

Quantifying Domain Shift in Foundation Models for Ground Vegetation Imagery

Stephen Foy^{1,2}, Simon McLoughlin²

¹*Atlantic Technological University, Galway, Ireland*

²*Technological University Dublin, Dublin, Ireland*

Abstract

This paper describes Team Atlantic’s work for the PlantCLEF 2026 Kaggle challenge, emphasising the role of domain adaptation in downstream species prediction. The challenge aims to improve ecological survey workflows through the automated analysis of high-resolution quadrat images. In PlantCLEF 2024, Team Atlantic ranked first on the official leaderboard, and the present work builds on that earlier participation. However, the focus of the 2026 paper is not solely on maximising leaderboard performance. Instead, it investigates how the domain gap between single-plant training images and multi-species vegetation plot images can be measured using foundation-model feature embeddings. In PlantCLEF 2026, the team ranked 22nd with a score of 0.34103 under the official macro-averaged per-sample F1 evaluation protocol. In this study, we implemented a lightweight morphotype-classification head atop the challenge-provided DINOv2 model to enable a systematic characterization and quantification of domain shift between a single-plant source domain and a vegetation-plot target domain. This audit compares the single-plant source imagery distributed across the PlantCLEF challenge lineage with the 2026 target quadrat imagery using frozen DINOv2 embeddings. The results show that source and target images are nearly perfectly separable in embedding space, even when mask-guided crops are used. They also suggest that crop-level adaptation, temporal consistency across repeated plots, and biologically informed priors over broad plant form are promising directions for improving transfer. Overall, this work suggests that models trained on large single-plant image collections should not be expected to transfer directly to community-level ecological inventory tasks without addressing both representation-level and label-level shift. Our code is available at <https://github.com/stevefoy/AtlanticAnalytica-PlantCLEF2026>

Keywords

Domain analysis, Deep learning, Plant detection, Domain adaptation, Vision transformer, Machine vision

1. Introduction

Quadrat sampling is a method employed by botanists to study plant species composition within a defined, limited area. The method involves placing a square or rectangular frame of known dimensions on the ground and manually recording the presence, identity, and abundance of plant species within the enclosed region. Quadrat-based observations support biodiversity assessment, long-term monitoring, habitat mapping, invasive species detection, restoration studies, and broader ecological analysis [1]. Traditional work has often focused on single-species patterns, whereas recent research has increasingly emphasised the need to understand multi-species structure and co-occurrence within plant communities [2, 3].

Automated plant recognition has made rapid progress through deep learning, especially in challenges such as PlantCLEF, where the scale of the data and the taxonomic diversity favour representation learning approaches over handcrafted features [4, 5, 6]. In the large-scale PlantCLEF challenges conducted in 2022, 2023 Vision Transformer (ViT) architectures consistently achieved superior performance compared to traditional convolutional neural network models [7]. The PlantCLEF 2024, 2025 and 2026 vegetation-plot editions then shifted the problem from single-plant recognition toward ecological inventory from community images, thereby placing much stronger emphasis on transfer learning and domain mismatch [8, 1]. Following the naming convention used in the PlantCLEF working notes,

CLEF 2026 Working Notes, 21–24 September 2026, Jena, Germany

✉ stephen.foy@atu.ie (S. Foy); simon.d.mcloughlin@tudublin.ie (S. McLoughlin)

🌐 <https://stevefoy.github.io/> (S. Foy)

🆔 0009-0001-9736-2264 (S. Foy); 0000-0002-9846-4617 (S. McLoughlin)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

we refer to the released DINOv2-based PlantCLEF classifiers as ViTD2PC24OC and ViTD2PC24All, where ViTD2PC24OC denotes the checkpoint with only the PlantCLEF classifier head trained and ViTD2PC24All denotes the checkpoint subsequently fine-tuned across all layers.

2. Related Work

Recent PlantCLEF working note papers show that successful vegetation-plot systems rarely rely on a single full-image prediction strategy. Numerous top-performing methods pair the common ViTD2PC24All model with additional pre- and post-processing techniques, including further fine-tuning through tiling or sliding-window inference, visual clustering, ecological priors, and segmentation-based filtering [9, 10, 11, 12]. These workflows motivate the present study because they show that quadrat analysis requires more than applying a strong classifier to the full image. Performance depends on region-level evidence, aggregation, and post-processing, while the persistent errors point to a substantial mismatch between the single-plant source domain and the multi-species target domain. We study the problem in an unsupervised domain-shift setting, where target-domain images are available without species labels.

In our earlier participation in PlantCLEF 2024, Team Atlantic achieved first place on the official private leaderboard with a sample-wise F1 score macro-averaged across multiple images of 0.2873 per sample/plot. In PlantCLEF 2026, the submission achieved 22nd place on the private leaderboard, with a score of 0.34103 under the official macro-averaged per-sample F1 evaluation protocol. In our earlier PlantCLEF 2024 study, we designed an inference pipeline which was ranked in first position and was based on the ViTD2PC24All PlantCLEF fine-tuned ViT classifier supplied by the challenge organisers [13]. The system used multi-scale sliding-window inference, probabilistic scoring, and segmentation-based filtering to improve species detection in high-resolution vegetation-plot images and to reduce false positives caused by soil, rocks, and other non-vegetation artefacts. This approach addressed the mismatch between the labelled source domain, which largely consisted of single-plant images captured from varied perspectives, and the target domain, which consisted of top-down, multi-species vegetation-plot images.

As in the previous vegetation-plot editions, the target images are available during development, but their species labels are hidden. This allows for the direct study of the target image distribution while treating the prediction task as unsupervised in relation to target labels. However, our earlier work treated domain adaptation primarily as an empirical pipeline problem, without directly measuring the mismatch between source and target representations. This paper aims to address that limitation through a more systematic analysis of the source-target gap. This focus also extends our previous work on turfgrass abnormality analysis, where representation learning was used to assess visual variation and changes in the condition of ground-level vegetation imagery [14].

The present work continues this line of research but shifts the emphasis from challenge-oriented inference optimisation to a systematic analysis of the domain mismatch itself. Here, the source domain is the labelled single-plant training distribution, while the target domain is the unlabelled vegetation-plot imagery used for evaluation. Rather than invoking domain shift as a general explanation for poor transfer, we measure where and how this mismatch appears in the representation space of the ViTD2PC24All PlantCLEF fine-tuned ViT classifier. We use frozen embeddings from this model as a diagnostic representation and evaluate the source-target gap through a series of controlled experiments, including same-domain controls, mask-guided target crops, crop-scale variation, morphotype probing, JPEG re-compression, and source-likeness scoring. This allows us to separate simple preprocessing effects from deeper representation-level differences between single-plant source images and quadrat-based ecological imagery.

3. Task Description

The experiments reported in this paper use data from the PlantCLEF 2026 Kaggle challenge [15], which forms part of the wider LifeCLEF 2026 evaluation campaign on AI challenges for biodiversity understanding and ecosystem management [16]. The source to target domain structure is therefore closely aligned with the setting described in recent PlantCLEF working-note papers, where participants combine single-plant training data, high-resolution quadrat test images, and pre-trained DINOv2 models supplied by the organisers [9, 10, 12]. The aim is to predict the full list of plant species present in high-resolution photographs of vegetation plots. This differs from conventional plant-identification benchmarks, where images usually depict a single plant or plant organ, because each quadrat image may contain multiple species at different scales, with overlapping leaves, seasonal change, soil, stones, shadows, and plot-frame artefacts.

The 2026 target data consist of top-view vegetation-plot images collected in several floristic contexts, including Pyrenean and Mediterranean vegetation, and produced under expert ecological survey protocols [17]. In the local Kaggle package used for inference, the target set contains 2,105 JPEG images. Each image represents an approximately 50×50 cm quadrat, or about 0.25 m^2 , and is typically around 3065×3065 pixels. The acquisition protocol often includes a wooden frame or measuring tape that delimits the plot. These non-plant structures, together with heterogeneous backgrounds and variable lighting, make direct transfer from single-plant imagery difficult.

The model-development data follow the complementary source domain used in recent PlantCLEF vegetation-plot challenges. The single-plant training set is a subset of Pl@ntNet observations focused on southwestern Europe, covering approximately 7.8k vascular plant species and about **1.4 million** labelled images [18].

The primary challenge metric is the macro-averaged sample-wise F1 score aggregated at the transect level. For each quadrat or plot image, the predicted species set is compared with the corresponding ground-truth inventory to compute a sample-level F1 score. These scores are first averaged across all images belonging to the same transect, and the final score is then computed as the macro-average across transects:

$$F1_{\text{avg-transect}} = \frac{1}{T} \sum_{k=1}^T \left(\frac{1}{Q_k} \sum_{j=1}^{Q_k} F1_j^{(k)} \right)$$

where T is the number of transects, Q_k is the number of quadrat or plot images associated with transect k , and $F1_j^{(k)}$ is the sample-wise F1 score for the j -th image in transect k . This hierarchical averaging prevents transects with more images from dominating the final score and better reflects the ecological structure of the evaluation protocol. The metric penalises both overprediction and underprediction, making it suitable for multi-label vegetation inventories [15].

3.1. Classifier Backbone and Source Data

The challenge organisers released two PlantCLEF 2024 species-classification checkpoints derived from a DINOv2 ViT-B/14 backbone with register tokens [19, 18]. Following the naming convention used in the PlantCLEF working notes, we refer to `vit_base_patch14_reg4_dinov2_lvd142m_onlyclassifier` as ViTD2PC24OC and to `vit_base_patch14_reg4_dinov2_lvd142m_onlyclassifier_then_all` as ViTD2PC24All. ViTD2PC24OC keeps the self-supervised DINOv2 backbone frozen and trains only the PlantCLEF species-classification head, whereas ViTD2PC24All is initialised from ViTD2PC24OC and then fully fine-tuned on the PlantCLEF 2024 single-plant data. Both checkpoints use a classification head over **7,806** source species classes and expect fixed-size image inputs of 518×518 pixels, which motivates the crop- and tile-based inference strategies used by many PlantCLEF systems, including ours.

For the domain-shift audit, we used a local copy of the labelled single-plant source data organised into species-specific folders. The training split stored in `\PlantCLEF2024\train` contains 7,806

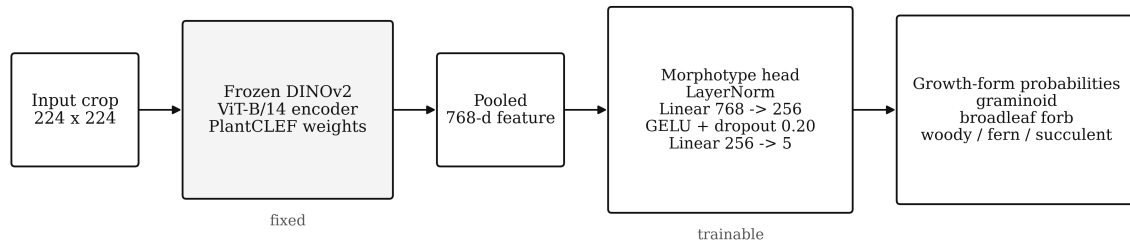


Figure 1: 5 Class, Morphotype head architecture

species folders and 1,308,901 image files, while the validation split in `val` contains 6,670 species folders and **51,194** image files. This distinction is important: the number 7,806 refers to the number of source species represented by the classifier, not to the number of training images. The source images are predominantly single-plant or organ-centred observations from the Pl@ntNet collaborative platform [20]. They differ substantially from the 2026 target quadrat images in viewpoint, composition, scale, background context, and label granularity. This mismatch is the core motivation for the domain-shift analysis presented in the next section.

4. Domain Shift Analysis

4.1. Data organisation and source–target setting

The source domain was derived from the single-plant PlantCLEF data used by the recent vegetation-plot challenge lineage and distributed as species-indexed folders in our local copy. Each subdirectory corresponds to a `species_id`, and each directory contains multiple image files for that species. As noted above, the training split contains **7,806** species folders and **1,308,901** image files, while the validation split contains **6,670** species folders and **51,194** image files. The source dataset is therefore highly imbalanced at image level, with some species represented by many more photographs than others.

4.2. Morphotype head

A lightweight five-class classifier was trained on the ViTD2PC24All model to predict broad plant-form categories, which we refer to as plant morphotype. The input resolution was changed from 518×518 to 224×224 . The morphotype classes were defined as follows: *graminoid* (grass-like, sedge-like, or rush-like plants), *broadleaf forb* (leafy non-woody herbs), *woody* (shrubs, trees, or dwarf shrubs), *fern*, and *succulent*. In a model-forward benchmark on an RTX 4080, the 224×224 morphotype model processed approximately 427 images/s, compared with approximately 61 images/s for the 518×518 PlantCLEF species classifier, corresponding to a roughly $7\times$ throughput increase. A control using the original 7,806-class species head at 224×224 achieved similar throughput, indicating that the gain is primarily due to reduced input resolution rather than the smaller **ViTD2PC24All+Morphotype** head.

The morphotype head consisted of a layer-normalisation stage followed by a two-layer multilayer perceptron: a $768 \rightarrow 256$ linear projection, GELU activation, dropout with $p = 0.20$, and a final $256 \rightarrow 5$ linear classifier over the five morphotype classes: *graminoid*, *broadleaf_forb*, *woody*, *fern*, and *succulent* (example plants in Figure 2). The model was trained for 20 epochs, with the best validation performance obtained at epoch 17. At this epoch, the model achieved a validation accuracy of 0.9291 and a validation macro-F1 score of 0.8401. At the best epoch, the per-class validation F1 scores were 0.8828 for *graminoid*, 0.9562 for *broadleaf_forb*, 0.7107 for *woody*, 0.8732 for *fern*, and 0.7778 for *succulent*. The trained morphotype model also provided a useful representation for qualitative inspection of domain structure. By extracting the pooled embedding before the morphotype classification head, the

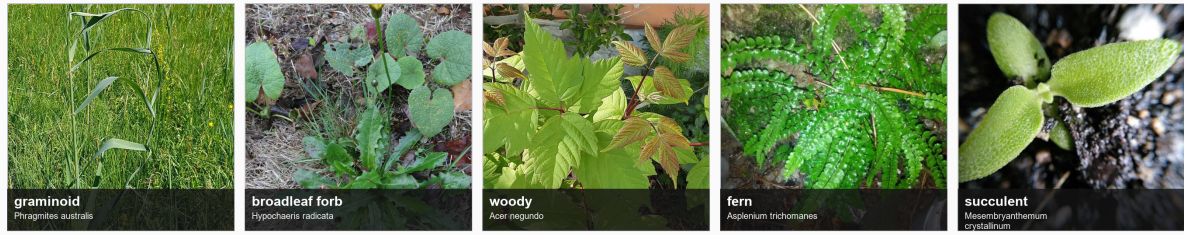


Figure 2: Representative examples of the five broad plant-form categories used in this study, selected from the single-plant source dataset: graminoid, broadleaf forb, woody, fern, and succulent.

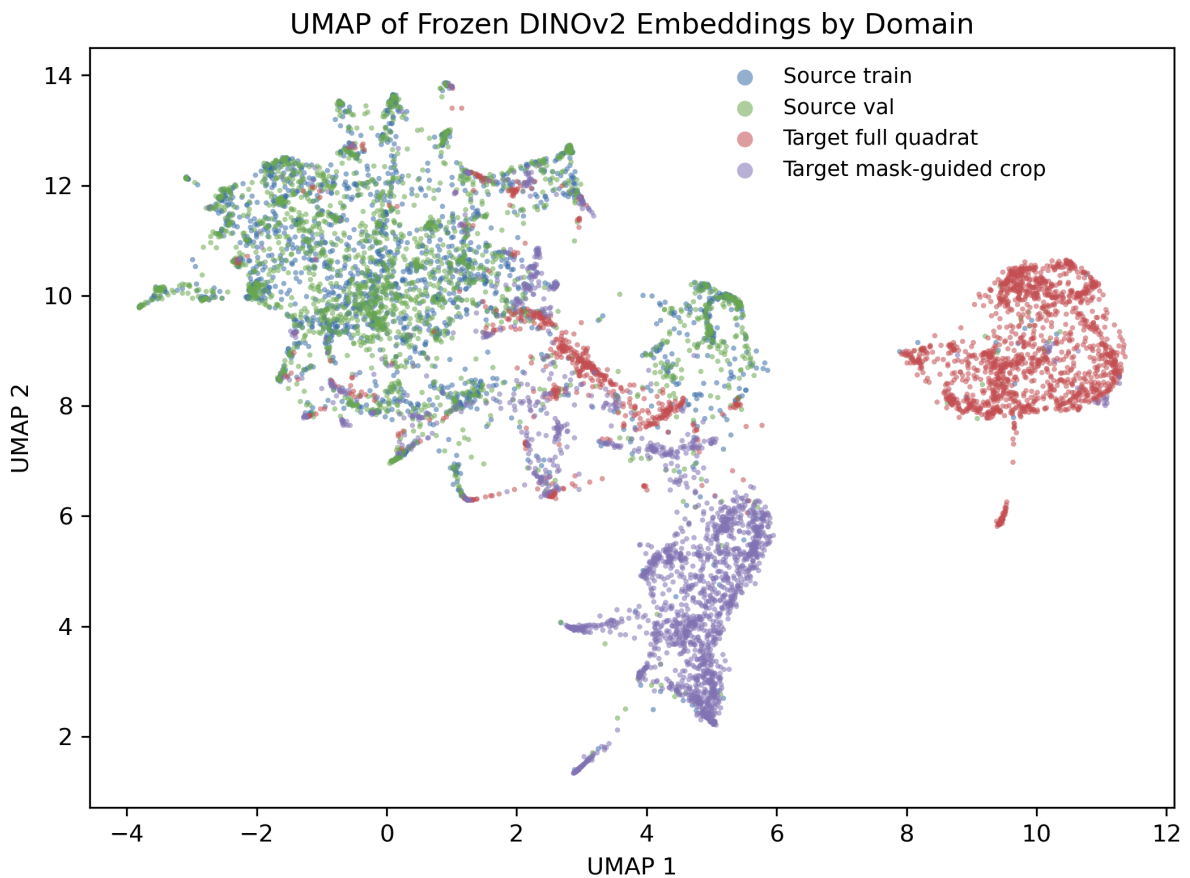


Figure 3: UMAP projection of pooled frozen DINOv2 embeddings extracted before the morphotype classification head. Points correspond to image samples and are coloured by domain or broad plant-form category. The projection illustrates the extent to which the learned representation separates different regions of the image distribution, providing a qualitative complement to the quantitative domain-shift metrics.

images could be projected into two dimensions using UMAP. As shown in Figure 3, this representation produced clear visual separation between parts of the data distribution, indicating that the learned feature space was sensitive to broad differences in plant form and image domain. This supported its use as an exploratory tool for identifying domain separation before applying the quantitative domain-shift metrics.

5. Evaluation Metrics and Diagnostic Indicators

To characterise the source and target domain mismatch, we used a set of complementary discrepancy measures computed from pooled frozen ViTD2PC24All embeddings extracted before the morphotype classification head. The first was the area under the receiver operating characteristic curve (AUC) of a binary domain classifier, which measures how easily source and target embeddings can be separated. Values close to 0.5 indicate substantial overlap, whereas values close to 1.0 indicate near-perfect separability. We also trained a lightweight morphology head on top of the same frozen backbone, using broad plant-form categories as a diagnostic probe of whether the source and target domains emphasised different types of visual plant structure.

Proxy A-distance was computed from the error rate of the binary domain classifier and used as an empirical measure of the divergence between the source and target embedding distributions. We also assessed local neighbourhood overlap using k -nearest-neighbour (k NN) mixing, defined as the proportion of nearest neighbours that come from the opposite domain. Higher k NN mixing therefore indicates greater local overlap between source and target samples.

To complement these classifier-based and local measures, we used Fréchet distance and maximum mean discrepancy (MMD) to compare the global structure of the embedding distributions. Fréchet distance measures differences in the means and covariance structure of the two domains, while MMD provides a kernel-based two-sample test for distribution mismatch. These metrics are not intended to predict leaderboard performance directly; rather, they provide complementary views of representation-level domain shift.

Recent work by Sorkhei et al. [21] examined domain adaptations. The work argues that k -nearest-neighbour accuracy in embedding space provides a simple and effective estimate of model transferability. Their large-scale evaluation, involving more than 42,000 experiments across 16 datasets. The results showed that nearest-neighbour score can outperform several established transferability metrics across a range of domains, tasks, and architectures. However, the original Sorkhei k NN score requires labelled target data, because it measures nearest-neighbour classification accuracy for the target task itself. In the PlantCLEF 2026 test setting, target species labels are not available. We therefore did not use the Sorkhei score directly. Instead, we used a related unsupervised k NN mixing measure in embedding space. This measure calculates the fraction of nearest neighbours that come from the opposite domain and was used as an estimate of local source–target overlap, rather than as a direct estimate of target-task transferability. Specifically, each image was represented using the pooled pre-logits representation from the final ViTD2PC24All layer, before the morphotype classification head. The morphotype logits and softmax probabilities were not used for the domain-mixing calculation.

5.1. Sampling strategy for the audit

Because the raw source training split contains over 1.3 million image files, a direct full-image audit would be computationally expensive and heavily biased toward species with many photographs. To obtain a computationally tractable and species-balanced estimate of domain shift, we constructed a reduced source audit sample by selecting one image per species folder from the source training split and **one image per species folder** from the validation split. This produced a balanced source audit set of **7,806 sampled training images** and **6,670** sampled validation images. These counts refer only to the audit sample and should not be confused with the total number of source image files available.

5.2. Feature extraction and metrics

Domain shift was quantified using the frozen PlantCLEF ViTD2PC24All backbone as a feature extractor. For each image, we extracted the DINO pre-logit embedding. The audit included four domains:

1. `source_train_singleplant`: one sampled image per species folder from the single-plant source training split;

Table 1

Quantitative domain-shift measurements computed on ViTD2PC24All+Morphotype embeddings.

Comparison	AUC	Proxy A-dist.	k NN Mixing	Fréchet	MMD
Source train (paired) vs. source train (paired)	0.5007	0.0228	0.5055	48.8	0.0012
Source train vs. source val	0.5070	0.0104	0.5004	49.5	0.0014
Source train vs. full quadrat	0.9985	1.9427	0.0906	486.7	0.1593
Source train vs. mask-guided crop	0.9992	1.9447	0.1158	391.7	0.1174
Full quadrat vs. mask-guided crop	1.0000	1.9935	0.0487	274.7	0.1350

2. `source_val_singleplant`: one sampled image per species folder from the single-plant source validation split;
3. `target_full_quadrat`: all 2,105 full quadrat images from the 2026 Kaggle challenge package;
4. `target_maskcrop_quadrat`: all 2,105 corresponding mask-guided crops.

A same-domain control (`source_train_singleplant` vs. `source_val_singleplant`) was included to calibrate the expected separability in the absence of a major domain shift.

5.3. Quantitative evidence of domain shift

As an additional sanity check, we compared two independently sampled image sets drawn from the same source training species folders, using different images from each folder whenever available. This paired within-split control again showed near-random separability (AUC = 0.5007), very small proxy A-distance (0.0228), balanced neighbourhood mixing (k NN mixing = 0.5055), low Fréchet distance (48.82), and low MMD (0.00118), confirming that the audit metrics remained stable under repeated sampling within the same source domain. The same-domain control behaved as expected. The comparison between sampled source training images and sampled source validation images yielded near-random separability, with a domain-classifier AUC of **0.5070**, proxy A-distance of **0.0104**, and k NN cross-domain mixing ratio of **0.5004**. This indicates that the feature representation does not artificially separate two subsets drawn from the same underlying single-plant source domain.

In contrast, the separation between the source single-plant images and the target quadrat imagery was extremely strong. Comparing sampled source training images to full quadrat images yielded a domain-classifier AUC of **0.9985** and a proxy A-distance of **1.9427**. This indicates that the two domains were almost perfectly separable in DINO embedding space. The comparison between sampled source training images and mask-guided crops yielded similarly strong separation, with AUC **0.9992** and proxy A-distance **1.9447**.

Distance-based measures showed the same pattern. Fréchet distance increased from **49.5** in the same-domain control to **486.7** for source versus full quadrats and **391.7** for source versus mask-guided crops. MMD increased from **0.0014** in the control setting to **0.1593** and **0.1174**, respectively. The k NN cross-domain mixing ratio fell from **0.5004** in the same-domain setting to **0.0906** for source versus full quadrats. Mask-guided crops improved this value slightly to **0.1158**, suggesting only partial local overlap between source and target neighbourhoods.

5.4. Lanczos and JPEG Target Normalisation

The work of Espitalier in 2025 challenge observed that effects of JPEG encoding. The encoding parameters constitute an additional source of domain shift in source domain versus the target domain. The PlantCLEF model was trained largely on images with specific JPEG quality and YCbCr subsampling characteristics, whereas the quadrat test images exhibited different compression properties. Running JPEG compression a second time significantly boosted the effectiveness of the top-performing method. Evidence presented in their work suggests that the model is sensitive to the entropy characteristics and color-channel statistical properties induced by the image encoding pipeline.[12].

Table 2

Domain-gap measurements for target normalisation variants using ViTD2PC24All+Morphotype embeddings. Each comparison uses a deterministic 500-image target-crop sample and an equally sized source sample. Lower AUC, proxy A-distance, Fréchet distance, and MMD, together with higher k NN mixing, indicate better source–target overlap.

Comparison	AUC	Proxy A-dist.	k NN Mixing	Fréchet	MMD
Source vs. original mask crop	0.9951	1.8693	0.1655	495.3	0.1230
Source vs. Lanczos 224	0.9932	1.8427	0.1705	486.7	0.1196
Source vs. Lanczos 224 + JPEG q85 4:2:2	0.9951	1.8587	0.1643	204.1	0.1222

Several high-performing Kaggle-style pipelines applied low-level image normalisation before tile inference, including high-quality resizing and JPEG re-compression. We therefore tested whether a similar preprocessing step changed the target-domain representation measured by the morphotype probe. For each target quadrat, we extracted the same mask-guided crop and compared three variants: the original crop, a 224×224 Lanczos-resized crop, and a 224×224 Lanczos-resized crop re-compressed as JPEG with quality 85 and YCbCr 4:2:2 chroma subsampling.

The experiment was carried out as a diagnostic test instead of as an official leaderboard entry. The small morphotype model was used because it provides a fast biological summary of broad plant-form evidence. Figure 4 shows that the preprocessing variants largely overlap in frozen DINOv2 embedding space, indicating that resize and re-compression alone do not create a separate target representation or remove the source–target domain gap. This morphotype-prior shift under target preprocessing is summarised in Figure 10.

Lanczos resizing slightly reduced AUC, proxy A-distance, Fréchet distance, and MMD, and slightly improved k NN mixing, suggesting a small movement toward the source distribution. Adding JPEG q85 4:2:2 did not further improve local overlap or classifier separability, although it substantially reduced the Fréchet distance. This indicates that JPEG re-compression changes global embedding moments, but does not by itself remove the source–target separability.

However, the morphotype probabilities changed measurably. On a deterministic sample of 500 target crops, the mean predicted graminoid probability increased from 46.7% for the original crop to 47.8% after Lanczos resizing and 50.1% after Lanczos plus JPEG re-compression. The broadleaf-forb probability decreased from 33.0% to 30.7% and then 27.8%, while woody probability increased slightly from 15.9% to 17.2%. Thus, low-level preprocessing did not solve the domain-shift problem, but it did alter the semantic prior exposed by the morphotype head.

This suggests that resize and compression choices can affect downstream ecological interpretation even when the embedding-space movement appears small. In practice, such preprocessing may be useful when combined with self-training or pseudo-labelling, where consistency of tile appearance can influence which target crops receive confident labels. In our setting, it is better interpreted as target normalisation than as sufficient domain adaptation.

5.5. Multi-scale tiling exploration

The sliding-window approach was used to address two practical limitations of direct full-image classification. First, many species occupy only a small region of the quadrat, so a single global crop may fail to preserve the relevant scale. Second, the ViTD2PC24All model was trained on fixed-size 518 inputs and therefore benefits from crops that better match the size and framing of the source data. Earlier detection systems such as OverFeat also used sliding-window style region generation for local object evidence [22].

The ViTD2PC24All model required an input resolution of 518×518 pixels with ImageNet-style normalisation. Adjustable parameters included window size, stride, scale factor, and border offset. A border offset was used to reduce interference from wooden frames and measuring devices at the edges of the quadrat. Each region proposal produced a ranked list of top species classes and associated confi-

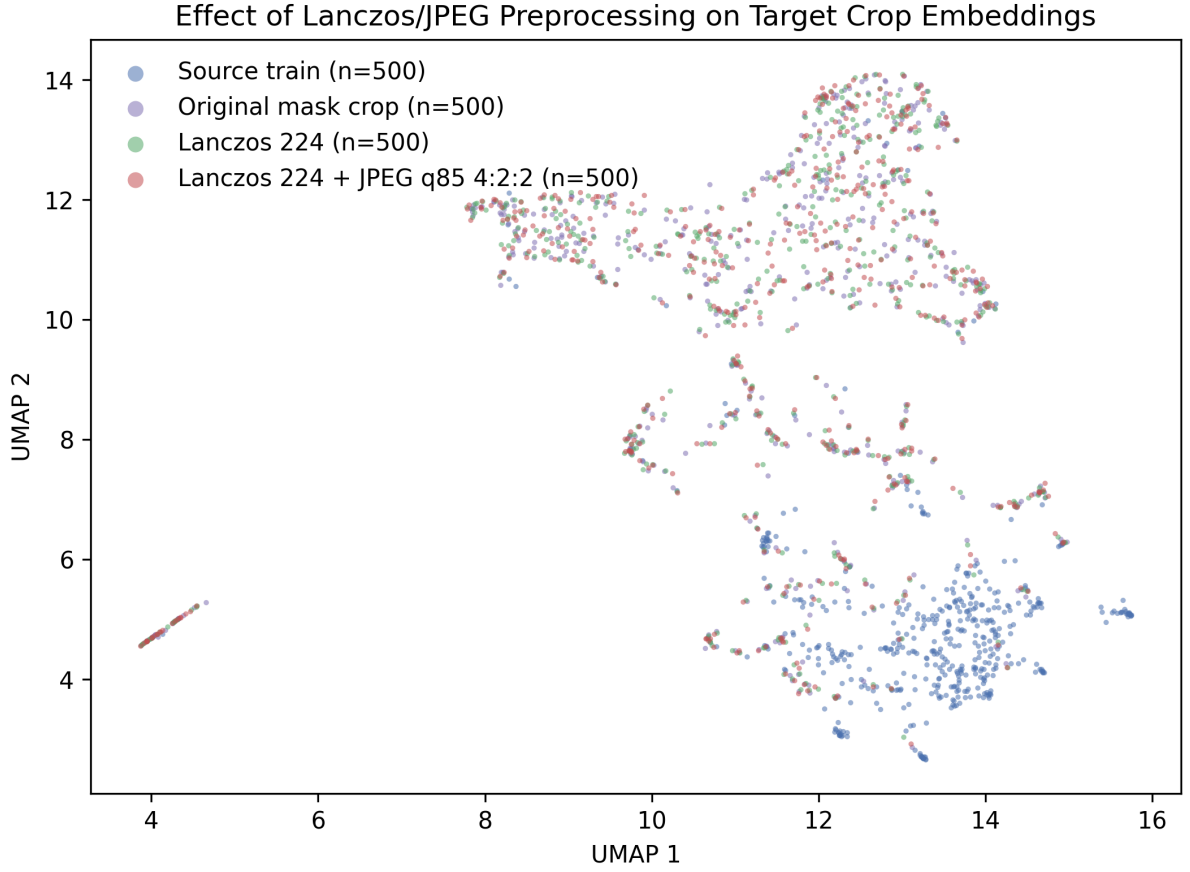


Figure 4: UMAP visualisation of target crop ViTD2PC24All+Morphotype embeddings under low-level preprocessing variants. Original mask-guided crops, Lanczos-resized crops, and Lanczos plus JPEG re-compressed crops largely overlap, suggesting that this preprocessing does not by itself remove the target-domain representation gap.

dence scores, which were stored in CSV format together with the crop coordinates. This representation allowed downstream aggregation across windows, images, and repeated plots.

The multi-scale crop audit showed that target crop size has a measurable effect on the representation learned by the frozen ViTD2PC24All backbone and on the biological prior exposed by the morphotype head (ViTD2PC24All+Morphotype). As the raw crop size increased from 259 to 1152 pixels, while the model input remained fixed at 224×224 , the predicted morphotype distribution shifted progressively away from broadleaf forb and toward graminoid and woody structure. The mean broadleaf-forb probability decreased from 40.2% at 259 pixels to 25.1% at 1152 pixels, whereas graminoid probability increased from 36.8% to 50.5% and woody probability increased from 16.5% to 20.3%. This suggests that crop scale changes not only the amount of visual context available to the model, but also the type of plant-form evidence that dominates the embedding. Smaller crops may emphasise local leaf texture and isolated herbaceous structures, while larger crops capture broader canopy patterns, grass-like cover, woody stems, and mixed community context. The domain-gap metrics followed a similar pattern: intermediate crop sizes produced slightly better source and target overlap, although all scales remained highly separable from the source domain. Thus, multi-scale tiling should be viewed not merely as a computational strategy for covering high-resolution images, but as a mechanism that changes the semantic and ecological prior presented to the classifier.

The multi-scale crop audit showed that target crop size has a measurable effect on the representation learned by the frozen DINOv2 backbone and on the biological prior exposed by the morphotype head. As shown in Table 3, very small crops were the most strongly separated from the source domain, with a domain AUC of 1.0000, proxy A-distance of 1.9883, and k NN mixing of only 0.0818 at a raw crop size

Table 3

Effect of raw target crop scale on source–target domain gap using embeddings. Anchor locations were selected with a 518×518 vegetation-mask window, while each raw crop was resized to the ViTD2PC24All+Morphotype model input size of 224×224 .

Raw crop size	AUC	Proxy A-dist.	kNN Mixing	Fréchet	MMD
259	1.0000	1.9883	0.0818	450.6	0.1487
384	0.9998	1.9746	0.0953	430.1	0.1405
518	0.9992	1.9518	0.1067	412.6	0.1315
672	0.9987	1.9375	0.1122	403.9	0.1243
896	0.9979	1.9141	0.1146	408.1	0.1267
1152	0.9985	1.9219	0.1094	423.0	0.1296

of 259 pixels. Increasing the crop size improved source–target overlap up to an intermediate range: the best AUC and proxy A-distance were observed at 896 pixels, while the lowest MMD occurred at 672 pixels. However, all AUC values remained above 0.997, indicating that crop scale alone does not remove the underlying domain gap. In parallel, the morphotype probabilities changed measurably with scale, shifting away from broadleaf forb and toward graminoid and woody structure as more plot context was included. This suggests that multi-scale tiling is not only a way to cover high-resolution images, but also a mechanism that changes the semantic and ecological prior presented to the classifier.

5.6. Interpretation

Taken together, these results show that the transfer from single-plant source imagery to 2026 quadrat-based community imagery constitutes a **severe domain adaptation problem** rather than a minor dataset shift. The observed separation is consistent with simultaneous differences in scale, scene composition, phenological stage, background complexity, and label granularity. In the source domain, images typically depict one plant or one organ prominently, whereas the target task requires reasoning over mixed-species communities in which small, partially occluded, or seasonally altered species may still contribute to the final species inventory.

The crop-based target representation reduced some aspects of the mismatch, especially morphotype prior divergence and certain geometric distribution distances, but it did not close the core gap. Indeed, full quadrat images and mask-guided crops were themselves almost perfectly separable (AUC **0.99996**), indicating that target representation choice is itself a substantial source of distribution shift.

6. Challenge Submission Methodology

The challenge submission pipeline used in this work was based on the Team Atlantic PlantCLEF 2024 system, which combined ViTD2PC24All inference, multi-scale sliding-window region proposals, probabilistic species scoring, and segmentation-based filtering for vegetation-plot analysis [13]. In the present 2026 work, this earlier pipeline was used as the baseline and extended with additional diagnostic and post-processing components, including new mask-guided crop selection, flower-aware crop proposals, morphotype priors, and a domain-critic signal. The baseline should therefore be understood as the 2024 DINOv2 crop-aggregation pipeline adapted to the 2026 dataset, rather than as a single full-image classifier.

The Kaggle submission pipeline combined DINOv2-based species classification, multi-scale crop proposals, plot-level aggregation, morphotype priors, and conservative post-processing. The aim was to improve species recall in vegetation quadrats while limiting the false positives that arise when a classifier trained on single-plant images is applied directly to dense plot-level imagery.

A separate baseline-only public leaderboard score was not retained in the saved submission records. Table 4 therefore reports the saved submission variants, where the term baseline refers to the 2024-derived DINOv2 crop-aggregation pipeline described above.

Table 4

Public and private leaderboard scores for selected PlantCLEF 2026 Kaggle submissions. Scores include both public leaderboard values and final private leaderboard values. Public scores were computed on approximately 11% of the hidden test set. The final private score was 0.34103, placing us in 22nd position in the rankings.

Strategy	Public score	Private score
Baseline + critic top 40	0.34885	0.33612
Baseline + critic top 35	0.34885	0.33501
Baseline + critic top 42	0.34885	0.33598
Baseline + critic top 38	0.34885	0.33611
Baseline + mask critic top 50	0.34750	0.34032
Baseline + mask critic top 40	0.34746	0.34051
Baseline + critic top 60	0.34721	0.33217
Morphotype/flower prior, looser	0.34636	0.34103
Morphotype/flower prior + LISAH, looser	0.34606	0.33858
Morphotype/flower prior + LISAH add 2	0.34575	0.33845
LISAH flower prior only	0.34575	0.33845
Domain critic filled from best baseline	0.24625	0.25052
Domain critic only	0.22412	0.22188

6.1. Multi-scale Region Proposals

The multi-scale proposal stage followed the strategy developed in our PlantCLEF 2024 system [13]. Direct full-image classification was avoided because many species occupy only small regions of a vegetation plot and may be visually diluted by surrounding vegetation, soil, litter, frame artefacts, or other species. Instead, each high-resolution quadrat image was decomposed into overlapping local crops at multiple spatial scales. Each crop was resized to the ViTD2PC24All input resolution and passed through the PlantCLEF species classifier.

For each crop, the classifier returned a ranked list of candidate species and their confidence scores. The crop coordinates, scale, image identifier, predicted species identifiers, and scores were stored for downstream aggregation. The resulting crop-level predictions formed a region-proposal database for each quadrat image. This allowed for the identification of species evidence from local regions instead of relying on a single global image representation.

The baseline aggregation stage converted these crop-level predictions into one plot-level species list. For each image, predictions with the same species identifier were grouped across all crops and scales. The retained score for each species was based on summary evidence across supporting crops, including the maximum crop confidence, the number of supporting crops, and percentile-based filtering of the crop-score distribution. Species with weak or isolated evidence were removed, while species appearing consistently across high-confidence crops were retained. A maximum species-count limit was also applied to avoid excessive over prediction. This baseline was then used as the reference submission to which the later morphotype, flower-aware, and domain-critic variants were compared.

6.2. Mask-guided and Flower-aware Crops

Alongside regular sliding-window crops, mask-guided and flower-aware crops were used to select more informative regions from each quadrat image. The mask was generated using a green vegetation index model, which identified areas with strong green plant material. This helped reduce the effect of soil, stones, litter, snow, frame edges, and other non-target background regions.

Candidate crops were ranked according to vegetation coverage and simple image-quality heuristics, with additional preference given to flower-like regions where present [23]. The purpose was not to create a perfect segmentation mask, but to guide the ViTD2PC24All classifier towards regions that were more likely to contain useful plant features for species prediction.

During visual inspection, we observed that flower-bearing regions could be under-represented by vegetation masks, even though flowers are often highly informative for species identification. For this

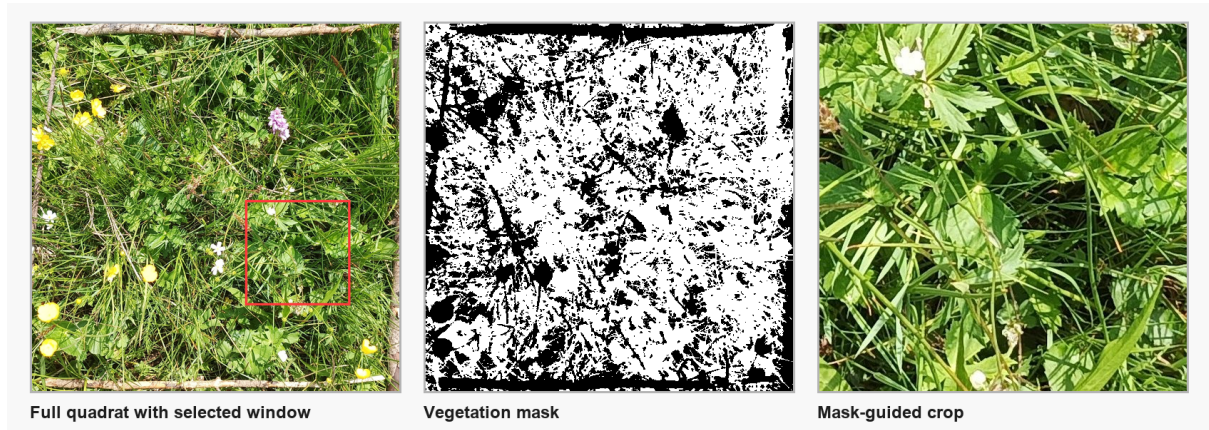


Figure 5: Illustration of the mask-guided crop strategy used for the target-domain representation in the domain-shift audit. A fixed 518×518 pixel window was slid across the quadrat image with a step size of 259 pixels and a border offset of 100 pixels. The final crop was chosen as the window with the highest vegetation-mask coverage, yielding a more plant-centred view than the full quadrat image.



Figure 6: Example output from the morphotype-guided proposal branch. Patch colour indicates the proposal score computed from broad plant-form evidence minus background evidence, with higher-scoring regions shown in red. The method suppresses large non-plant regions such as exposed rock while retaining vegetation-rich areas as candidate crops for the full 518×518 PlantCLEF species classifier.

reason, flower-rich or visually distinctive regions were treated as useful proposal candidates rather than background. Several submitted variants therefore tested flower-aware add-back rules and morphotype-informed blending. These improved some submissions modestly, although the gains were limited by the public leaderboard variance and by the difficulty of detecting small flowers reliably across the full test set.

Inspired by the work of Filho et al. in CLEF2025, we explored a fast proposal branch based on the ViTD2PC24A11+Morphotype model. Rather than using prototype-guided attention directly, each quadrat was scanned at 224×224 resolution and patches were scored using broad plant-form evidence together with a small patch-semantic background head. The proposal score was defined as plant evi-

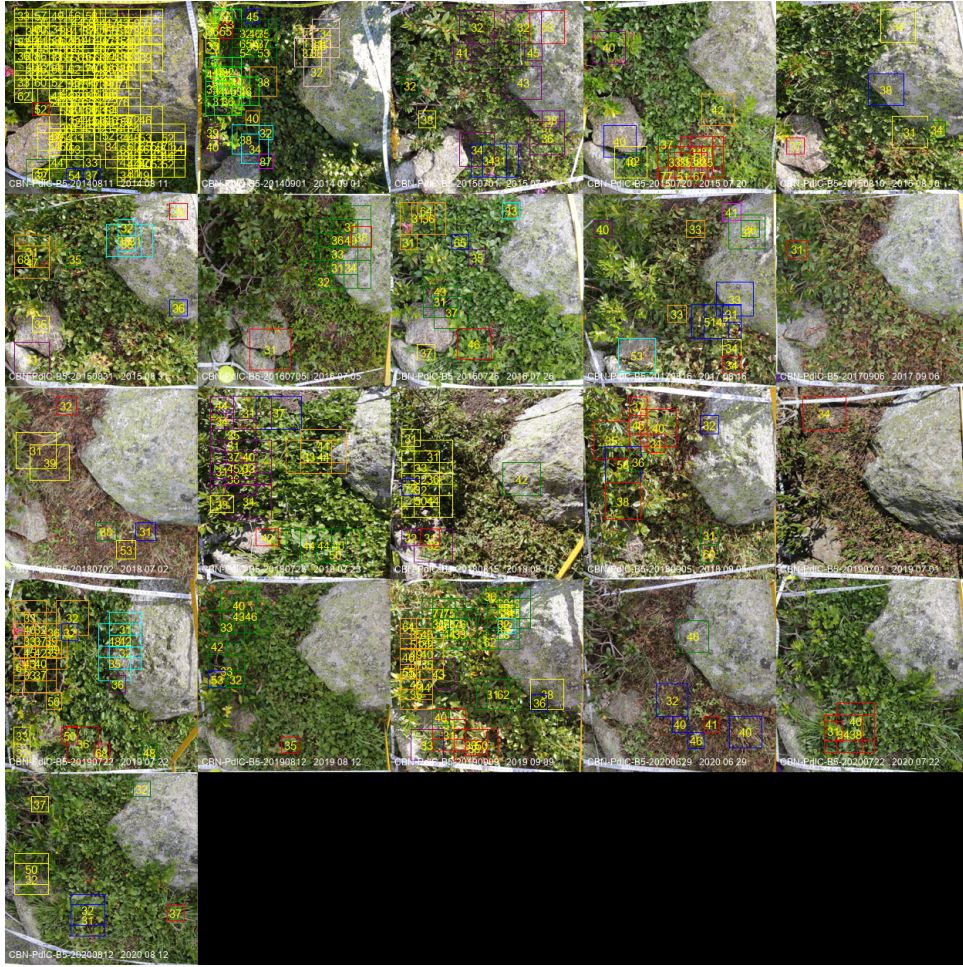


Figure 7: Example debug image showing sliding-window region proposals, coloured by predicted class ID, with class scores shown in yellow.

dence minus background evidence, suppressing regions dominated by rock, soil, snow, frame, or tape artefacts. The highest-scoring patches were then retained as candidate crops for the full 518×518 PlantCLEF species classifier. Figure 9 shows an example of the top five mask-guided crop proposals.

6.3. Plot-level Aggregation

The multi-scale region proposals produced many crop-level predictions for each quadrat image. A direct union of these predictions led to excessive false positives, so a conservative aggregation procedure was used. First, all crop predictions were grouped by image identifier and species identifier. Second, species-level evidence was summarised using the maximum crop confidence, the mean confidence across supporting crops, and the number of crops in which the species appeared. Third, percentile-based filtering was applied, retaining only species with strong relative evidence within a plot. Finally, a maximum number of species was imposed for each submission variant.

Where filename tokens indicated repeated observations of the same physical plot, predictions were also aggregated across temporal plot groups. Species that appeared consistently across repeated observations could be reinforced, while low-confidence species appearing in only one crop or on one date were removed unless their confidence exceeded a conservative threshold. This temporal grouping followed the logic of the 2024 pipeline, where repeated plot observations provided useful evidence for stabilising species predictions [13]. The aggregation therefore combined local crop evidence, confidence-based filtering, and repeated-plot consistency to produce the final plot-level prediction list.

6.4. False-positive Control

A major source of error was high-confidence species prediction on non-vegetative material. Rocks, soil, dead stems, snow, and quadrat artefacts could all trigger plausible but incorrect species predictions. To reduce this, several rejection strategies were tested. Simple colour rules based on green vegetation fraction were useful but brittle, especially under variable lighting and when snow or pale rocks were misclassified as vegetation.

We therefore also explored semantic patch filtering and mask-based rejection. Patches were clustered and manually assigned broad labels such as green vegetation, broadleaf vegetation, dead litter, soil, rock, snow, mixed vegetation, and ignore. These labels were then used to test whether crop proposals dominated by non-plant material should be removed or down-weighted. Figure 8 illustrates an example of using elevated Green channel and grayscale intensity values as a strategy to generate a binary classification mask distinguishing soil and rock from foliage. In practice, this filtering was helpful for diagnosis but did not consistently improve the leaderboard score when applied too aggressively, because some visually mixed crops still contained useful species evidence.

6.5. Morphotype and Domain-critic Signals

The morphotype and domain-critic signals were added as extensions to the 2024 Team Atlantic DINOv2 crop-aggregation baseline [13]. The morphotype branch was not used as a replacement species classifier. Instead, it acted as a lightweight diagnostic and post-processing signal for broad plant-form evidence. A frozen ViTD2PC24All backbone was used to extract pooled feature embeddings from 224×224 crops. These embeddings were passed through a small classification head consisting of layer normalisation, a $768 \rightarrow 256$ linear projection, GELU activation, dropout with $p = 0.20$, and a final $256 \rightarrow 5$ classifier. The five morphotype classes were graminoid, broadleaf forb, woody, fern, and succulent. Figure 11 shows the morphotype-guided proposal branch, where high plant-form evidence is retained and non-plant regions are suppressed.

The domain-critic branch was trained as a binary classifier over frozen ViTD2PC24All embeddings to distinguish source-domain single-plant imagery from target-domain quadrat crops. The critic did not predict species labels directly and did not update the ViTD2PC24All backbone. Its role was to provide a weak compatibility score indicating whether a crop-level representation was closer to the target-domain distribution. During post-processing, candidate species were first generated by the baseline crop-aggregation pipeline. Species supported by high-confidence crops and high target-domain critic compatibility were then retained or added back in controlled top- K variants. Used alone, the domain critic produced weak submissions, but when blended with the 2024-derived baseline it produced the best public leaderboard variants. The best recorded public submissions resulted from blends of the baseline and critic-selected species additions, achieving a public score of approximately 0.34885.

7. Summary and Conclusions

This paper presented Team Atlantic’s PlantCLEF 2026 Kaggle challenge work as both a challenge submission and a domain-shift audit. The submitted system combined ViTD2PC24All DINOv2-based species classification, multi-scale crop proposals, mask-guided target crops, plot-level temporal aggregation, ViTD2PC24All+Morphotype priors, and conservative post-processing. The strongest public leaderboard submissions were obtained by retaining the high-confidence baseline predictions while adding a limited number of species supported by domain-critic evidence. The best recorded public score was 0.34885.

The main finding is that the PlantCLEF vegetation-plot task is dominated by a severe source–target domain shift. To investigate this efficiently, we developed a lightweight diagnostic workflow based on a reduced-input morphotype model. By using 224×224 inputs instead of the larger 518×518 PlantCLEF species-classifier input, the morphotype model provided a faster way to extract representations and visualise domain structure. This made it possible to generate UMAP projections, compare source and

target embeddings, and inspect broad plant-form differences without running the full species classifier for every diagnostic experiment.

Mask-guided cropping, multi-scale tiling, and low-level target normalisation each modified the target representation and, in some cases, reduced specific discrepancy measures. However, none of these operations eliminated the embedding-space domain gap. The morphotype head provided an interpretable diagnostic of semantic prior shift, showing that the source data was dominated by broadleaf-forb imagery while the target quadrats contained much higher predicted graminoid and woody structure. This suggests that the problem involves both visual domain shift and label-prior shift.

The key contribution of this working note is not limited to the PlantCLEF submission pipeline. It also presents a rapid and computationally efficient method for quantifying the source-to-target domain gap between single-plant imagery and vegetation-plot imagery. The results show that this gap remains severe even after mask-guided cropping, resizing, and JPEG normalisation. This indicates that future systems should not rely solely on direct transfer from single-plant classifiers, but should instead combine crop-level adaptation, temporal consistency, and biologically informed priors to improve robustness in quadrat-based species prediction.

8. Acknowledgments

We extend our gratitude to the PlantCLEF organisers for releasing the DINOv2 classifiers and associated challenge material. This made it possible to explore strong vegetation-plot baselines and domain adaptation strategies on limited local compute, while still working with a modern vision foundation model.

9. Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT5 and Writefull for grammar and spelling checks and for drafting assistance during revision. After using these tools, the author(s) reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] H. Goëau, V. Espitalier, P. Bonnet, A. Joly, Overview of PlantCLEF 2024: Multi-Species Plant Identification in Vegetation Plot Images, in: Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum, 2024.
- [2] F. He, K. J. Gaston, Estimating Species Abundance from Occurrence., *The American Naturalist* 156 (2000) 553–559. URL: <https://www.journals.uchicago.edu/doi/10.1086/303403>. doi:10.1086/303403.
- [3] Z. Liao, J. Zhou, T.-J. Shen, Y. Chen, Inferring single- and multi-species distributional aggregation using quadrat sampling, *Ecological Indicators* 156 (2023) 111085. URL: <https://www.sciencedirect.com/science/article/pii/S1470160X2301227X>. doi:10.1016/j.ecolind.2023.111085.
- [4] J. Wäldchen, P. Mäder, Plant Species Identification Using Computer Vision Techniques: A Systematic Literature Review, *Archives of Computational Methods in Engineering* 25 (2018) 507–543. URL: <http://link.springer.com/10.1007/s11831-016-9206-z>. doi:10.1007/s11831-016-9206-z.
- [5] F. Hu, P. Wang, Y. Li, C. Duan, Z. Zhu, Y. Li, X.-S. Wei, A Deep Learning based Solution to FungiCLEF2023, in: M. Aliannejadi, G. Faggioli, N. Ferro, M. Vlachos (Eds.), Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2023), volume 3497 of *CEUR Workshop Proceedings*, CEUR, Thessaloniki, Greece, 2023, pp. 2051–2059. URL: <https://ceur-ws.org/Vol-3497/#paper-173>.
- [6] M. Xu, S. Yoon, C. Wu, J. Baek, D. S. Park, PlantCLEF2023: A Bigger Training Dataset Contributes More than Advanced Pretraining Methods for Plant Identification, in: M. Aliannejadi, G. Faggioli, N. Ferro, M. Vlachos (Eds.), Working Notes of the Conference and Labs of the Evaluation Forum

- (CLEF 2023), volume 3497 of *CEUR Workshop Proceedings*, CEUR, Thessaloniki, Greece, 2023, pp. 2168–2180. URL: <https://ceur-ws.org/Vol-3497/#paper-183>.
- [7] H. Goëau, P. Bonnet, A. Joly, Overview of PlantCLEF 2023: Image-based Plant Identification at Global Scale, in: M. Aliannejadi, G. Faggioli, N. Ferro, M. Vlachos (Eds.), *Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2023)*, volume 3497 of *CEUR Workshop Proceedings*, CEUR, Thessaloniki, Greece, 2023, pp. 1972–1981. URL: <https://ceur-ws.org/Vol-3497/#paper-167>.
 - [8] 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. Hruz, M. Servajean, J. Matas, others, 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.
 - [9] M. Gustineli, A. Miyaguchi, A. Cheung, D. Khattak, Tile-Based ViT Inference with Visual-Cluster Priors for Zero-Shot Multi-Species Plant Identification, in: G. Faggioli, N. Ferro, P. Rosso, D. Spina (Eds.), *CLEF 2025 Working Notes*, volume 4038 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2025, pp. 3029–3038. URL: https://ceur-ws.org/Vol-4038/paper_242.pdf.
 - [10] N. Semenova, Optimizing a Vision Transformer with Ecological Context for Multi-Label Plant Species Identification, in: G. Faggioli, N. Ferro, P. Rosso, D. Spina (Eds.), *CLEF 2025 Working Notes*, volume 4038 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2025, pp. 3166–3178. URL: https://ceur-ws.org/Vol-4038/paper_392.pdf.
 - [11] L. A. D. Filho, A. M. d. S. Neto, R. P. David, R. T. Calumby, Zero-Shot Segmentation through Prototype-Guidance for Multi-Label Plant Species Identification, in: G. Faggioli, N. Ferro, P. Rosso, D. Spina (Eds.), *CLEF 2025 Working Notes*, volume 4038 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2025, pp. 3019–3028. URL: https://ceur-ws.org/Vol-4038/paper_241.pdf.
 - [12] V. Espitalier, Preprocessing Is All You Need: TheHeartOfNoise Submission to PlantCLEF 2025, in: G. Faggioli, N. Ferro, P. Rosso, D. Spina (Eds.), *CLEF 2025 Working Notes*, volume 4038 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2025, pp. 2979–3000. URL: https://ceur-ws.org/Vol-4038/paper_238.pdf.
 - [13] S. Foy, S. McLoughlin, Utilising DINOv2 for Domain Adaptation in Vegetation Plot Analysis, in: *Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024)*, volume 3740 of *CEUR Workshop Proceedings*, CEUR-WS.org, Grenoble, France, 2024, pp. 2083–2092. URL: <https://ceur-ws.org/Vol-3740/paper-196.pdf>.
 - [14] S. Foy, S. McLoughlin, A self-supervised approach for the detection and classification of turfgrass abnormalities, *IET Conference Proceedings 2024 (2024)* 234–241. doi:10.1049/icp.2024.3310.
 - [15] G. Martellucci, I. Moummad, H. Goëau, P. Bonnet, F. Vinatier, A. Joly, Overview of PlantCLEF 2026: Identify Multi-species Plants in Images of Vegetation Plots, in: *Working Notes of CLEF 2026 - Conference and Labs of the Evaluation Forum*, 2026.
 - [16] L. Picek, L. Adam, S. Kahl, R. Bossy, L. Chrobak, H. Goëau, K. Papafitsoros, H. Klinck, W.-P. Veltinga, R. Planqué, T. Denton, K. Barnard, C. Nédellec, L. Deléger, M. Courtin, G. Martellucci, I. Moummad, F. Vinatier, P. Bonnet, A. Joly, Overview of LifeCLEF 2026: AI Challenges for Biodiversity Understanding and Ecosystem Management, in: *International Conference of the Cross-Language Evaluation Forum for European Languages*, Springer, 2026.
 - [17] G. Martellucci, H. Goëau, P. Bonnet, F. Vinatier, A. Joly, Overview of PlantCLEF 2025: Multi-Species Plant Identification in Vegetation Quadrat Images, in: G. Faggioli, N. Ferro, P. Rosso, D. Spina (Eds.), *CLEF 2025 Working Notes*, volume 4038 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2025, pp. 2942–2954. URL: https://ceur-ws.org/Vol-4038/paper_235.pdf.
 - [18] H. Goëau, J.-C. Lombardo, A. Affouard, V. Espitalier, P. Bonnet, A. Joly, PlantCLEF 2024 pretrained models on the flora of the south western Europe based on a subset of Pl@ntNet collaborative images and a ViT base patch 14 dinoV2 (2024). URL: <https://zenodo.org/records/10848263>.
 - [19] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, M. Assran, N. Ballas, W. Galuba, R. Howes, P.-Y. Huang, S.-W. Li, I. Misra, M. Rabbat, V. Sharma, G. Synnaeve, H. Xu, H. Jegou, J. Mairal, P. Labatut, A. Joulin, P. Bojanowski, DINOv2: Learning Robust Visual Features without Supervision, 2024. URL: <http://arxiv.org/abs/>

2304.07193. doi:10.48550/arXiv.2304.07193, arXiv:2304.07193 [cs].

- [20] PlantNet, PlantNet <https://plantnet.org/en/>, 2019. URL: <https://plantnet.org/en/>.
- [21] M. Sorkhei, C. Matsoukas, J. F. Haslum, E. Konuk, K. Smith, k-NN as a Simple and Effective Estimator of Transferability, Transactions on Machine Learning Research (2025). URL: <https://openreview.net/forum?id=hGlkjP1zHc>.
- [22] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, Y. LeCun, OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks, 2014. URL: <http://arxiv.org/abs/1312.6229>. doi:10.48550/arXiv.1312.6229, arXiv:1312.6229 [cs].
- [23] A. Patrignani, T. E. Ochsner, Canopeo: A Powerful New Tool for Measuring Fractional Green Canopy Cover, Agronomy Journal 107 (2015) 2312–2320. URL: <https://onlinelibrary.wiley.com/doi/abs/10.2134/agronj15.0150>. doi:10.2134/agronj15.0150, _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.2134/agronj15.0150>.

A. Appendix: Additional Debug Data

Three sliding-window configurations were tested during early inference development. Set 1 used a raw 518×518 window with a stride of 172 pixels and a 150-pixel border offset; inference for this configuration took approximately 12 hours on a single RTX 4080. Set 2 used a smaller raw window of 259×259 pixels, corresponding to half the DINOv2 input size, and upsampled each crop to 518×518 before inference; this used the same 172-pixel stride and 150-pixel border offset and took approximately 18 hours. Set 3 used a raw 224×224 window with a stride of 112 pixels, upsampled crops to 518×518 , and evaluated $+90^\circ$ and -90° rotations without border padding; this configuration required more than 24 hours.

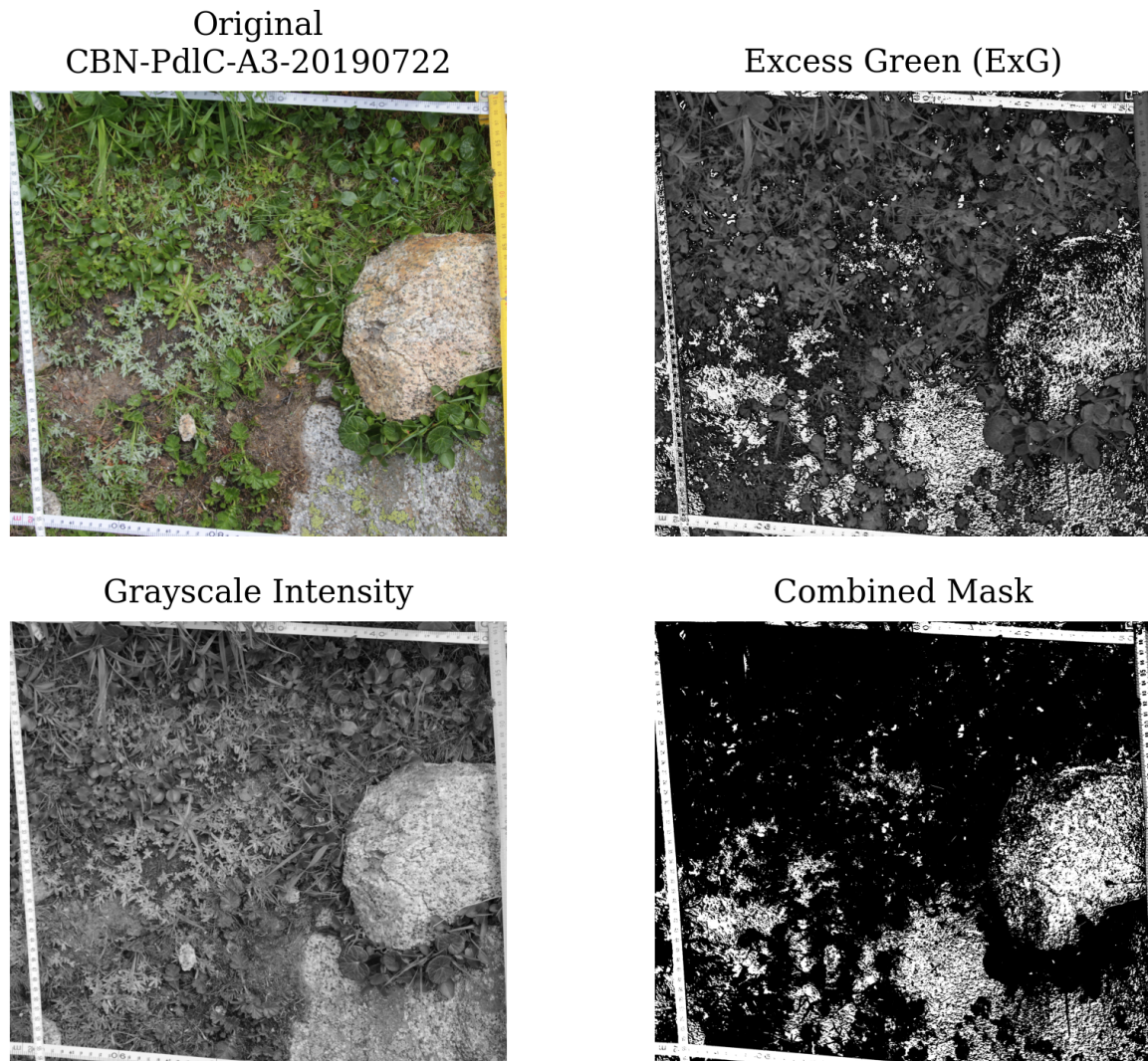


Figure 8: Excess Green and Grayscale intensity strategy to generate a soil and rock binary mask

Top-5 diverse mask-guided crops from one quadrat image

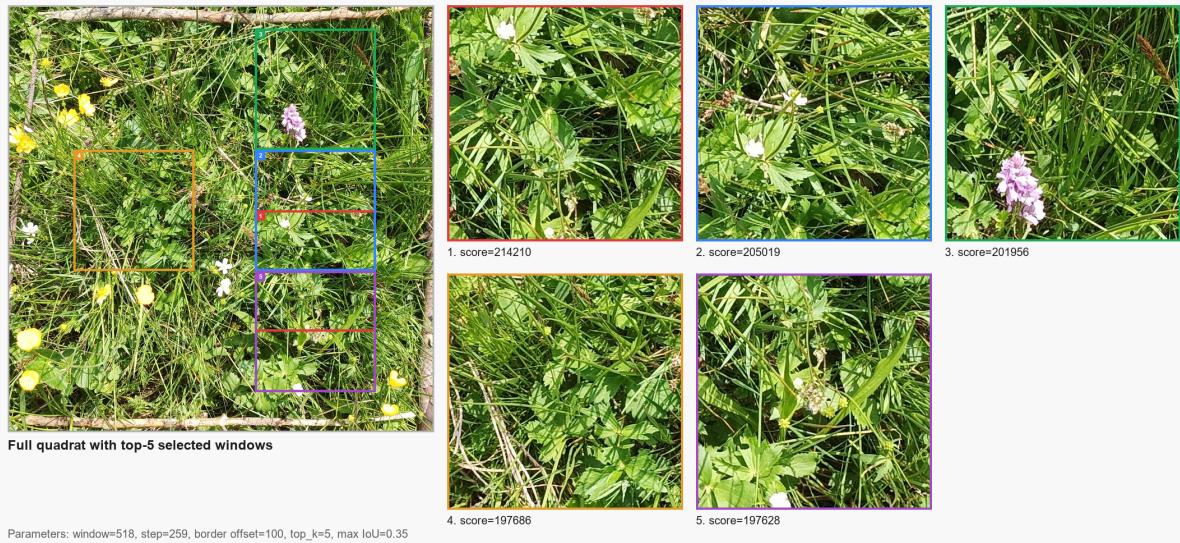


Figure 9: Example region proposals produced by the mask-guided crop strategy. The selected windows concentrate the fixed-size classifier input on vegetation-rich parts of the quadrat rather than on the full image frame.

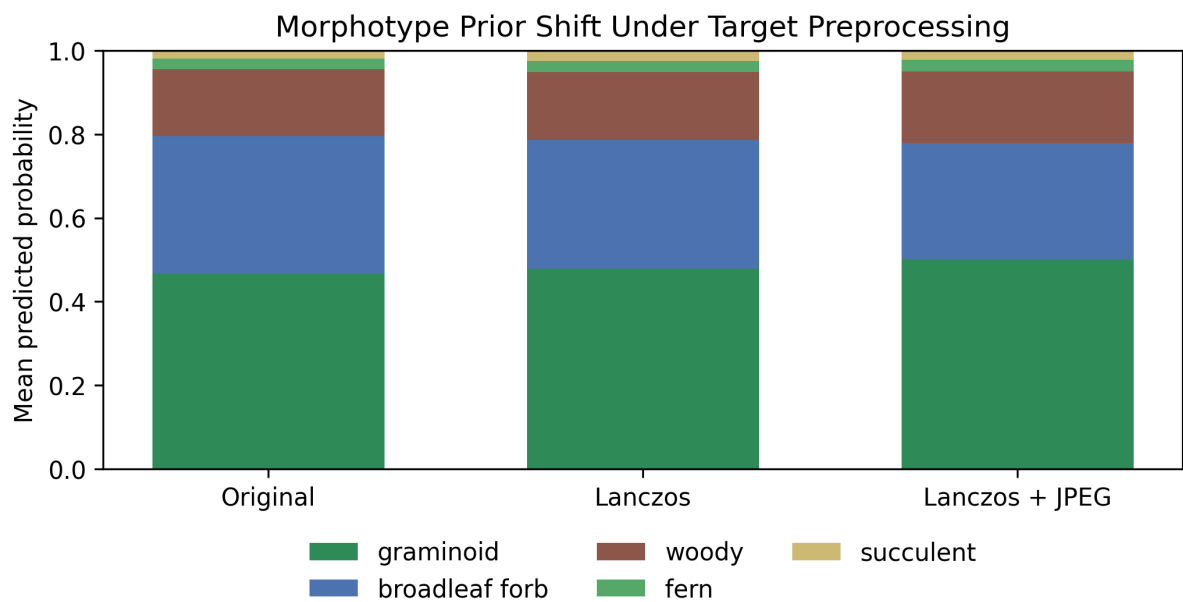


Figure 10: Morphotype prior shift under target preprocessing. Lanczos resizing and JPEG re-compression increased the mean predicted graminoid and woody probabilities while reducing broadleaf-forb probability on a deterministic 500-image target-crop sample.

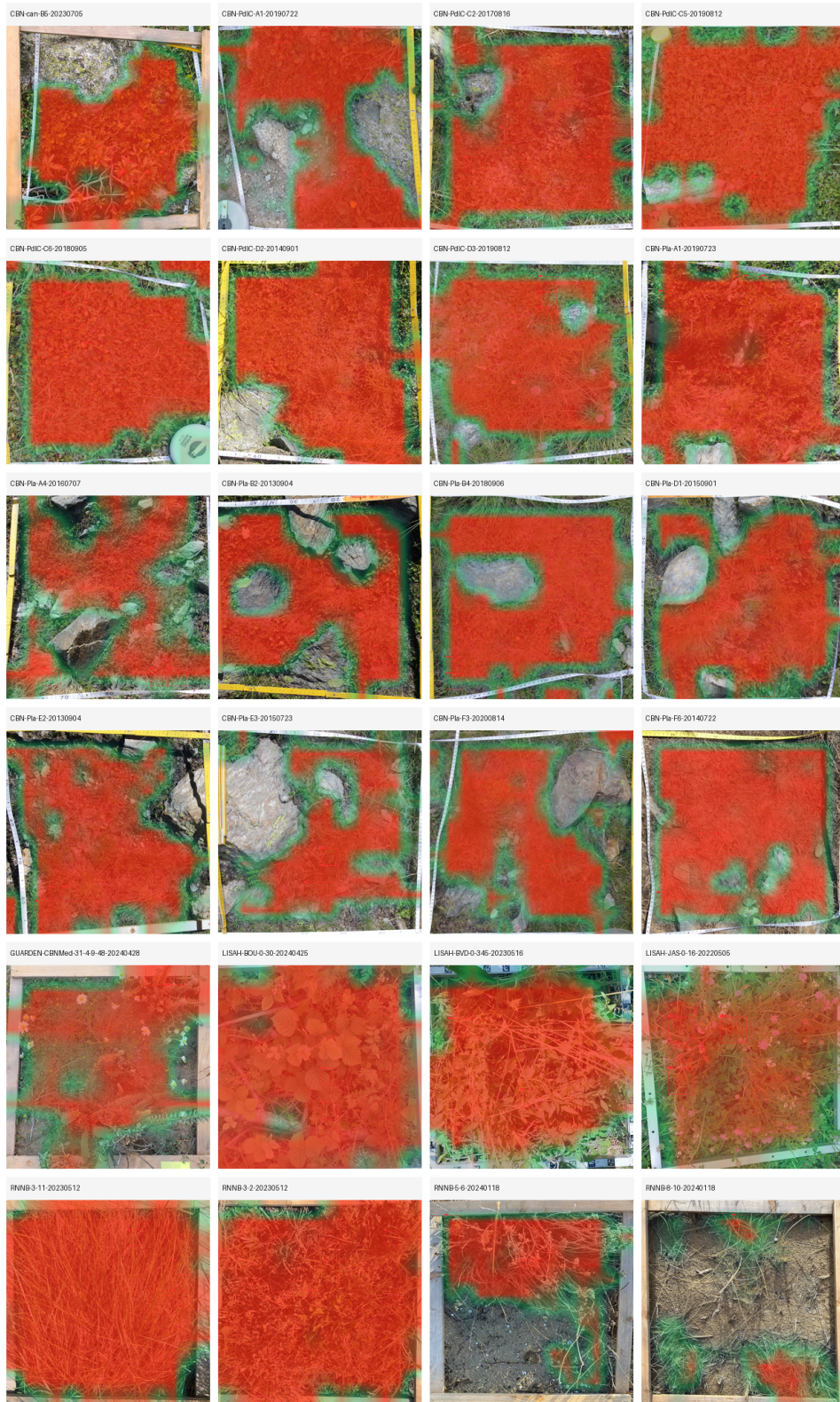


Figure 11: Examples output from the morphotype-guided proposal branch. Patch colour indicates the proposal score computed from broad plant-form evidence minus background evidence, with higher-scoring regions shown in red. The method suppresses large non-plant regions such as exposed rock while retaining vegetation-rich areas as candidate crops for the full 518×518 PlantCLEF species classifier.