Exploring Biodiversity: A Multi-Model Approach to Multi-Label Plant Species Prediction

Darren Rawlings[1,*,†], Tim Chopard[2,*,†]

[1] *University of Groningen, Netherlands*
[2] *University of Edinburgh, United Kingdom*

Abstract

The research aims to develop a multi-label classification method for predicting plant species based on environmental factors such as satellite images, climate data, and other environmental variables. Utilizing a dataset of plant surveys in Europe, the study addresses challenges such as collating image, tabular and time series data from a variety of sources, multi-modal learning, with variable label counts. The approach employs dimensionality reduction via PCA, and an ensemble of machine learning models used to predict both species pseudo-probabilities and also the counts of species present in a given area. The potential contributions of this research include advancing our understanding of ecological systems, informing conservation efforts, and promoting the preservation of plant diversity — all essential components of a comprehensive approach to safeguarding the natural world.

Keywords

Multi-Label classification, Principal component analysis, ResNet, Vision Transformer, Swin Transformer, Gradient Boosting, XGBoost, GeoLifeCLEF 2024, CEUR-WS

1. Introduction

In this research, we aim to develop a model for predicting plant species in a specific location and time using various environmental factors as predictors. These predictors include satellite images, climatic time series, and other rasterized environmental data such as land cover, human footprint, bioclimatic variables, and soil characteristics. Our motivation behind this challenge is the potential usefulness of accurate plant species prediction in various scenarios related to biodiversity management and conservation, species identification and inventory tools, and education.

We were provided a large-scale training dataset of approximately 5 million plant occurrences in Europe, as well as train and test sets with 88987 and 4716 surveys, respectively. The predicted output will be multi-label, presence-absence data for all present species at each plot. The data covered 11255 different plant species, which created significant challenges associated with this task, including learning from single positive labels, dealing with strong class imbalance, multi-modal learning, and handling large-scale datasets.

The potential applications of accurate plant species prediction are numerous. High-resolution maps of species composition and related biodiversity indicators can be created to aid in scientific ecology studies and conservation efforts. The accuracy of species identification tools can be improved by reducing the list of candidate species observable at a given site. Additionally, location-based recommendation services and educational applications with features such as quests or contextualized educational pathways can be developed to facilitate biodiversity inventories and promote environmental education. We believe that our research will contribute to the advancement of plant species prediction and its practical applications in various fields.

The research was conducted as part of the GeoLifeCLEF 2024 competition on Kaggle [1], which is a part of the LifeCLEF initiative [2]. The competition aims to develop models for predicting plant species in a specific location and time using various environmental factors as predictors.

CLEF 2024: Conference and Labs of the Evaluation Forum, September 09–12, 2024, Grenoble, France

[*] GeoLifeCLEF Kaggle team BernIgen https://github.com/timchopard/ml-geolife

[†] These authors contributed equally.

d.rawlings@student.rug.nl (D. Rawlings); timchopard@pm.me (T. Chopard)
https://startung.github.io (D. Rawlings); https://cloudberries.io (T. Chopard)

2. Background

The GeoLifeCLEF challenge has been running for several of years. Each year, participants are tasked with predicting species distribution, but the challenge has evolved over time, with new datasets, evaluation metrics, and research questions introduced each year. Here, we provide an overview of the some of the recent submissions to GeoLifeCLEF challenge and summarize the key contributions.

In 2021, the GeoLifeCLEF challenge [3] focused on fine-grained visual categorization using remote sensing data. The winning submission by [4] leveraged contrastive learning to improve species distribution modeling (SDM) from remote sensing imagery. The authors explored the effectiveness of using only RGB imagery and the impact of adding altitude imagery to the model's performance. They introduced a new consistency-based model selection metric to enhance the model's generalization capabilities. The paper outlined potential areas for further research, including the impact of transformations and the utility of the consistency metric.

In 2022, the GeoLifeCLEF challenge [5] shifted its focus to predicting species distribution across the U.S. and France using remote sensing data and other covariates. The second-place submission by [6] proposed a classification approach with a spatial block-label swap regularization during training and an ensemble of deep learning models. Their method achieved a top-30 accuracy of 31.22% on the private test set, securing second place in the competition. The authors reflected on the results and suggested potential improvements and the importance of species distribution modeling for ecological research.

In 2023, the GeoLifeCLEF challenge [7] introduced a new dataset with single positive labels for each location, making multi-label prediction challenging. The winning submission by [8] proposed a three-step training strategy to leverage the single positive labels effectively. The authors introduced several CNN-based models and demonstrated their effectiveness compared to a simple baseline. The paper discussed the challenges of the new dataset and the proposed model's performance, providing detailed results and comparisons.

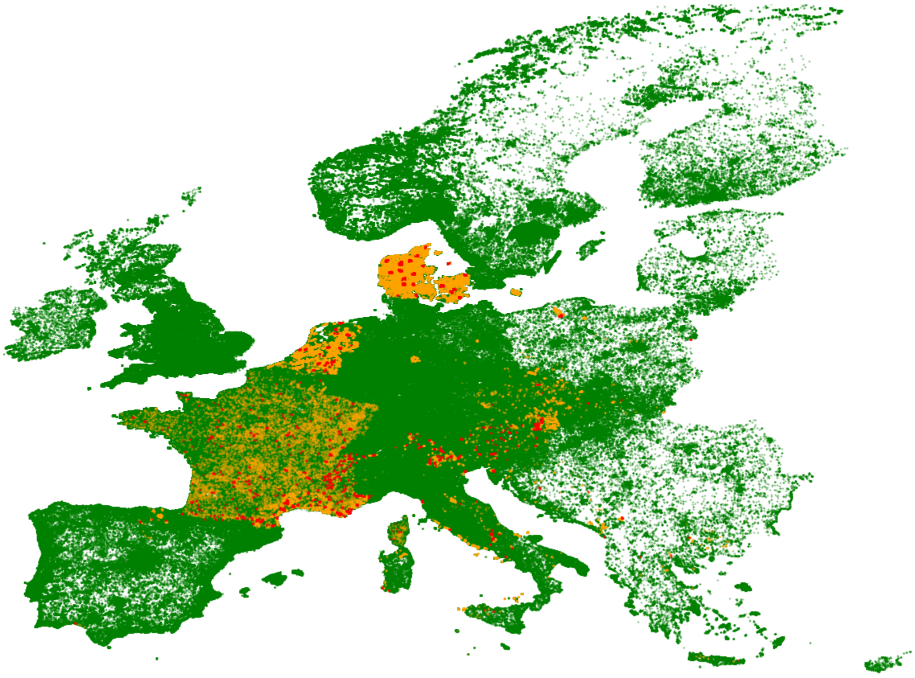

Figure 1: Survey latitude and longitude with PO surveys in green, PA training surveys in orange and PA test surveys in red.

3. Data

The training data comprises species observations and environmental data. Below, we explain the data in detail.

3.1. Observations data

The species-related training data for this study was gathered using two different methods: Presence-Absence (PA) surveys and Presence-Only (PO) occurrences. The different methodologies resulted in datasets that differed in both scale and utility.

The PA surveys included 88987 surveys with 5016 species from the European flora. This data was provided to address the issue of false absences in PO data and calibrate models to avoid associated biases.

The PO occurrences consisted of approximately five million observations gathered from various datasets available through the Global Biodiversity Information Facility (GBIF). This data covered all countries in the study area and constituted the larger portion of the training data. The PO also contained 11255 species, notably more than the PA. However, it was collected opportunistically, without a standardized sampling protocol, leading to various sampling biases. It is important to note that the local absence of a species in PO data does not necessarily indicate its true absence. An observer may have failed to report a species due to difficulties in seeing or identifying it at a particular time of year, or because it was not the target of monitoring efforts.

The testing data comprised of only PA surveys. The geographical spread of PA, PO and test data can be seen in Figure 1.

3.2. Environmental data

Besides species data, we had access to spatialized geographic and environmental data, these were used as additional input variables.

Satellite image patches which consist of 3-band (RGB) and 1-band (NIR) 128x128 JPEG images, a color JPEG file for RGB data and a grayscale one for Near-Infrared images at 10m resolution. The source for these images is Sentinel2 remote sensing data pre-processed by the Ecodatacube platform. An example of this data can be seen in Figure 2.

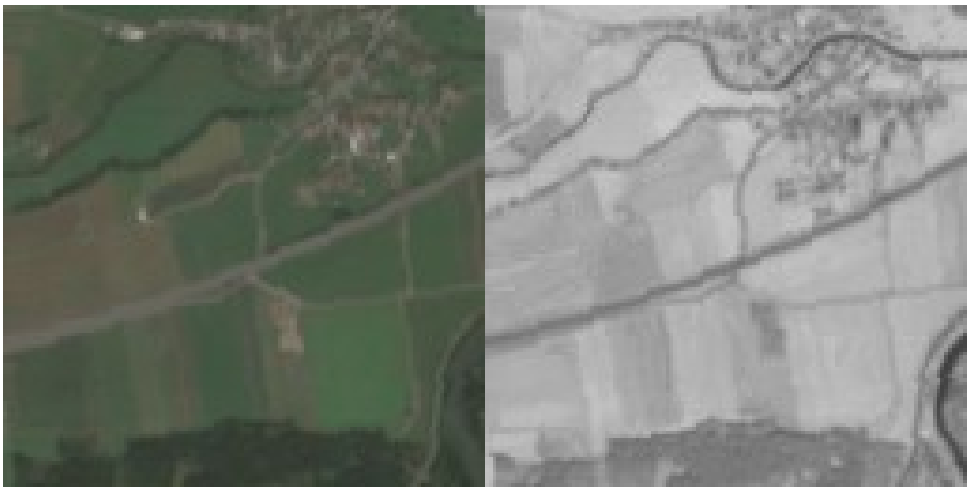

Figure 2: Satellite image patches. Left: Red, Green, Blue (RGB); Right: Infra-Red (IR).

Satellite time series comprises up to 20 years of values for six satellite bands (R, G, B, NIR, SWIR1, and SWIR2). Each observation is associated with the time series of the satellite median point values over each season since the winter of 1999 for six satellite bands (R, G, B, NIR, SWIR1, and SWIR2). This data carries a high-resolution local signature of the past 20 years' succession of seasonal vegetation changes, potential extreme natural events (fires), or land use changes. The original satellite data has a resolution of 30m per pixel. The source for this is the Landsat remote sensing data pre-processed by the Ecodatacube platform

Environmental rasters are formed of various climatic, pedologic, land use, and human footprint variables at the European scale. These were provided as scalar values, time-series, and the original rasters.

Environmental rasters, for each observation, we were provided additional environmental data such as GeoTIFF rasters and scalar values were already extracted from the rasters. We were provided CSV files, one per band raster type, i.e., Climate, Elevation, Human Footprint, LandCover, and SoilGrids. Further details can be found in Appendix A.

4. Method

For the creation of our proposed solution, a multi-stage development process was adopted. Initially, we performed data pre-processing. Next, we trained and tested individual models. In order to improve performance, an ensemble approach was then implemented by combining an XGBoost model, with a Multi-Modal model. The output was tuned to predict the number of species per survey using an additional XGBoost regression model. Finally, additional tests were performed using different weightings for each model in the ensemble step and different weightings added to the species count prediction. We also tested combinations of the same Multi-Modal model trained on different seeds.

4.1. Data Processing

In order to improve the model performance and reduce noise [9] the data was processed prior to use in the models. The flow of the data processing is shown in Figure 3.

4.1.1. Initial Processing

The raw PA and Environmental Rasters data included some missing, or infinite values that needed to be processed prior to model fitting. When a column was both deemed important to model fitting and contained such values, these values were replaced by median values so as to avoid excessive influence from outliers without overly effecting the shape of the data [10]. Both the Country and Region columns were one-hot encoded to remove the categorical variables. The following columns were then dropped from the data:

- Environmental Rasters from before 2008, retaining a 12 year window for further processing.
- All countries and regions with fewer than 250 occurrences, in order to minimize overfitting to the few occurrences present in the data.
- The combined Human Footprint data as well as that for Navigable water, Roads were removed, as they were each missing more than 5% of their values.

4.1.2. Principal Component Analysis

Principal Component Analysis (PCA) is a method for reducing dimensionality [11]. It functions by projecting the high dimensional data onto the direction of maximum variance, thus retaining key features and reducing noise. Single component PCAs were used on several of the Environmental Rasters:

- Monthly precipitation values were combined for the years 2008-2019, giving a single value per month.
- Monthly mean, minimum and maximum temperature values were combined for the years 2008-2019, giving a single value per month.

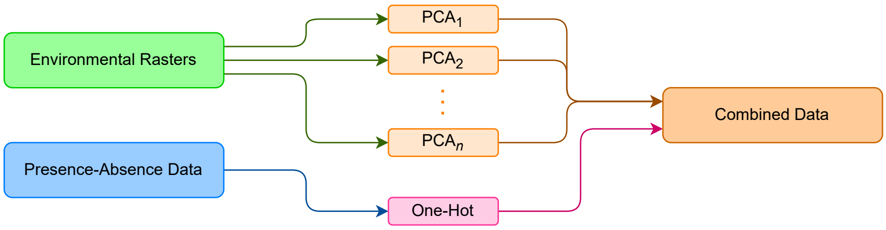

Figure 3: Pre-processing of the PA and Environmental Rasters data to create a combined dataset for the XGBoost models.

- Other high covariance columns (in Appendix Table 6) were combined using PCA into single values.

4.1.3. Species Reduction

The number of viable output species was reduced from 11255 (the total species present across all data) down to 1141 (the total species with more than 100 occurrences in the PA data). This reduction served to remove several edge cases and focus the models more on likelier species.

4.2. Models

To handle the multi-label classification necessary for this study, we employed an ensemble approach. Ensemble methods have been utilized to enhance neural network performance on challenging tasks [12]. Our ensemble consists of various model types, including an XGBoost model and a Multi-Modal model comprised of multiple neural networks, as well as multiple instances of the same Multi-Modal model trained with different seeds. The complex nature and stochastic initialization of the multi-modal model can result in its loss function settling into local minima during training. By using multiple instances of the same multi-modal model trained with different seeds and aggregating their outputs, we aim to improve performance at the cost of increased training time. The pseudo-probability outputs of the models that formed the ensemble were then weighted and combined. The species with the combined highest score were then selected with the number determined by a further XGBoost model referred to as the Count model. This technique of combining models in to an ensemble has been used in Kaggle competitions to optimize performance [13].

4.2.1. XGBoost

XGBoost is an open source gradient tree boosting package [14]. For this research, we used the XGB regression model. It has shown broad success in a wide range of tasks, performing on par with or better than most equivalent and Automated Machine learning approaches [15].

A multi-label regression XGBoost model was chosen as this generated predictions in the form of pseudo-probabilities for every species class. This also allowed the model to be used in an ensemble with other models, as the pseudo-probabilities could be combined in a weighted sum.

The XGBoost Regression model was trained on a combination of the processed data and the Landsat data. as shown in Figure 4. Hyperparameter tuning was performed using a Grid Search Cross Validation approach [16].

4.2.2. Multi-Modal

The multi-modal model is the Sentinel+Landsat+Bioclim baseline model [17]. This model uses ResNet18 to process all data except the satellite images which are processed by a Swin Transformer. This used the full PA data to train.

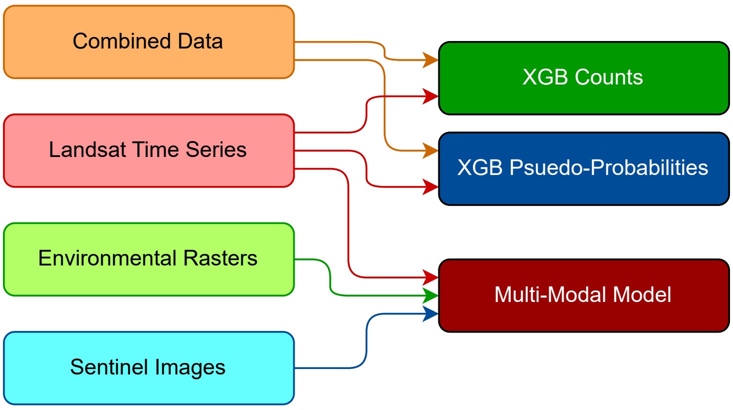

Figure 4: The flow of data into the models.

ResNet (Residual Network) is an architecture for deep neural networks that speeds up the training process through the use of residual connections [18]. For the purposes of this research we focused on ResNet18, the 18 layer deep variation with some modifications to accommodate the shape of the input data and the required output.

Swin Transformers, which were introduced in 2021 by researchers at Microsoft [19], are a type of vision transformer architecture that use a hierarchical structure to process images. Unlike traditional convolutional neural networks (CNNs), which use fixed-size filters to scan an image and extract features, Swin Transformers use multi-head self-attention mechanisms to dynamically learn relationships between different parts of an image. This allows them to capture more complex patterns and dependencies in the data.

Swin Transformers divide an image into smaller patches and process them in a hierarchical manner, starting with small patches and gradually merging them into larger ones. This allows the network to learn features at different scales and improve its performance on tasks that require fine-grained detail, as well as coarse-grained context. Swin Transformers have been shown to outperform other state-of-the-art vision transformer architectures on several benchmark datasets, making them a promising approach for computer vision applications.

4.3. Count Prediction

In order to select the number of species per survey two methods will be used.

Firstly, a top-K approach using fixed values in the range $[\lfloor \mu_{counts} \rfloor, 25]$. This range was selected as the mean count μ_{count} provided a good lower bound and 25 allowed us to test against the baseline models provided by the GeoLifeCLEF organizers. [1].

Secondly a dynamic approach was used based on a XGBoost Regression model trained on the pre-processed PA-Metadata and Environmental Rasters. The training labels used were the count of species recorded in each PA survey. The predicted counts $\hat{\mathbf{y}}$ then had a weight added and were rounded to the nearest integer as shown in equation 1.

$$\mathbf{c} = \lfloor \hat{\mathbf{y}} + n + 0.5 \rfloor \tag{1}$$

The weight n was consistent across all surveys and the values used were [0, 2, 2.5, 3, 3.5, 4, 4.5, 5]. These values were selected as during initial testing there was consistently a drop-off from values above 4.5 or below 2.5. The raw predicted count was also included to demonstrate the value of adding an n.

4.4. Scoring Metric

The main scoring metric used was the micro-averaged F1 score as shown in equation 2 which is calculated from the precision $\frac{\text{TP}}{\text{TP+FP}}$ and the recall $\frac{\text{TP}}{\text{TP+FN}}$ for each individual class i. Where TP are the true positives, FP are the false positives, and FN are the false negatives.

$$\text{F1}_{micro} = \frac{1}{N}\sum_{i=1}^{N}\frac{2\cdot\text{TP}_i}{2\cdot\text{TP}_i+\text{FP}_i+\text{FN}_i} \tag{2}$$

5. Results

Table 1
F1 scores per model using the top N species selection.

Model	15	17	19	21	23	25
XGB Regression	0.30664	0.30986	**0.31035**	0.31025	0.30958	0.30826
Multi-Modal Model	0.30876	0.31216	0.31420	0.31531	**0.31577**	0.31514
Ensemble Model 1:1	0.32509	0.32838	0.32967	***0.32972***	0.32859	0.32699

Table 2
F1 scores per model using the Count Prediction model with additional weighting.

Model	Raw	+2	+2.5	+3	+3.5	+4	+4.5	+5
XGB Regression	0.31957	0.32178	0.32214	**0.32234**	0.32217	0.32233	0.32198	0.32184
Multi-Modal Model	0.32165	0.32523	0.32566	0.32664	0.32721	**0.32762**	0.32750	0.32751
Ensemble Model 1:1	0.34078	0.34302	0.34335	**0.34378**	0.34323	0.34329	0.34316	0.34338
Ensemble Model 2:3	0.34077	0.34306	0.34297	0.34304	0.34303	**0.34341**	0.34289	0.34093
Ensemble Model 5:4	0.34116	0.34349	0.34340	0.34396	***0.34407***	0.34402	0.34378	0.34372

5.1. Top-K Species Counts

The Top-K approach was used to provide a benchmark for each model approach. Here the base XGB Regression and Multi-Modal models were tested as well as an equally weighted combination of the two. The results for this are shown in Table 1. The optimal configuration was using the equally weighted ensemble model with a K value of 21. This led to an F1 score of 0.32927.

5.2. XGBoost Species Counts

The second round of testing used the dynamic XGBoost Count model in place of Top-K species selection. In addition to the models tested with Top-K, different weightings of the XGB Regression to Multi-Modal models were tested as shown in table 2. Here the model weighting is denoted as X:Y, where the X weight is applied to the Multi-Modal model and the Y weight to the XGB Regression model. An additional weight was applied to the XGBoost Count model on top of the predicted species count.

The best performance was achieved with the Ensemble Model with a 5:4 weighting, using a species weight of +3.5. This resulted in an F1 score of 0.34407, an increase of 0.01435 over the best Top-K model result.

5.3. Multiple Multi-Modal Models

Using a combination approach, with a mean prediction between multiple Multi-Modal models trained with different seeds, as shown in Table 3, outperformed the other approaches. The best result was achieved with a mean output of 6 different seeds equally weighted against the XGB Regression model,

Table 3
Mean F1 scores across 3 tests per ensemble using a 1:1 split of the mean output of Multi-Modal models and the output of the XGB Regression model, with additional species count weighting.

Multi-Modal Count	+3	+3.5	+4
2	0.349850	0.349743	**0.350037**
3	0.353183	0.353063	**0.353303**
4	**0.353797**	0.353547	0.353637
5	**0.355133**	0.354817	0.354873
6	***0.355177***	0.354803	0.354643

using a species count weight of +3. This gave a final F1 score of 3.55177, and increase of 0.011107 over the best single Multi-Modal model result and 0.025457 over the best Top-K model result. A more detailed breakdown of these results, including seeds, can be found in Appendix Table 7.

6. Conclusion

The data presented in Tables 1 and 2 suggest that using values higher than the mean species count of 15.396 and the predicted species count of 15.762 consistently resulted in better performance. In our tests a value of 3.5 added to the XGB Count prediction values provided the best result. This can in part be attributed to the use of the micro F1 score as the evaluation metric. This metric prioritizes models that are able to accurately predict a higher number of true positives even if it comes at the cost of introducing additional false positives, as the number of false negatives decreases. The results also show how combining multiple independent models can be better than the sum of their parts.

Increasing the number of instances (trained with different seeds) of the Multi-Modal models in the ensemble further improved the results. Each additional model improved the score, albeit with diminishing returns. The addition of further models comes at a cost of time and computational complexity. The required resources for training the additional models scales linearly with each additional model added, but the results only add marginal improvements. Whilst testing higher numbers of additional models may improve the score further, we would suggest the testing of different optimizers and tuning hyperparameters may allow a single model to achieve similar or better results without the computational penalty. There is also the possibility for improvements through adjusting the sub-models within the Multi-Modal model. Alternatively using larger ResNet models in the Multi-Modal model trained over a greater number of epochs could potentially enhance performance.

Another potential area for further investigation is the acquisition of additional data. In this study, the plant data was anonymized in order to protect sensitive information about rare species, which may have limited the ability to analyze species-specific performance. Future datasets that include information about plant families or genera could provide valuable context for improving model performance in specific areas. Additionally, the Presence Only data was not utilized in this investigation. With proper pre-processing, this information could be used to refine and optimize our models.

This research has laid the groundwork for future studies with more time and access to greater computational or data resources to build upon.

References

[1] L. Picek, C. Botella, M. Servajean, B. Deneu, D. Marcos Gonzalez, R. Palard, T. Larcher, C. Leblanc, J. Estopinan, P. Bonnet, A. Joly, Overview of GeoLifeCLEF 2024: Species presence prediction based on occurrence data and high-resolution remote sensing images, in: Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum, 2024.

[2] A. Joly, L. Picek, S. Kahl, H. Goëau, V. Espitalier, C. Botella, B. Deneu, D. Marcos, J. Estopinan, C. Leblanc, T. Larcher, M. Šulc, M. Hrúz, M. Servajean, et al., Overview of lifeclef 2024: Challenges

on species distribution prediction and identification, in: International Conference of the Cross-Language Evaluation Forum for European Languages, Springer, 2024.

[3] A. Joly, T. Lorieul, E. Cole, B. Deneu, M. Servajean, P. Bonnet, Overview of geolifeclef 2021: Predicting species distribution from 2 million remote sensing images, in: CLEF (Working Notes), 2021, pp. 1451–1462.

[4] S. Seneviratne, Contrastive representation learning for natural world imagery: Habitat prediction for 30,000 species, in: CLEF-2021, 2021, pp. 1639–1648. URL: http://ceur-ws.org/Vol-2936/#paper-140.

[5] A. Joly, T. Lorieul, E. Cole, B. Deneu, M. Servajean, P. Bonnet, Overview of geolifeclef 2022: Predicting species presence from multi-modal remote sensing, bioclimatic and pedologic data, in: CLEF (Working Notes), 2022, pp. 1940–1956.

[6] B. Kellenberger, T. Devis, Block label swap for species distribution modelling, in: CLEF2022, 2022, pp. 2103–2114. URL: http://ceur-ws.org/Vol-3180/#paper-167.

[7] C. Botella, B. Deneu, D. M. Gonzalez, M. Servajean, T. Larcher, C. Leblanc, J. Estopinan, P. Bonnet, A. Joly, Overview of geolifeclef 2023: Species composition prediction with high spatial resolution at continental scale using remote sensing, in: CLEF 2023: Conference and Labs of the Evaluation Forum, 2023.

[8] H. Q. Ung, R. Kojima, S. Wada, Leverage samples with single positive labels to train cnn-based models for multi-label plant species prediction, in: CLEF2023, 2023, pp. 2149–2158. URL: http://ceur-ws.org/Vol-3497/#paper-181.

[9] A. Famili, W.-M. Shen, R. Weber, E. Simoudis, Data preprocessing and intelligent data analysis, Intelligent data analysis 1 (1997) 3–23.

[10] E. Acuna, C. Rodriguez, The treatment of missing values and its effect on classifier accuracy, in: Classification, Clustering, and Data Mining Applications: Proceedings of the Meeting of the International Federation of Classification Societies (IFCS), Illinois Institute of Technology, Chicago, 15–18 July 2004, Springer, 2004, pp. 639–647.

[11] K. Pearson, On lines and planes of closest fit to systems of points in space, in: Proceedings of the 17th ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems (SIGMOD), 1901, p. 19.

[12] M. P. Perrone, L. N. Cooper, When networks disagree: Ensemble methods for hybrid neural networks, in: How we learn; How we remember: Toward an understanding of brain and neural systems: Selected papers of Leon N Cooper, World Scientific, 1995, pp. 342–358.

[13] N. Ketkar, E. Santana, Deep learning with Python, volume 1, Springer, 2017.

[14] T. Chen, C. Guestrin, XGBoost: A scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, ACM, New York, NY, USA, 2016, pp. 785–794. URL: http://doi.acm.org/10.1145/2939672.2939785. doi:10.1145/2939672.2939785.

[15] L. Ferreira, A. Pilastri, C. M. Martins, P. M. Pires, P. Cortez, A comparison of automl tools for machine learning, deep learning and xgboost, in: 2021 International Joint Conference on Neural Networks (IJCNN), 2021, pp. 1–8. doi:10.1109/IJCNN52387.2021.9534091.

[16] M. Adnan, A. A. S. Alarood, M. I. Uddin, I. ur Rehman, Utilizing grid search cross-validation with adaptive boosting for augmenting performance of machine learning models, PeerJ Computer Science 8 (2022) e803.

[17] L. Picek, Sentinel+landsat+bioclim baseline [0.31626], 2024. URL: https://www.kaggle.com/code/picekl/sentinel-landsat-bioclim-baseline-0-31626.

[18] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, CoRR abs/1512.03385 (2015). URL: http://arxiv.org/abs/1512.03385. arXiv:1512.03385.

[19] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, CoRR abs/2103.14030 (2021). URL: https://arxiv.org/abs/2103.14030. arXiv:2103.14030.

A. Appendix — Data

Country & Region Data

Table 4
The number of surveys per country in the training and test data.
* Data removed for the XGBoost Model.

Country	Training Data	Test Data
Andorra*	59	0
Austria	43821	204
Belgium	31620	71
Bosnia and Herzegovina*	1428	7
Bulgaria*	2738	2
Croatia*	4837	42
Czech Republic	24481	10
Denmark	791993	1453
France	247739	1386
Germany	24701	2
Greece*	2268	4
Hungary*	1355	0
Ireland*	107	0
Italy	46625	433
Latvia	11723	0
Luxembourg*	38	0
Monaco*	150	0
Montenegro	20652	19
Netherlands	179457	846
North Macedonia*	122	0
Norway*	36	0
Poland	13697	75
Portugal*	79	0
Romania	8114	0
Serbia*	1803	0
Slovakia*	4850	4
Slovenia*	1104	39
Spain	11574	16
Switzerland	6466	103

Table 5
The number of surveys per region in the training and test data.
* Data removed for the XGBoost model. ** Removed through country filtering.

Region	Training Data	Test Data
ALPINE	1838	773
ATLANTIC	36357	1402
BOREAL**	73	0
BLACK SEA*	8	0
CONTINENTAL	43147	1449
MEDITERRANEAN	7333	1092
PANNONIAN*	219	0
STEPPIC*	12	0

Environmental Rasters

Bioclimatic rasters: 19 low-resolution rasters covering Europe; commonly used in species distribution modeling. Provided in longitude/latitude coordinates (WGS84). These were provided as GeoTIFF files with compression and CSV file with extracted values, with a resolution of 30 arcsec ($\sim$ 1 kilometer). The source for these rasters is the CHELSA climate dataset.

Soil rasters: Nine pedologic low-resolution rasters covering Europe. Provided variables describe the soil properties from 5 to 15cm depth and are determinant of plant species distributions. These included numerous values such as pH, clay, organic carbon and nitrogen contents. The format is GeoTIFF files with compression and CSV file with extracted values, with a resolution of $\sim$ 1 kilometer. The source for these rasters is Soilgrids.

Elevation: High-resolution raster covering Europe. Provided as a GeoTIFF file and CSV file with extracted values, with a resolution of 1 arc second ($\sim$ 30 meters). The source for this raster is the ASTER Global Digital Elevation Model V3.

Land Cover: A medium-resolution multi-band land cover raster covering Europe. Each band describes either the land cover class prediction or its confidence under various classifications. The format is GeoTIFF file with compression and CSV file with extracted values, with a resolution of $\sim$ 500 meters. The source for this raster is MODIS Terra+Aqua 500m.

Human footprint: Several low-resolution rasters describing human footprint, encapsulating seven pressures on the environment (e.g., nighlight level, population density) induced by human presence and activity, are provided for two time periods, the early 90's ($\sim$ 1993) and late 2000's ($\sim$ 2009). The format is GeoTIFF files with compression and CSV file with extracted values, with a resolution of $\sim$ 1 kilometer.

PCA Columns

Table 6

High Covariance columns combined with PCA.

For the columns with months denoted by MM PCA was applied per month with year values YYYY in the range 2008-2019, giving 12 precipitation and 12 temperature columns of PCA output values.

Column
HumanFootprint-Built1994
HumanFootprint-Built2009
HumanFootprint-croplands1992
HumanFootprint-croplands2005
HumanFootprint-Lights1994
HumanFootprint-Lights2009
HumanFootprint-Pasture1993
HumanFootprint-Pasture2009
HumanFootprint-Popdensity1990
HumanFootprint-Popdensity2010
HumanFootprint-Railways
Bio-tas_MM_YYYY
Bio-tasmax_MM_YYYY
Bio-tasmin_MM_YYYY
Bio-pr_MM_YYY

B. Appendix — Extended Results

Table 7
F1 scores of mean outputs of Multi-Modal models weighted 1:1 with the XGB Regression output.

Addition	Number of Networks	Seeds	F1-Score
3	2	489304 556925	0.34836
3.5	2	489304 556925	0.34802
4	2	489304 556925	0.34836
3	2	481765 356592	0.35082
3.5	2	481765 356592	0.35076
4	2	481765 356592	0.35084
3	2	489304 356592	0.35037
3.5	2	489304 356592	0.35045
4	2	489304 356592	0.35091
3	3	356592 20438 489304	0.35337
3.5	3	356592 20438 489304	0.35351
4	3	356592 20438 489304	0.35366
3	3	278872 481765 831651	0.35346
3.5	3	278872 481765 831651	0.35352
4	3	278872 481765 831651	0.35346
3	3	356592 937905 216510	0.35272
3.5	3	356592 937905 216510	0.35216
4	3	356592 937905 216510	0.35279
3	4	278872 937905 556925 216510	0.35349
3.5	4	278872 937905 556925 216510	0.35329
4	4	278872 937905 556925 216510	0.35330
3	4	278872 216510 831651 556925	0.35306
3.5	4	278872 216510 831651 556925	0.35269
4	4	278872 216510 831651 556925	0.35290
3	4	216510 356592 489304 831651	0.35484
3.5	4	216510 356592 489304 831651	0.35466
4	4	216510 356592 489304 831651	0.35471
3	5	20438 489304 831651 729571 216510	0.35502
3.5	5	20438 489304 831651 729571 216510	0.35452
4	5	20438 489304 831651 729571 216510	0.35427
3	5	481765 937905 216510 356592 489304	0.35512
3.5	5	481765 937905 216510 356592 489304	0.35478
4	5	481765 937905 216510 356592 489304	0.35534
3	5	216510 831651 481765 356592 489304	0.35526
3.5	5	216510 831651 481765 356592 489304	0.35515
4	5	216510 831651 481765 356592 489304	0.35501
3	6	356592 729571 937905 831651 278872 20438	0.35598
3.5	6	356592 729571 937905 831651 278872 20438	0.35540
4	6	356592 729571 937905 831651 278872 20438	0.35509
3	6	556925 356592 831651 481765 489304 20438	0.35544
3.5	6	556925 356592 831651 481765 489304 20438	0.35507
4	6	556925 356592 831651 481765 489304 20438	0.35493
3	6	556925 356592 278872 20438 216510 937905	0.35411
3.5	6	556925 356592 278872 20438 216510 937905	0.35394
4	6	556925 356592 278872 20438 216510 937905	0.35391

Addition: The addition to the count prediction used to select species
Number of Networks: The number of unique seeds of the same network used in combination

C. Appendix — Hyper-parameters

All models were trained on GPUs using CUDA.

Table 8
Hyper-parameters used to train the XGBoost models.

Hyper-Parameter	Pseudo-Probability	Count
reg lambda	12	10
learning rate	0.1	0.1
min split loss	0	0
reg alpha	0	0

Table 9
Hyper-parameters used to train the Multi-Modal model.

Hyper-Parameter	Value
learning rate	0.00025
epochs	10
positive weight factor	1.0
training batch size	64
optimizer	Adam W
scheduler	Cosine Annealing LR
T$_{max}$	25