGBoost model remains the highest performing classifier, whilst Adaboost performs the poorest. As such XGBoost is the highest performing metadata classifier across both evaluation metrics.

Figures 14a, 15a, 16a, 17a in the Appendix describe the model f1-scores at the individual level across the taxonomy, in alignment with the quantity of observations of each class. The figures showcase, that with reduced data quantities, even as low as 10 observations, the metadata models still form accurate classifiers. However, the greater quantities of available data are always aligned with comparatively higher performing models.

B. Taxonomic Image Classification Trend

Figure 9 visualizes the mean balanced accuracy of the image classification models at each taxonomic level. The figure contains an identical red vertical line fulfilling the same purpose as described previously in Section VI-A. The figure showcases a trend whereby balanced accuracy decreases as the taxonomic level decreases. However, there seems to be an increase at the subspecies taxonomy. The bars represent

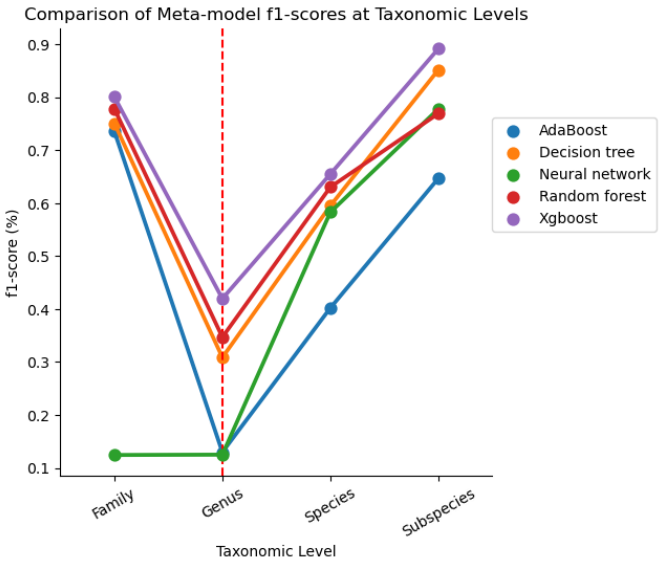


Fig. 8: Metadata Model f1-score per Taxonomic Level

percentile interval for each taxonomic level, showcasing the 95% confidence interval of where the data falls. The percentile interval bars increase with decreasing taxonomic levels, with the subspecies taxonomy having the largest percentile interval.

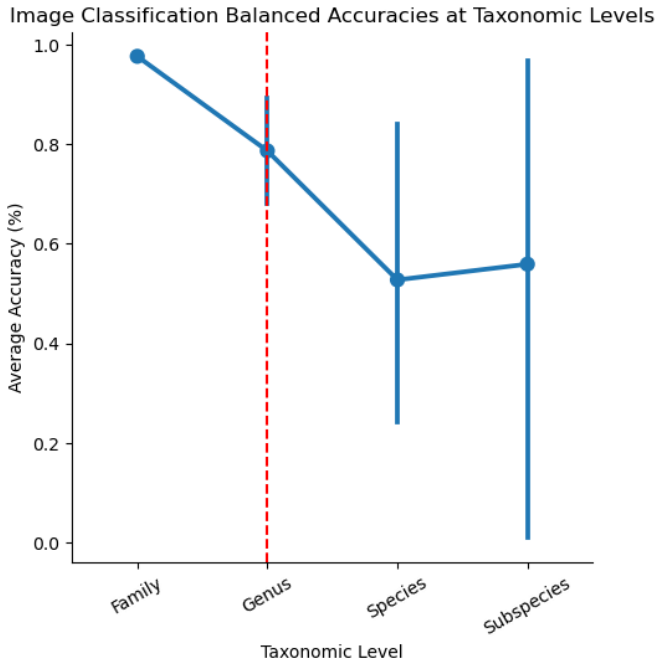


Fig. 9: Average Balanced Accuracy Image Classification Performance per Taxonomic Level

Figure 10 follows the same trend identified within Figure 9, showcasing a decreasing performance across precision, recall, and f1-score metrics as the taxonomic level decreases, with a spike in performance at the subspecies taxonomic level.

Figures 14b, 15b, 16b, 17b in the Appendix describe the

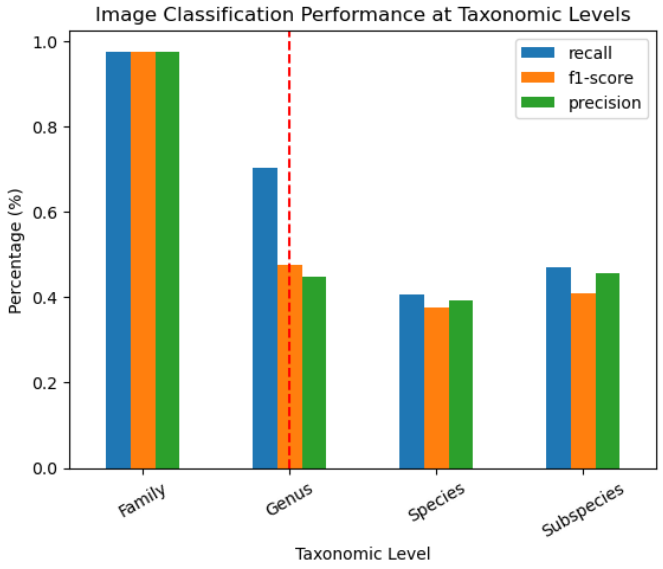


Fig. 10: Image Classification Performance per Taxonomic Level

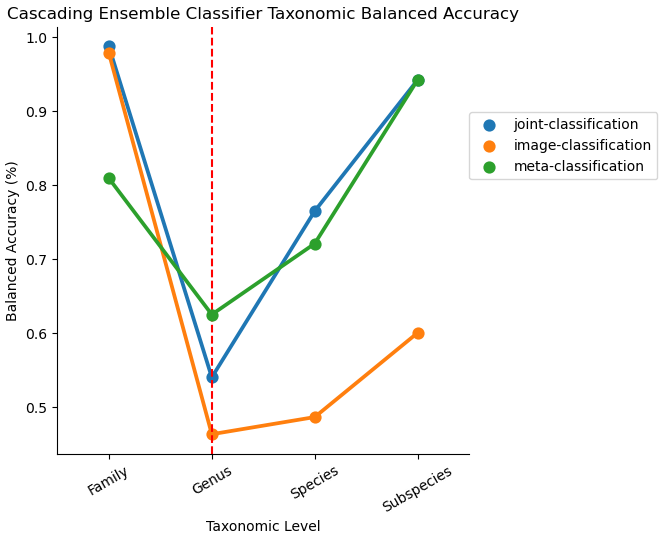


Fig. 11: Cascading Ensemble Classifier Prediction Breakdown

individual model f1-scores at each taxonomic level. The figures show an increase in zero-performance per class occurrences, aligned with comparatively less observations within the dataset.

C. Cascading Ensemble Classifier Joint Performance

Figure 11 showcases a near 100% percent balanced accuracy within the initial family taxonomy. However, as stated within VI-A this represents a skewed and incomplete metric, hence the presence of the red vertical line. Each of the models within the figure experiences an increasing trend in balanced accuracy as the taxonomic level decreases. At the genus taxonomic level, the image classifier experiences the poorest performance at approximately 50% percent balanced accuracy, followed by the joint-classification with 58% percent balanced accuracy, and the best performing metadata-classification at 65% balanced accuracy. At the species taxonomic level, the joint-classifier experiences the best performance with approximately 80% percent balanced accuracy, overtaking both metadata and image classifier's balanced accuracies. At the final subspecies taxonomic level the joint-classification and metadata classification contain equal balanced accuracies, significantly outperforming the image classification by almost 20% percent balanced accuracy.

Figure 12 presents an increasing trend in the difference between the joint-classification and the image model performance, as the taxonomic level decreases. Showcasing an average 35% percent joint-classification improvement over image classification at the subspecies level. Despite the metadata-classification exceeding the joint-classification at the genus level in Figure 11, the joint-classification shows an almost 10% percent performance improvement in comparison to image classification.

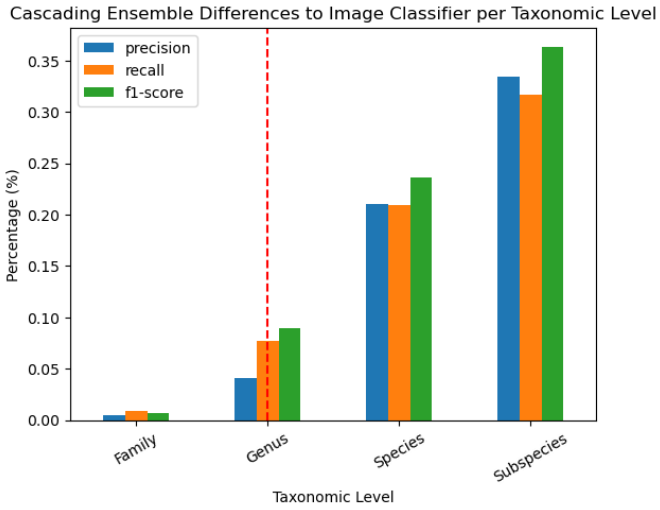


Fig. 12: Differences between Joint-classification and Image Classification

D. Cascading Ensemble Classifier Baseline Comparison Results

Figure 13 showcases the baseline performance of the image classification, the metadata classification, and the novel CE classifier on the same evaluation set. The metadata classification achieves an accuracy of approximately 38%, the image classifier achieves an accuracy of approximately 8%, and the cascading ensemble classifier achieves an accuracy of approximately 84%, performing at nearly 10 times the performance of the baseline image classifier and 2.5 times that of the baseline metadata classifier.

VII. Discussion

There exists a trend of increasing metadata classification performance as taxonomic level decreases. This is best de-

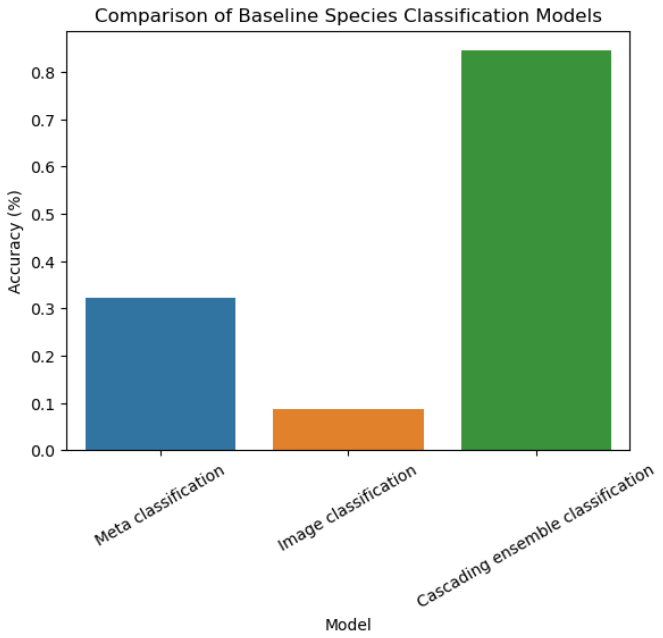


Fig. 13: Species Classification Baseline Comparison against Cascading Ensemble Classifier

scribed as a linear relation. The rational behind this occurrence is that the metadata captures environmental/ behavioural snapshots that become increasingly unique to each taxonomic level as they decrease.

From a biological perspective, each set of child nodes within the taxonomic tree represent a genetically distinct group of individuals/ individual. Genetic expression dictates wildlife environmental conditions and behavioural traits which are captured by spatiotemporal snapshots.

Higher taxonomic levels contain overlapping spatiotemporal information due to the variety of wildlife they encapsulate. This provides a challenge to spatiotemporal classification. Lower taxonomic levels provide niche spatiotemporal snapshots that uniquely represent the wildlife, providing increased capability to perform accurate classification. The cascade from high to low taxonomic levels provides a continually narrowing spatiotemporal domain (removing overlaps), enabling the niche spatiotemporal snapshots to form.

For example, Figure 3a showcases only the geographic distribution of *Elephantidae* species, in which there is clear separation of environmental traits enabling classification.

The resulting metadata classification performance significantly exceeds that of image classification at the species and subspecies taxonomy by at least 20% as seen in Figure 11.

Allopatric[1] and sympatric[2] speciation bear consideration in further research to determine the biological limitations that determine the usefulness of spatiotemporal classification. Such that metadata models may experience better performance for allopatric wildlife, as compared to sympatric wildlife in which the metadata models could potentially experience significantly worse performance.

There exists a trend of decreasing image classification performance as taxonomic level decreases. An opposite linear relation to that of metadata classification. This relationship can be attributed to the levels of abstraction dictated by taxonomic level, and the availability of data within the taxonomic structure.

Classification at higher levels of the taxonomic tree provides greater visual differences, such as the distinction between the Kingdoms of *Animalia*, *Plantae*, *Fungi*, *Protista*, and *Monera*, or to the family levels of *Felidae* and *Elephantidae*. The classification balanced accuracy result between the latter in Figure 9 is near 95% supporting this rationale. Decreasing taxonomic levels (genus, species, and subspecies) commonly increase the visual similarity within taxonomic child nodes, requiring an acute extraction of visual detail in order to perform accurate classification.

The availability and quality of data within the taxonomic structure must be considered in parallel to the abstraction level. Higher taxonomic levels generally contain a greater quantity of data as they encompassing all child nodes, as compared to a single child node. The training of classification models at higher taxonomic levels thereby has access to greater data quantities, in which greater visual distinctions are more easily extracted despite potentially poor data quality. Classification models at lower taxonomic levels, access a reduced data quantity in which the quality of the data may obscure acute details required for accurate classification, resulting in decreased performance. The increased number of zero-performance per class occurrences aligned with comparatively less observations (Figures 14b, 15b, 16b, 17b) and the increasing confidence intervals (Figure 9) support argumentation for the decreasing image classification performance as taxonomic levels decrease.

However, the dataset provided is considered challenging for image classification due to the variable image quality as a result of the pre-processing. Image classification performance may achieve improved results at all taxonomic levels on alternate datasets.

The novel CE classifier outperforms traditional wildlife flat-classification methodologies by a margin of 2.5 - 10 times (Figure 13). The improved performance measure can be attributed to two factors. Firstly, the cascading classification structure. Applied to image classification alone is expected to outperform the flat-classification image model, based on Figure 11, showcasing the potential for classification improvement when used.

Secondly, the joint-classification weighting both the image and metadata model results at each taxonomic node, to enhance their strengths and mitigate the weaknesses. Specifically, the metadata model component determining unique spatiotemporal snapshots at the species and subspecies level. This provides reliable low taxonomic level classification, where image classification typically fails within a challenging dataset.

[1]The evolution of distinct species due to geographic separation

[2]The evolution of a distinct species within the same geographic region

The image classification model structure and hyper-parameters are identical within the flat-classifier and the novel CE classifier. This is to provide a comparison focusing solely on the cascading structure and joint-classification differences in the methodology. In both uses of image classification, the models could be improved through hyper-parameter tuning, data augmentation, or an alternative state-of-the-art model.

The metadata flat-classification performance hints at unique spatiotemporal snapshots present in the dataset, that allow a 30% (Figure Figure 13) percent accuracy at the species taxonomy. The metadata classification model performance at all taxonomic levels may decrease if extended to species that contain overlapping spatiotemporal features. The cascading classification structure should minimize the potential for this occurrence, but it bears consideration that spatiotemporal information may not always provide the high level of classification performance as seen within this study.

Despite the above considerations, the relatively poor performance of the image and metadata classifiers in comparison to the novel cascading ensemble method, showcase its capability to accurately classify wildlife above when presented with a challenging dataset.

VIII. Conclusion and Future Work

The use of spatiotemporal metadata within the novel cascading ensemble classifier significantly improves the accuracy of wildlife classification.

The study creates and uses a novel dataset, combining wildlife observations and their taxonomic labels with the spatiotemporal conditions of the observation.

The use of the novel dataset within the study lead to the following statements regarding metadata and image classifiers through the taxonomic levels: metadata classification models exhibit an increasing performance trend as taxonomic level decreases, providing strong classification capability at lower taxonomic levels such as species and subspecies; image classification experiences a decreasing performance trend as taxonomic level decreases, providing accurate classification at higher taxonomic levels such as family and genus.

The discovered image and metadata classification trends within the taxonomic structure enables an effective weighted joint-classification, combining the individual elements to produce an output greater than the sum of its parts. The joint-classification is the primary concept of the novel cascading ensemble methodology, providing increased accuracy within the range of 2.5-10 times that of traditional metadata and image classifiers on challenging datasets. This result showcases the enormous potential underlying the taxonomic structure of wildlife and the essential inclusion of metadata in wildlife classification using methods such as the CE classifier.

The novel cascading classifier is limited by the enormous computational time, complexity, and memory involved, proving infeasible as a real-time classifier within its current form. However, the potential classification performance enhancements implores further research. Further research avenues include: the biological underpinnings, such as allopatric and sympatric speciation, leading to successful metadata classification as a means of bolstering wildlife classification; the comparison of the novel cascading classifier on a standardized wildlife dataset, such as the iNaturalist challenge to provide a concrete comparative benchmark; and further research into reducing the computational and memory demands of cascading classifiers, considering the underlying structure of the biological domain and the enormous potential benefit.

References

[1] Sara Beery, Dan Morris, and Siyu Yang. Efficient pipeline for camera trap image review. *arXiv preprint arXiv:1907.06772*, 2019.

[2] Ruilong Chen, Ruth Little, Lyudmila Mihaylova, Richard Delahay, and Ruth Cox. Wildlife surveillance using deep learning methods. *Ecology and Evolution*, 9(17):9453–9466, 2019.

[3] Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. *CoRR*, abs/1603.02754, 2016.

[4] Travis Dawson. Spatiotemporal wildlife dataset, 2023.

[5] Jeffrey S. Ellen, Casey A. Graff, and Mark D. Ohman. Improving plankton image classification using context metadata. *Limnology and Oceanography: Methods*, 17(8):439–461, 2019.

[6] Francisco Gomez-Donoso, Félix Escalona, Ferran Pérez-Esteve, and Miguel Cazorla. Accurate multilevel classification for wildlife images. *Computational Intelligence and Neuroscience*, 2021:1–11, 2021.

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

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[9] H. Hersbach, B. Bell, P. Berrisford, G. Biavati, A. Horányi, J. Muñoz Sabater, J. Nicolas, C. Peubey, R. Radu, I. Rozum, D. Schepers, A. Simmons, C. Soci, D. Dee, and J-N. Thépaut. ERA5 hourly data on single levels from 1959 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). (Updated daily), 2018. Licensed under the Creative Commons Attribution 4.0 International License https://creativecommons.org/licenses/by/4.0/.

[10] iNaturalist. inaturalist, 2008.

[11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[12] Standford Vision Lab. Imagenet, 2020.

[13] Guillaume Lemaître, Fernando Nogueira, and Christos K. Aridas. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, 18(17):1–5, 2017.

[14] Shie Mannor, Xin Jin, Jiawei Han, Xin Jin, Jiawei Han, Xin Jin, Jiawei Han, and Xinhua Zhang. K-means clustering. *Encyclopedia of Machine Learning*, page 563–564, 2011.

[15] Zhongqi Miao, Kaitlyn M. Gaynor, Jiayun Wang, Ziwei Liu, Oliver Muellerklein, Mohammad Sadegh Norouzzadeh, Alex McInturff, Rauri C. Bowie, Ran Nathan, Stella X. Yu, and et al. Insights and approaches using deep learning to classify wildlife. *Scientific Reports*, 9(1), 2019.

[16] Sajid Nazir and Muhammad Kaleem. Advances in image acquisition and processing technologies transforming animal ecological studies. *Ecological Informatics*, 61:101212, 2021.

[17] University of Michigan. Snapshot safari project, 2010.

[18] Open-Meteo. Open-Meteo Website. https://open-meteo.com/, n.d. Accessed during the period January 2023-May 2023.

[19] Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *CoRR*, abs/1506.02640, 2015.

[20] Pâmela M Rezende, Joicymara S Xavier, David B Ascher, Gabriel R Fernandes, and Douglas E V Pires. Evaluating hierarchical machine learning approaches to classify biological databases. *Briefings in Bioinformatics*, 23(4), 06 2022. bbac216.

[21] Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.
[22] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.
[23] Robert E Schapire. Explaining Adaboost. In *Empirical inference*, pages 37–52. Springer, 2013.
[24] Stefan Schneider, Saul Greenberg, Graham W. Taylor, and Stefan C. Kremer. Three critical factors affecting automated image species recognition performance for camera traps. *Ecology and Evolution*, 10(7):3503–3517, 2020.
[25] Carlos Silla and Alex Freitas. A survey of hierarchical classification across different application domains. *Data Mining and Knowledge Discovery*, 22:31–72, 01 2011.
[26] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. sep 2014. Unpublished article. Originating from ILSVRC Challenge.
[27] Fanny Simões, Charles Bouveyron, and Frédéric Precioso. Deepwild: Wildlife identification, localisation and estimation on camera trap videos using deep learning. *Ecological Informatics*, 75:102095, Apr 2023.
[28] Mingxing Tan and Quoc V Le. Efficientnet: Rethinking model scaling for convolutional neural networks. may 2019. Unpublished article. Originating from ILSVRC Challenge.
[29] Kevin Tang, Manohar Paluri, Li Fei-Fei, Rob Fergus, and Lubomir Bourdev. Improving image classification with location context. *2015 IEEE International Conference on Computer Vision (ICCV)*, page 1008–1016, 2015.
[30] J. Christopher Terry, Helen E. Roy, and Tom A. August. Thinking like a naturalist: Enhancing computer vision of citizen science images by harnessing contextual data. *Methods in Ecology and Evolution*, 11:303–315, 2019.
[31] Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The inaturalist species classification and detection dataset. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun 2018.
[32] Jana Wäldchen and Patrick Mäder. Machine learning for image based species identification. *Methods in Ecology and Evolution*, 9(11):2216–2225, 2018.
[33] Xiaohua Zhai, Alexander Kolesnikov, Neil Houlsby, and Lucas Beyer. Scaling vision transformers. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
[34] Zooniverse. Zooniverse, 2009.

IX. Appendix

A. Additional Information

1) Novel Cascading Ensemble Classifier Components Discussion: The joint-classification and its individual components (image and metadata classifiers) shows a consistent increasing performance trend (Figure 11) as taxonomic levels decrease. This occurrence can be attributed to the cascading structure of the experiment. Only observations of sufficient quality enabling correct classification cascade to the next classifier. This results in increasing data quality as taxonomic levels decrease, promoting improved classification metrics. Figure 11 and 12 shows joint-classification significantly outperforming image classification at the genus and subspecies level by a margin as significant as 35%. This result is expected due to the reduced image quantity and the acute level of detail required to perform accurate image classification at lower taxonomic levels.

However, at the genus level, image classification is expected to outperform metadata classification due to the higher level of visual abstraction, but showed the worse performance. The anomalous result could potentially be attributed to unique spatiotemporal metadata snapshots at the genus level. It is expected that when applied to further taxonomic genera that the image classification results would significantly increase due to the higher level of visual discrepancy between the genera, whereas significant spatiotemporal overlap would potentially occur.

B. Tables and Figures

TABLE II: Dataset Taxonomic Family Breakdown

Taxonomic Name	Observation Count	Image Count
Felidae	44710	26668
Elephantidae	11292	14831

TABLE III: Dataset Taxonomic Genus Breakdown

Taxonomic Name	Observation Count	Image Count	Family
Lynx	20139	12121	*Felidae*
Panthera	12411	7485	*Felidae*
Puma	5239	2725	*Felidae*
Leopardus	2220	1391	*Felidae*
Acinonyx	1872	1245	*Felidae*
Felis	991	633	*Felidae*
Caracal	595	323	*Felidae*
Herpailurus	448	315	*Felidae*
Leptailurus	442	279	*Felidae*
Prionailurus	291	124	*Felidae*
Neofelis	24	11	*Felidae*
Otocolobus	20	6	*Felidae*
Pardofelis	11	4	*Felidae*
Catopuma	7	4	*Felidae*
Loxodonta	9407	12810	*Elephantidae*
Elephas	1885	2021	*Elephantidae*

TABLE IV: Dataset Taxonomic Species Breakdown

Taxonomic Name	Observation Count	Image Count	Family
Lynx rufus	19043	11641	*Felidae*
Panthera leo	6782	4935	*Felidae*
Puma concolor	5239	2725	*Felidae*
Panthera pardus	2964	1360	*Felidae*
Acinonyx jubatus	1872	1245	*Felidae*
Panthera onca	1831	745	*Felidae*
Leopardus pardalis	1324	866	*Felidae*
Panthera tigris	788	433	*Felidae*
Lynx canadensis	727	355	*Felidae*
Caracal caracal	573	299	*Felidae*
Herpailurus yagouaroundi	448	315	*Felidae*
Leptailurus serval	442	279	*Felidae*
Felis lybica	372	252	*Felidae*
Leopardus weidii	362	209	*Felidae*
Felis silvestris	345	228	*Felidae*
Leopardus geoffroyi	288	174	*Felidae*
Lynx lynx	260	75	*Felidae*
Felis chaus	228	138	*Felidae*
Prionailurus bengalensis	178	76	*Felidae*
Lynx pardinus	109	50	*Felidae*
Prionailurus javanensis	77	31	*Felidae*
Leopardus tigrinus	67	36	*Felidae*
Leopardus guigna	55	37	*Felidae*
Leopardus guttulus	47	33	*Felidae*
Panthera uncia	46	12	*Felidae*
Caracal aurata	22	24	*Felidae*
Otocolombus manul	20	6	*Felidae*
Neofelis diardi	19	8	*Felidae*
Felis nigripes	17	4	*Felidae*
Prionailurus viverrinus	17	9	*Felidae*
Leopardus pajeros	16	10	*Felidae*
Leopardus colocola	16	7	*Felidae*
Leopardus garleppi	12	6	*Felidae*
Prionailurus rubiginosus	12	5	*Felidae*
Pardofelis marmorata	11	3	*Felidae*
Leopardus emiliae	10	5	*Felidae*
Felis margarita	10	1	*Felidae*
Leopardus braccatus	9	5	*Felidae*
Leopardus jacobita	7	2	*Felidae*
Prionailurus planiceps	7	3	*Felidae*
Catopuma temminckii	6	2	*Felidae*
Neofelis nebulosa	5	2	*Felidae*
Felis bieti	4	0	*Felidae*
Leopardus fasciatus	2	0	*Felidae*
Catopuma badia	1	0	*Felidae*
Loxodonta africana	8939	12313	*Elephantidae*
Elephas maximus	1885	2021	*Elephantidae*
Loxodonta cyclotis	222	214	*Elephantidae*

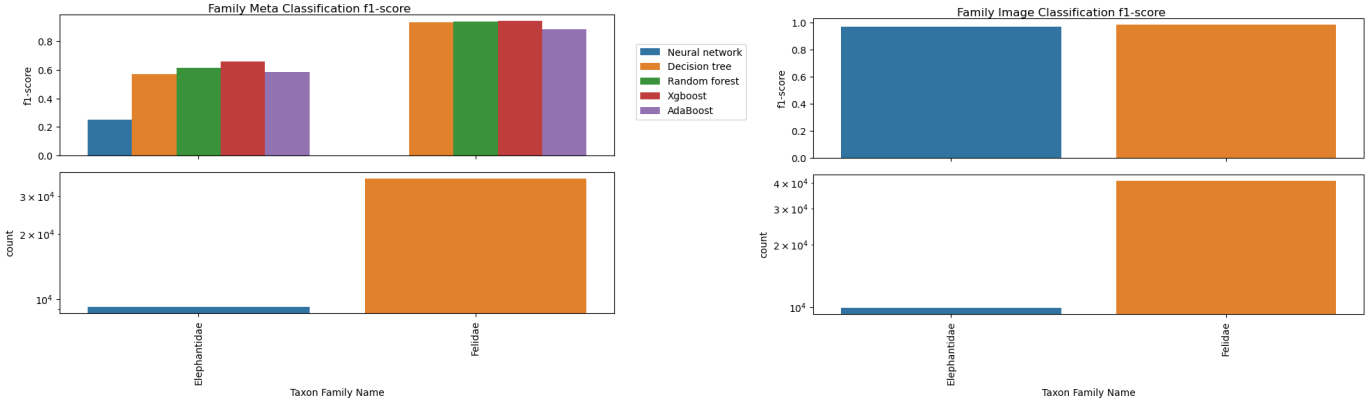
TABLE V: Dataset Taxonomic Subspecies Breakdown

Taxonomic Name	Observation Count	Image Count	Family
Panthera leo melanochaita	5587	4035	*Felidae*
Panthera pardus pardus	2449	1160	*Felidae*
Acinonyx jubatus jubatus	1642	1079	*Felidae*
Panthera leo leo	786	672	*Felidae*
Panthera tigris tigris	710	430	*Felidae*
Lynx rufus fasciatus	410	263	*Felidae*
Puma concolor couguar	357	172	*Felidae*
Caracal caracal caracal	349	163	*Felidae*
Lynx rufus rufus	262	163	*Felidae*
Panthera pardus fusca	222	107	*Felidae*
Puma concolor concolor	167	99	*Felidae*
Felis lybica cafra	167	112	*Felidae*
Panthera pardus kotiya	164	77	*Felidae*
Leptailurus serval lipostictus	152	96	*Felidae*
Felis silvestris silvestris	144	98	*Felidae*
Felis chaus affinis	105	69	*Felidae*
Leptailurus serval serval	87	54	*Felidae*
Felis lybica lybica	81	74	*Felidae*
Leopardus pardalis mitis	58	44	*Felidae*
Leptailurus serval constantina	57	46	*Felidae*
Acinonyx jubatus hecki	46	37	*Felidae*
Prionailurus javanensis sumatranus	38	16	*Felidae*
Prionailurus bengalensis euptilurus	32	16	*Felidae*
Lynx rufus escuinapae	27	16	*Felidae*
Felis lybica ornata	26	16	*Felidae*
Prionailurus bengalensis bengalensis	22	13	*Felidae*
Neofelis diardi borneensis	16	8	*Felidae*
Panthera tigris sondaica	15	0	*Felidae*
Panthera pardus delacouri	15	10	*Felidae*
Lynx lynx lynx	14	6	*Felidae*
Leopardus guigna guigna	13	9	*Felidae*
Leopardus pardalis pardalis	11	8	*Felidae*
Panthera pardus tulliana	11	5	*Felidae*
Panthera tigris altaica	11	0	*Felidae*
Panthera pardus melas	10	1	*Felidae*
Caracal caracal nubicus	10	9	*Felidae*
Felis silvestris caucasica	9	7	*Felidae*
Prionailurus javanensis javanensis	8	1	*Felidae*
Prionailurus viverrinus viverrinus	8	5	*Felidae*
Lynx lynx dinniki	8	7	*Felidae*
Leopardus guigna tigrillo	5	4	*Felidae*
Lynx lynx isabellinus	5	0	*Felidae*
Lynx lynx carpathicus	5	4	*Felidae*
Leopardus tigrinus oncilla	4	0	*Felidae*
Pardofelis marmorata marmorata	4	2	*Felidae*
Elephas maximus indicus	877	920	*Elephantidae*
Elephas maximus maximus	567	740	*Elephantidae*
Elephas maximus borneensis	180	146	*Elephantidae*
Elephas maximus sumatranus	18	12	*Elephantidae*

TABLE VI: metadata Features

Feature	Description	Unit/ Format	Timeframe
Observed on	Time of observation	ISO8601	Instant
Coordinates	World Geodetic System (WGS84)	(latitude, longitude)	Instant
Positional accuracy	Publicly available positional accuracy	R	Instant
Elevation	Meters above sea level	Meters (m)	Instant
Ground temperature (2m)	Air temperature 2 meters above ground	Celsius (° C)	Hourly
Relative humidity (2m)	Humidity 2 meters above ground	Percentage (%)	Hourly
Dew point (2m)	Dew point 2 meters above ground	Celsius (° C)	Hourly
Apparent temperature	Real feel temperature considering additional factors	Celsius (° C)	Hourly
Surface pressure	Atmospheric air pressure at the surface.	Hectopascal (hPa)	Hourly
Cloudcover	Cloudcover within the immediate area	% of area covered	Hourly
Low cloudcover	Cloudcover and fog up to an altitude of 3 kilometers	% of area covered	Hourly
Mid cloudcover	Cloudcover from $3-8$ kilometers altitude	% of area covered	Hourly
High cloudcover	Cloudcover from 8 kilometers altitude	% of area covered	Hourly
Wind speed (10m)	Wind speed at 10 meters above ground	kilometers per hour (km/h)	Hourly
Wind speed (100m)	Wind speed at 100 meters above ground	kilometers per hour (km/h)	Hourly
Wind direction (10m)	Wind direction at 10 meters above ground	Degrees (∘)	Hourly
Wind direction (100m)	Wind direction at 100 meters above ground	Degrees (∘)	Hourly
Wind gusts (10m)	Maximum wind speed of the preceding hour	kilometers per hour (km/h)	Hourly
Shortwave radiation	Average shortwave solar radiation of the preceding hour	Watt per square meter (W/m^2)	Hourly
Direct radiation	Average direct solar radiation of the preceding hour	Watt per square meter (W/m^2)	Hourly
Diffuse radiation	Average diffuse solar radiation of the preceding hour	Watt per square meter (W/m^2)	Hourly
Vapor pressure dificit	A high VPD effects the transpiration of plants	Kilopascal (kPa)	Hourly
Evapotranspiration	Water evaporation into the atmosphere	Millimeters (mm)	Hourly
ET0 FAO Evapotranspiration	Metric estimating required irregation for plants	Millimeters (mm)	Hourly
Precipitation	Hourly precipitation sum (rain, showers, snow)	Millimeters (mm)	Hourly
Snowfall	Hourly snowfall sum	Centimeters (cm)	Hourly
Rain	Large scale weather systems resulting rain	Millimeters (mm)	Hourly
Hourly Weather code	WMO numeric weather code	WMO code	Hourly
Soil temperature (0cm-7cm)	Temperature in the soil at 0-7 centimeters	Celsius (°)	Hourly
Soil temperature (7cm-28cm)	Temperature in the soil at 7-28 centimeters	Celsius (°)	Hourly
Soil temperature (28cm-100cm)	Temperature in the soil at 28-100 centimeters	Celsius (°)	Hourly
Soil moisture (0cm-7cm)	Average water content in the soil at 0-7 centimeters	Meter cubed per meter cubed (m^3/m^3)	Hourly
Soil moisture (7cm-28cm)	Average water content in the soil at 7-28 centimeters	Meter cubed per meter cubed (m^3/m^3)	Hourly
Soil moisture (28cm-100cm)	Average water content in the soil at 28-100 centimeters	Meter cubed per meter cubed (m^3/m^3)	Hourly
Daily Weather code	WMO numeric weather code	WMO code	Daily
Max temperature (2m)	Maximum daily temperature at 2 meters above ground	Degrees (∘)	Daily
Min temperature (2m)	Minimum daily temperature at 2 meters above ground	Degrees (∘)	Daily
Apparent temperature max	Maximum real-feel temperature at 2 meters above ground	Degrees (∘)	Daily
Apparent temperature min	Minimum real-feel temperature at 2 meters above ground	Degrees (∘)	Daily
Precipitation sum	The sum of daily precipitation (rain, showers, snowfall)	Millimeters (mm)	Daily
Rain sum	Sum of daily rain	Millimeters (mm)	Daily
Snowfall sum	Sum of daily snowfall	Centimeters (cm)	Daily
Precipitation hours	The number of hours with rain in a day	Z	Daily
Sunrise	Local sunrise time	ISO 8601	Daily
Sunset	Local sunset time	ISO 8601	Daily
Wind speed max (10m)	Maximum daily wind speed 10 meters above ground	Kilometers per hour (km/h)	Daily
Wind gusts (10m)	Maximum daily gust speed at 10 meters above ground	Kilometers per hour	Daily
Dominant wind direction	Dominant daily wind direction for winds at 10 meters	Kilometers per hour (km/h)	Daily
Shortwave radiation sum	The daily sum of short wave radiation	Megajoules per meter squared (MJ/m^2)	Daily
Daily evapotranspiration	Sum of daily evapotranspiration	Millimeters (mm)	Daily
Terrestrial	Terrestrial or aquatic observation	$\{0, 1\}$	Instant
Hemisphere	Location lies in the northern/ southern hemisphere	$\{0, 1\}$	Instant
Day	Sighting occurrence in light/ dark	$\{0, 1\}$	Instant
Season	Season of sighting, dependent on hemisphere	Season	Instant

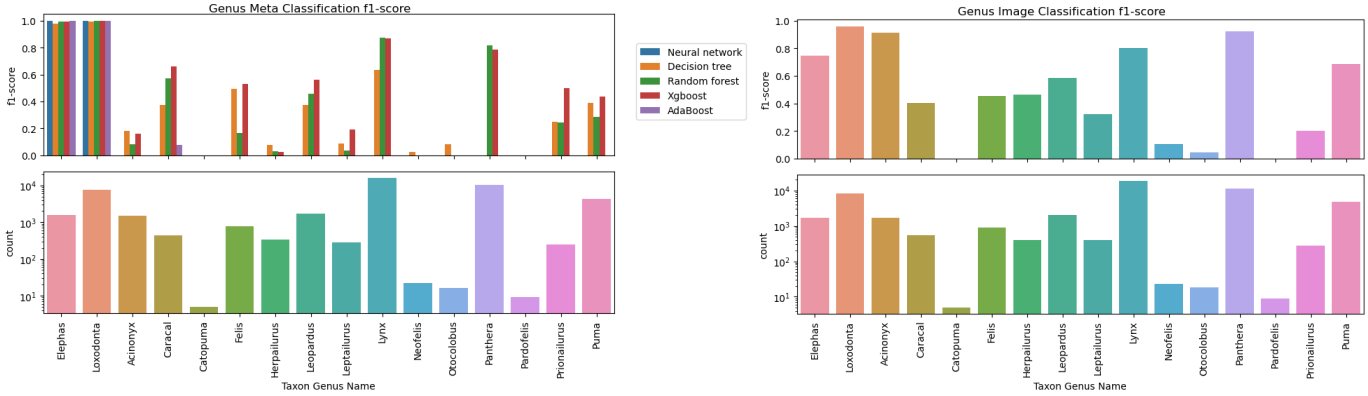
The table occurs in sections including: primary metadata, secondary hourly metadata, secondary daily aggregate metadata, and further derived secondary metadata. Descriptions of the hourly and daily metadata are sourced from the Open-Meteo API documentation, for a further and more detailed resource please review the API documentation.



(a) Metadata-model Taxon Family Performance

(b) Image-model Taxon Family Performance

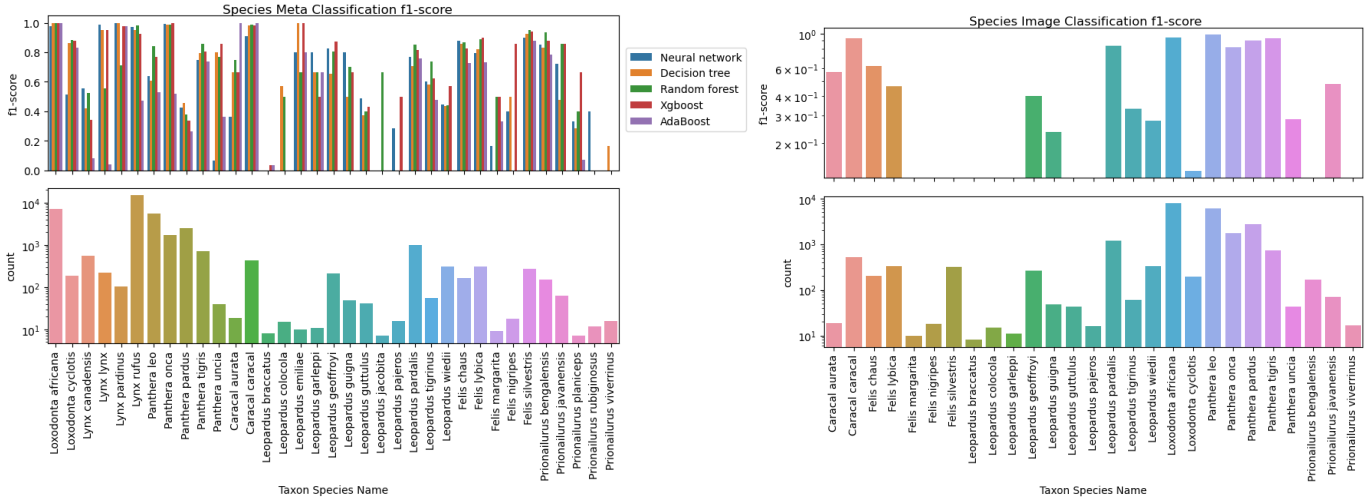
Fig. 14: Taxonomic Family Metadata and Image Model f1-score



(a) Metadata-model Taxon Genus Performance

(b) Image-model Taxon Genus Performance

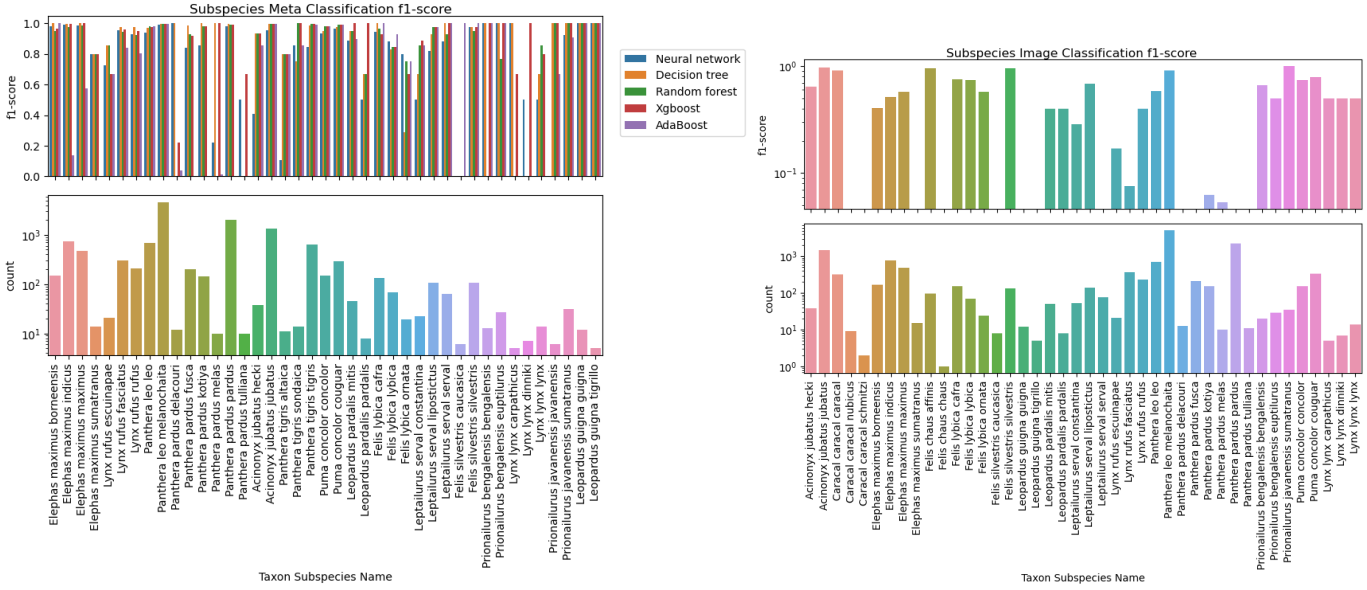
Fig. 15: Taxonomic Genus Metadata and Image Model f1-score



(a) Metadata-model Taxon Species Performance

(b) Image-model Taxon Species Performance

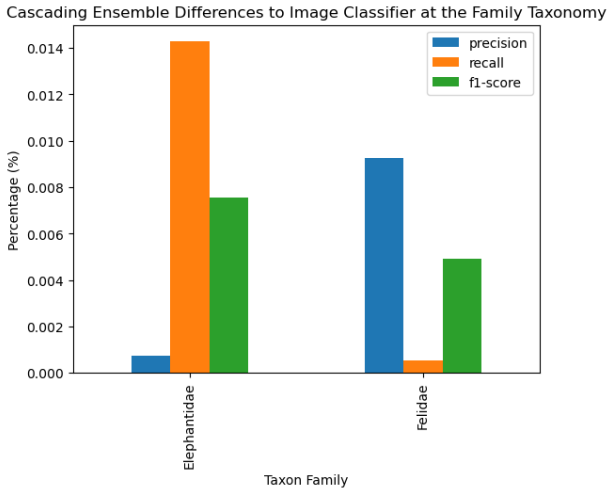
Fig. 16: Taxonomic Species Metadata and Image Model f1-score



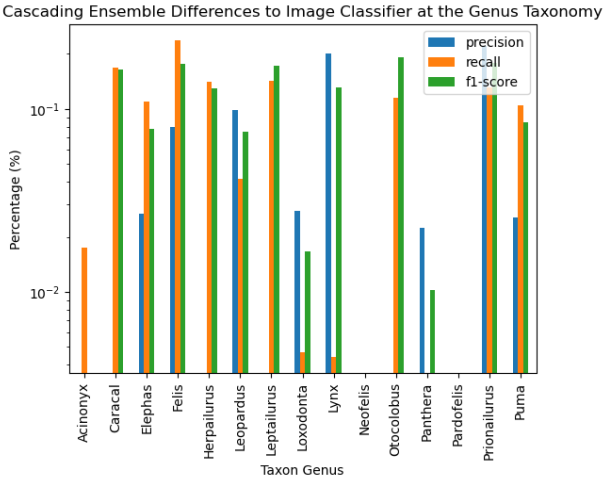
(a) Metadata-model Taxon Subspecies Performance

(b) Image-model Taxon Subspecies Performance

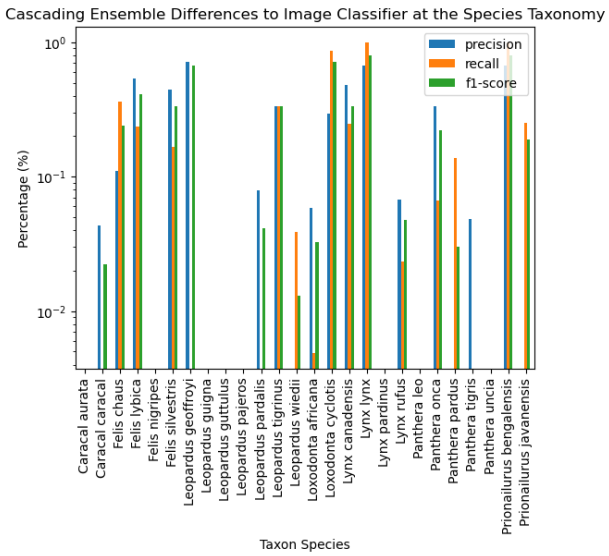
Fig. 17: Taxonomic Subspecies Metadata and Image Model f1-score



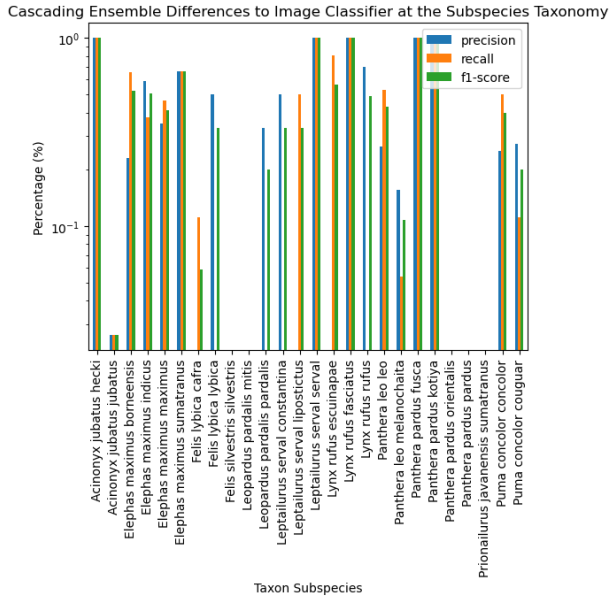
(a) Performance Differences between the Cascading Ensemble Classifier and the Image Classifier at the Family Taxonomy



(b) Performance Differences between the Cascading Ensemble Classifier and the Image Classifier at the Genus Taxonomy



(c) Performance Differences between the Cascading Ensemble Classifier and the Image Classifier at the Species Taxonomy



(d) Performance Differences between the Cascading Ensemble Classifier and the Image Classifier at the Subspecies Taxonomy