

Plant Species Identification using Probability Tree Approach of Deep Learning Models

Anantharaman Karun¹, Krishnakumar Divyasri¹ and Sella Veluswami Jansi Rani¹

¹Department of CSE, Sri Sivasubramaniya Nadar College of Engineering, Chennai, Tamilnadu, India

Abstract

This paper presents the submissions made by our team SVJ-SSN-CSE to PlantCLEF 2022[5]. The challenge's goal was to identify plant species based on the test set made from only plant images in the field, given a training dataset consisting of 80k species and up to 100 images each. We submitted 2 runs, one run implemented a ResNet50 which was fine-tuned by adding a flattened layer and a 2 dense hidden layers along with an output layer. The second run implemented a multi-level classification model emulating taxonomic classification tree using probabilistic tree approach. We achieved a Mean Reciprocal Rank (MRR) of 0.00005 for the run 1 which was a single fine-tuned ResNet50 classification and comparably, we achieved an MRR of 0.000158 for the second run which is a multi-level probabilistic tree approach using fine-tuned ResNet50 nodes.

Keywords

Plant species, Probabilistic approach, ResNet50, Image classification, Image identification, PlantCLEF 2022, Single Level Classification, Multi Level Classification,

1. Introduction

In recent times, the automation of tasks like identifying and associating the living world using visual cues has received considerable momentum. This assumes importance in ecological studies because, for accessing information pertinent to a species, or for recording such information in the first place, specimen identification is a prerequisite. But these are not straightforward tasks by any means, owing to the complicated nature of the form and appearance of biomass in general, which entails large levels of expertise and accuracy. This problem faced by taxonomic researchers, is touted to be one of the most critical hindrances to the Convention on Biological Diversity, according to the Rio Conference of 1992.

The most significant roadblock is the sheer number of the species under a category, with there being more than 10000 birds, about 30000 fishes, about 400000 flowering plants, and around 1.2 million invertebrates. Presently, a possible solution gaining increasing traction is that of content-based visual identification of specimen images. It requires to present the viewer with simplistic and comprehensive access to archived biological data, while gathering particulars and integrating the database in accelerated time-frames.

Recent activities show promising signs towards achieving this objective, as seen in the


CLEF 2022: Conference and Labs of the Evaluation Forum, September 5–8, 2022, Bologna, Italy

✉ jansi.research@gmail.com (S. V. Jansi Rani)

🆔 0000-0003-2863-5465 (S. V. Jansi Rani)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

LifeCLEF 2017[7] Plant Identification challenge¹, where datasets of 10,000 species were recorded with about 90% accuracy, by means of deep learning models. There exist a multitude of methods employed for this task, a few of which are based on multimedia data analysis and machine learning, as described by Gaston and O’Neill in 2004 [4]. They also postulated that the biologists could detect novel species if they were to circumvent the generation of voluminous training datasets, which preclude the efficient use of resources, and calculate the error rates, and set out with scaled-up automatic approaches to reduce them, then they could discover novel species effectively. To contribute in the direction towards attaining such an end, this paper presents the plant identification task that was organized within ImageCLEF 2022 for the system-oriented evaluation of visual-based plant identification.

The PlantCLEF2022[5] challenge edition proposes to take a step in progressing automatic identification of plant species by tackling a multi-image (and metadata) classification problem with a very large number of classes. The training dataset that has been made use of, could be distinguished into 2 main categories, in "trusted" and "web", i.e. with, and without species labels provided and checked by human experts, respectively, comprising a total of 4 million images between them. The "trusted" training dataset is based on a specific assortment of more than 2.9 million pictures across 80000 plant species. In contrast, the second data set is based on a collection of web images provided by the search engines, Google and Bing. The “trusted” dataset sized up to nearly 200 GB, and the “web” dataset measured around 100 GB of data, which is considerably large and necessitated remarkable amounts of computing power and time to be trained.

Through an initial literature survey, we came across ResNet50 - a pre trained transfer learning model, and Inception v3. These models were worked with, as they could be trained faster technically, and their subsequent failure will be discussed in this paper. Another method incorporated in the study was a probability tree approach in which the dataset was subdivided and trained separately, to result in a final score of 0.00015, and this behavior will also be elaborated on further.

2. Dataset



Figure 1: Sample image from the trusted dataset

This “trusted” dataset proved to be challenging to navigate through, owing to its significantly greater size, in comparison to those of other tasks, as well as the version from the previous year (it had 800,000 images, as opposed to the 2.8 million this year). 200 GB of training data

¹<https://www.imageclef.org/LifeCLEF2017>

consisted of images provided by the Global Biodiversity Information Facility (GBIF), and the Encyclopedia of Life (EOL). Sources have listed as museums, academic institutions, and other online platforms like iNaturalist and Pl@ntNet.

The dataset is unbalanced, with a minimum of only 101 images in one biological class(Ginkgoopsida), while the maximum number was found to be 2.1 million in another biological class(Magnoliopsida). The various classes, orders, families, genera, and species under each class, are depicted along with their count in Figure 1. The “web” dataset consists of 1.1 million images in total, and is also unbalanced in nature. Its largest class is Magnoliopsida, with 0.77 million images, and its smallest class is Ginkgoopsida, with only 25 images in it.

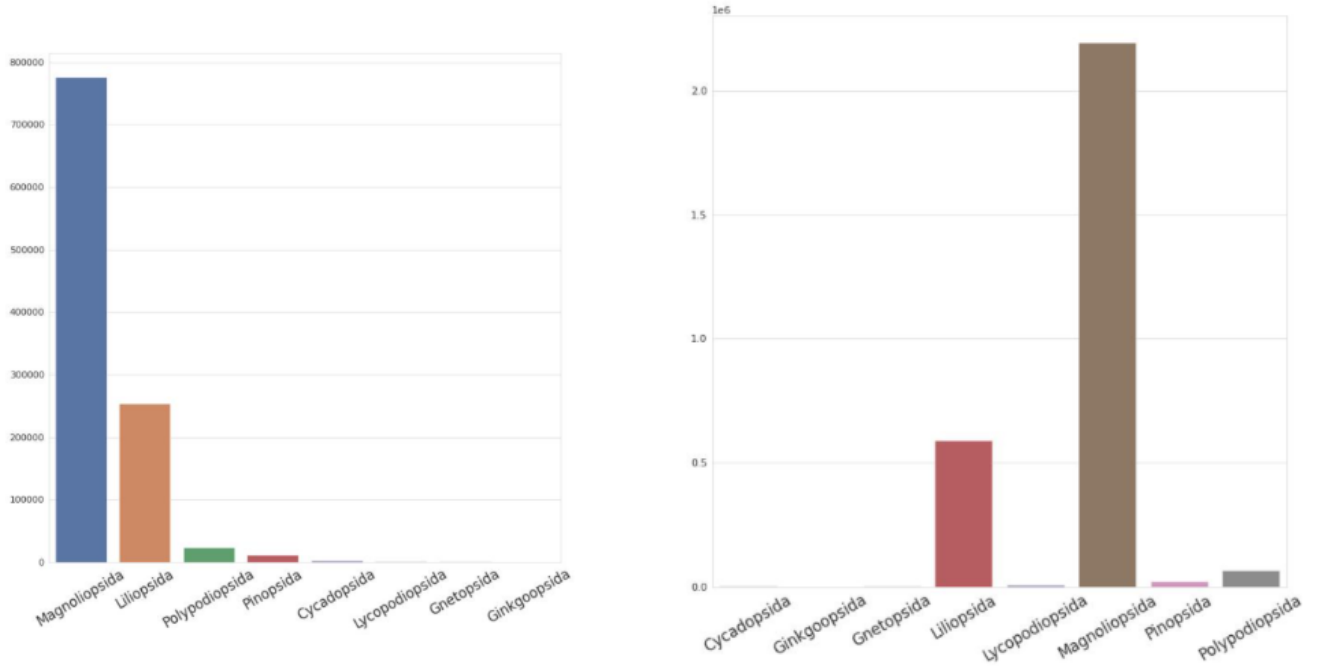


Figure 2: Web Dataset Distribution

Figure 3: Trusted Dataset Distribution

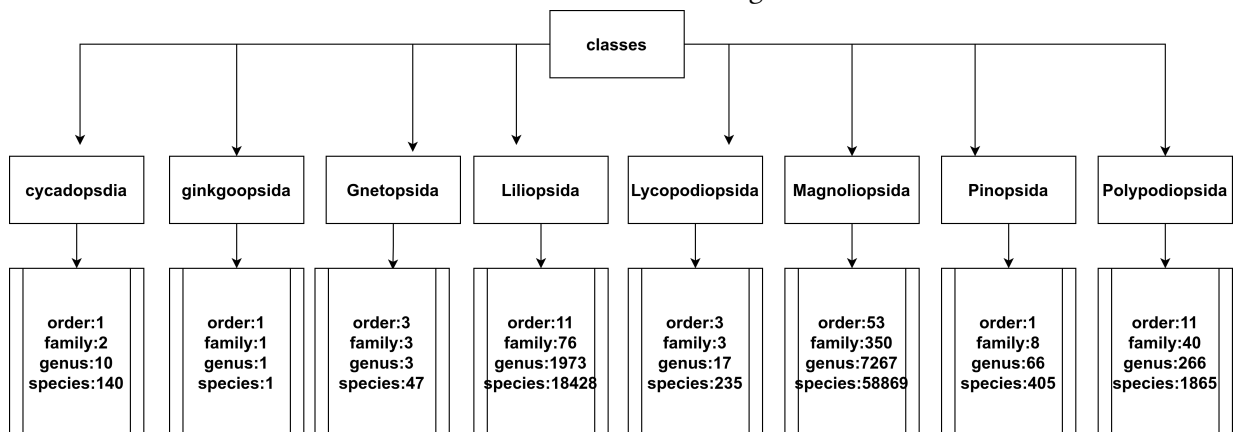


Figure 4: Class Split

3. Related works

On a literature review of the previous year submissions to this shared task, a few well-performing models were observed. In the years from 2014 to 2021, the working notes from the organizers claimed that the external dataset did not actually improve the performance but rather, it was counterproductive owing to it being not properly labelled, of poor quality and introducing a lot of noise into the training dataset.

In the year 2019, a team named LIRMM from the university of Montpellier [9], France modeled a deep convolutional neural network modeled on top of the Inception - v3 model, they trained the model as a classifier, using a cross entropy loss function and a softmax layer to get the output. The team who secured the third place in the same year, were from SSN College and VIT Chennai[10]. They had used probabilistic models, which say that the probability of encountering a species is proportional to its frequency in the training dataset. They combined it with traditional machine learning techniques like random forest, and XGBoost, and the authors identified 5 different feature groups and trained each one individually using a neural network, and combined them using artificial neural network techniques. A team from Turkey uses SesNet and DensNet pre trained models and uses Error Correcting Output Codes (ECOC) as an ensemble model achieving an accuracy of 87

Another team from Romania [1] also took a similar approach, where they used standard machine learning techniques like XGBoost, K-NN, and random forest, and combined them using ensembling techniques.

On referring to the submissions to the PlantCLEF2020 tasks, we found that a team from the USA [12], who had secured the first rank, used a Partial Domain Adaptation (PDA) technique. They used a NASNetLarge pretrained model to extract the features, and then develop a novel Adversarial Consistent Learning (ACL) approach through a unified deep architecture. It combined 3 different loss functions - an adversarial loss, feature consistency loss and source domain classification loss. The adversarial loss aims to learn domain-invariant features while the feature consistent loss helps in preserving the fine-grained feature transition between the two domains.

Neuron AI who secured the second rank used a triple-loss network[3]. They used two inception v4, and a ResNet model, which were further fine-tuned in different parts of the dataset. They were then converted into embeddings. They used a cosine function as a distant metric and it was then transformed with Inverse Distance Weighting, into probabilities for ranking the classes. They used many augmentation techniques on the raw dataset, along with ensemble techniques between the triple-loss functions.

The Costa Rica Institute of Technology (ITCR) [11] team from Costa Rica and France used Few Shot Adversarial Domain Adaptation (FSADA) approach for a few of their runs. They also used ResNet50 without the final classifier part. [8] Another team from SSN College also used ResNet50 in a similar manner.

In the task for the year 2021, only 2 teams had participated. Of those,[2]one team developed their own model based on a two-streamed Herbarium-Field Triplet Loss Network (HTFL), and trained a complimentary One-Streamed Mixed network (OSM) which is similar to a triple network.

From the above, we understood that ResNet50 trains faster generally due to its missing net property. It is also used extensively for such classification tasks. Inception v3 could also be

used as it shows a higher accuracy rate. So, these transfer learning, pre-trained models could be fine-tuned with the existing training datasets provided and an ensemble could be created between these two models. A probability tree approach could also be employed, where a set of 2000 species under each class are separately trained, and modeled using fine-tuned ResNet50.

4. Methodology

4.1. ResNet50 architecture

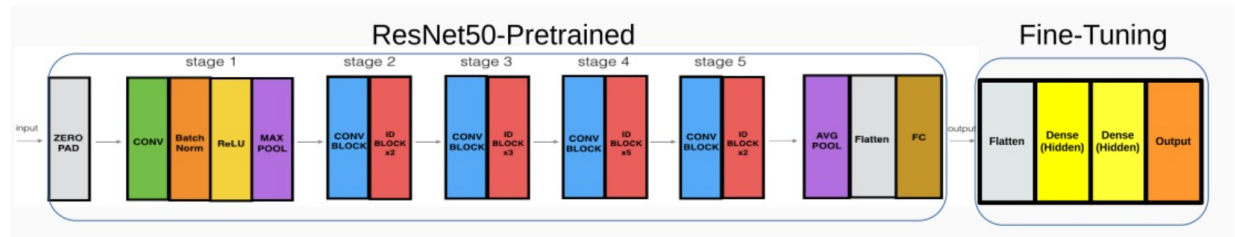


Figure 5: Architecture of ResNet50 including the finetune layer

ResNet50 is an innovative variant of the ResNet (Residual Network), which only had 50 layers. It was designed initially and documented in the 2015 computer vision research paper titled, "Deep Residual Learning For Image Recognition", authored by Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun[6]. Such a pretrained, transfer learning model was employed to train the model.

It is a general rule of thumb in machine learning to counter complex problems by inflating the number of layers in the neural network used. This brings about saturation in the accuracy, leading to the model being degraded. Residual Network model works to obviate this scenario. Such model relies on its fundamental strength in the concept of "skip connections". They are found at the core of the residual blocks, which essentially form the basis used by the deep residual nets to enhance the precision and accuracy of the models.

The work of the aforementioned skip connections is twofold: tackling the vanishing gradient problem, and associating an identity function with the model. The vanishing gradient is solved by means of channeling the gradient passage through an alternative short-cut, and subsequently back-propagating them to the earlier layers. The identity function helps the model to show similar levels of performance in the higher layers, as compared to the lower ones. In summation, the residual blocks are instrumental in easing the process of layers learning identity functions. Therefore, ResNet minimizes error occurrences and increases the efficiency of deep neural networks through increased number of neural layers. It starts off with a block of a stack of 3-layer bottleneck block

The convolution networks in this model have a 3x3 filter. It has 50 weighted layers and achieves a performance of 3.8 bn Floating point operations per second. A flatten layer and a couple dense layer was added on top of it followed by an output layer. The ResNet-50 has over 23 million trainable parameters.

4.2. Experimental Analysis

The Run 1 is a Single-level Fine-tuned ResNet50 classification. The batch size of the ResNet50 was set to 32, and the Learning Rate was set to $1e-5$, and a Sparse categorical cross-entropy was also employed, with an SGD optimizer. The model was set to run for 30 epochs, for which, and 2.4 million images of size $224 \times 224 \times 3$ mm were fed into the model. The model was run on a GPU Tesla V100, and the first epoch took 24 hrs to execute with the runtime seeing a gradual reduction with each passing epoch. The partially trained model gave a Mean Reciprocal Rank (MRR) score of 0, as the datasets it was trained on were highly unbalanced, and no augmentation techniques had been employed either. The second approach is a probability tree built using Multi-level Fine-tuned ResNet50.

4.3. Techniques used

4.3.1. ResNet50 and InceptionV3

Our initial approach to this problem was to employ transfer learning techniques with ResNet50 and Inception V3 models which were initialised with pre-trained ImageNet weights. The output layers of both models were flattened and 2 hidden layers (Dense) were added. An output layer consisting of 80000 neurons, representing each species was also appended to the above.

The pre-trained weights were left untouched, while the new layers were trained on the full trusted dataset, while the web dataset was used for validation. Ultimately, ResNet50 yielded a higher accuracy among the two, thus we picked the model for Run 1. A softmax layer was used to convert the output from the models to probabilities. The species with the top 30 probabilities, averaged across all images of an observation, were taken as the prediction for that observation.

The use sparse categorical could be pertained to the fact that each plant belongs to one definitive category. The task being a ranking task needed the use of cross entropy as the probability distribution needs to be concentrated, as close to the result as possible. On retrospection and re-examination MRR seems like a preferable metric considering this is a ranking task.

The poor results yielded by this model due to extreme imbalance of the data brought us to the conclusion that a single level general classifier would not suffice for this problem. Thus, we moved onto a multi-level classifier.

4.3.2. Probabilistic Tree Approach

A multi-level classification model was used, emulating the taxonomic classification tree, made up of Fine-tuned ResNet50 models, trained on various parts of the dataset, relevant to the node they are in. The probability of a species (a leaf node in the taxonomic tree, is calculated at the product of probabilities of each node on the path from the root to the leaf).

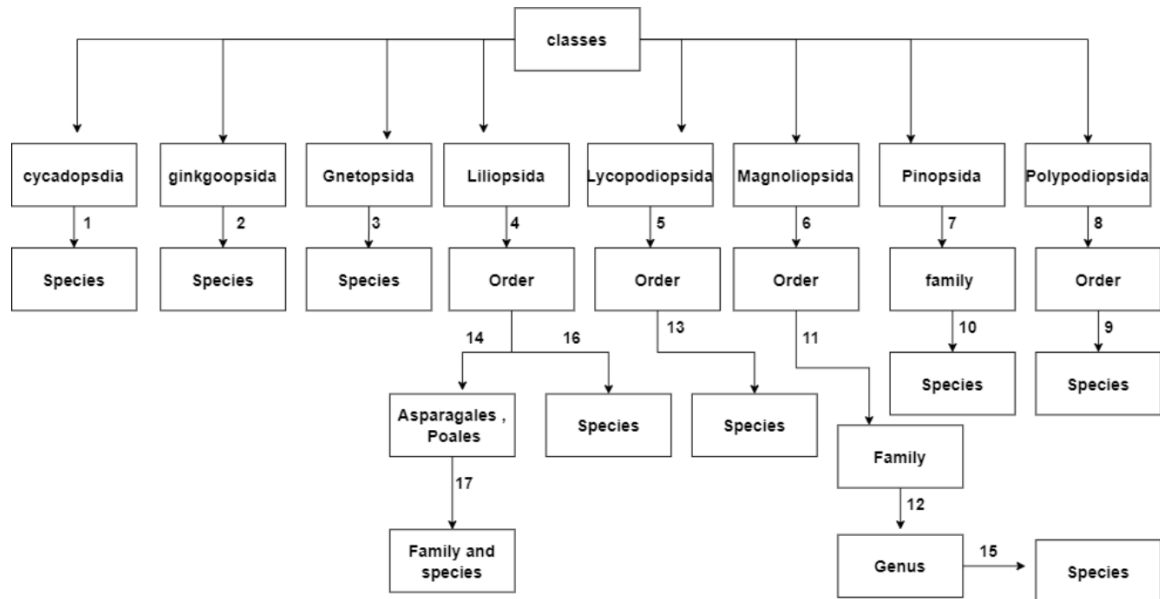


Figure 6: Architecture of the probability tree

The model consists of a 6-level depth tree starting with an abstract root node as the basis of classification, wherein all probabilities start at $1.25e-05$. From the root, images are classified into the broadest level of classification, the biological ‘class’. A maximum of 2000 images are taken in one level of classification. The given dataset has 8 different classes. The first three classes Cycadopsdia, Ginkgoopsida and Gnetopsida have about 2000 images under them, so they are retained as such, while the last 4 classes, Lycopodiopsida, Magnoliopsida, Pinopsida, and Polypodiopsida have more than 2000, images so they are further divided and classified into the order, family, and genus level. The class Liliopsida has about 5.9 million images, and it comprises the orders Asparagales and Poales, which have a huge number of images, and are separated into a distinct level of classification. These levels of classification are shown in Figure 2. This taxonomic approach allows us to keep closely related species with similar attributes together, while distancing them from species of wildly different properties. Taking probabilities at each level, instead of deciding on a specific class, allows us to account for errors at each level, and the overall probability computed at the end will serve to mitigate these errors.

The objective was to identify the top 30 most probable species an observation may belong to. An observation consists of multiple images relating to a single plant species.

For each individual image, predictions across the whole tree was computed. Since the task was multi-image classification, we had to combine the predictions for multiple images within each observation. We did this by taking the average of all class probabilities for the images within the observation. Then among those, the top 30 relevant classes were picked and ranked.

4.4. Result

The ranking metric used during the challenge was the MA-MRR, the MRR averaged by species. The MRR is a statistical metric for assessing any process that generates a list of potential answers to a sample of queries arranged by likelihood that the answers are correct. The multiplicative

inverse of the rank of the first accurate response is the reciprocal rank of a query response. For the entire test set, the MRR is the average of the reciprocal ranks:

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank_i}$$

where $|Q|$ is the total number of query occurrences in the test set.

The following are the final MRR score achieved from the 3 runs using the probability tree approach and the ResNet50 model table here.

| Run | MA MRR Score |
|-------------------|--------------|
| SVJ-SSN-CSE Run 1 | 0.00005 |
| SVJ-SSN-CSE Run 2 | 0.00000 |
| SVJ-SSN-CSE Run 3 | 0.00015 |

4.5. Conclusion

In this paper we proposed and elaborated on a novel probabilistic approach to solve the plant classification and identification problem given by the organizers [5]. Our model employed multi level classification using fine tuned ResNet50. We also employed standard CNN models like ResNet50 and Inception v3 which with low parameters did not perform well and took a long time to train. The low score could be due to the unbalanced distribution of data. Further improvements to the model could include training the weights from scratch instead of just fine-tuning the model. Additionally data and image augmentation techniques could be employed to balance the data set, along with better pre-processing of data for more efficient training.

References

- [1] Sara Ahmed, Berrin Yanikoglu, Erchan Aptoula, Ipek Ganiyusufoglu, Aras Yıldız, Kerem Yıldırım, Barış Sevilmiş, and M Şen. Plant identification with deep learning ensembles in expertlifecyclef 2018. 09 2018.
- [2] Sophia Chulif and Yang Loong Chang. Improved herbarium-field triplet network for cross-domain plant identification: Neuron submission to lifecyclef 2021 plant. 09 2021.
- [3] Sophia Chulif, Kiat Jing Heng, Teck Wei Chan, Mohammad Abdullah Al Monnaf, and Yang Loong Chang. Plant identification on amazonian and guiana shield flora: Neuron submission to lifecyclef 2019 plant. In *CLEF*, 2019.
- [4] Gaston, K.J., O'Neill, and M.A. Iautomated species identification: why not? philosophical transactions of the royal society of london b: Biological sciences. 09 2004.
- [5] Hervé Goëau, Pierre Bonnet, and Alexis Joly. Overview of PlantCLEF 2022: Image-based plant identification at global scale. In *Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum*, 2022.
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

- [7] Alexis Joly, Hervé Goëau, Hervé Glotin, Concetto Spampinato and Pierre Bonnet, Willem-Pier Vellinga, Jean-Christophe Lombardo, Robert Planque, and Henning Müller Simone Palazzo. Lifeclef 2017 lab overview: Multimedia species identification challenges. 09 2017.
- [8] Nanda H. Krishna, Rakesh M, and Ram Kaushik R. Plant species identification using transfer learning - plantclef 2020. In Linda Cappellato, Carsten Eickhoff, Nicola Ferro, and Aurélie Névéol, editors, *Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, September 22-25, 2020*, volume 2696 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2020.
- [9] N.H. Krishna, R. Praveen Kumar, R. Ram Kaushik, P. Mirunalini, C. Aravindan, and Jaisakthi S M. Species recommendation using machine learning - geolifeclef 2019. 09 2019.
- [10] G.E.H.M.D.T.T.W Si-Moussi, Sara. Species recommendation using environment and biotic associations. proceedings of clef 2019. 09 2019.
- [11] Juan Villacis-Llobet, Hervé Goëau, Pierre Bonnet, Alexis Joly, and Erick Mata-Montero. Domain adaptation in the context of herbarium collections: A submission to plantclef 2020. In *CLEF*, 2020.
- [12] Youshan Zhang and Brian Davison. Adversarial consistent learning on partial domain adaptation of plantclef 2020 challenge, 09 2020.