

Contextual Multi-Label Identification of Plant Species in Multi-Species Plots^{*}

Notebook for the LifeCLEF Lab at CLEF 2026

Andrea Menco-Tovar^{1,*†}, Jairo E. Serrano^{1,†}, Juan Carlos Martinez-Santos^{1,†} and Edwin Puertas^{1,†}

¹Universidad Tecnologica de Bolivar, Ternera km 1, Cartagena, Bolivar, Colombia

Abstract

This work presents the official submission of the team ANDREAMENCOTOVAR, hereafter referred to as VerbaNexAI Lab, through a hybrid pipeline for multi-label plant species prediction at the quadrat level in PlantCLEF 2026. The proposal combines multiview visual inference with DINOv2, global prototype-based retrieval, and site-date ecological consensus in order to integrate local visual evidence, global similarity, and community context. Unlike direct classification of the full image, the system decomposes each plot into complementary views and consolidates predictions through adaptive species selection and conservative post-processing. In the official evaluation, the method achieved an F1 score of 0.33106 on the public set and 0.32980 on the private set, ranking 28th and 27th, respectively, among 103 participants. The results show that combining visual evidence and ecological structure improves the coherence of species lists in complex multi-species scenarios.

Keywords

PlantCLEF 2026, Multi-species identification, DINOv2, Ecological consensus, Multiview inference

1. Introduction

Ecological monitoring is an essential tool for understanding ecosystem dynamics and guiding conservation strategies. Among the most widely used methods are vegetation inventories, which enable biodiversity assessment through standardized sampling and long-term monitoring. These inventories have provided key information for environmental management and evidence-based decision-making. However, traditional methods based on visual species identification present significant limitations, as they require substantial time investment and the involvement of highly trained specialists, which restricts the frequency and coverage of ecological studies [1, 2, 3].

The incorporation of artificial intelligence (AI) techniques offers a promising alternative to overcome these restrictions by facilitating automatic species identification and expanding study coverage [4, 5, 6]. In recent years, platforms such as Pl@ntNet and iNaturalist have demonstrated the potential of citizen participation in generating and annotating botanical observations, enabling the identification of individual species from isolated photographs and fostering the large-scale collection of data by experts and non-experts [7, 8]. Nevertheless, the simultaneous identification of multiple species in high-resolution plot images remains a considerable technological challenge. Deep learning models require large annotated datasets, but there is a marked disparity between training images—generally of individual plants with single labels—and ecological plot images, which reflect complex floristic contexts such as those of the Pyrenean and Mediterranean flora. This difficulty calls for the development of robust models capable of performing multilabel classification in diverse ecological scenarios [9, 10, 11].

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

^{*}Corresponding author.

[†]These authors contributed equally.

✉ amenco@utb.edu.co (A. Menco-Tovar); jserrano@utb.edu.co (J. E. Serrano); jcmartinezs@utb.edu.co (J. C. Martinez-Santos); epuerta@utb.edu.co (E. Puertas)

ORCID 0000-0002-6861-7547 (A. Menco-Tovar); 0000-0001-8165-7343 (J. E. Serrano); 0000-0003-2755-0718 (J. C. Martinez-Santos); 0000-0002-0758-1851 (E. Puertas)



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

Within this framework, the PlantCLEF 2026 challenge focuses on multilabel prediction of plant species in high-resolution quadrat images [12, 13]. The proposed task seeks to advance the ability of computer vision models to recognize multiple species in complex ecological contexts, contributing to the development of more efficient and scalable monitoring systems. The evaluation is based on metrics such as sample-wise macro F1, which balances precision and recall, avoiding both overprediction and underprediction. This approach ensures a fair comparison between methods and promotes solutions applicable to real-world ecological monitoring scenarios.

This manuscript describes our participation in the PlantCLEF 2026 challenge, detailing the methodology employed and the results obtained. Our approach is based on the use of deep learning models and inference strategies adapted to the complexity of plot images, with the aim of contributing to the advancement of ecological monitoring systems that are more efficient, scalable, and capable of supporting biodiversity conservation across different floristic contexts.

2. Related Work

Several studies have demonstrated the usefulness of vegetation plot inventories for assessing biodiversity and guiding conservation strategies [1, 7, 8]. However, the literature consistently points out that these methods have important limitations related to the time required and the need for specialized experts [6]. The application of artificial intelligence techniques has been explored in multiple works as an alternative to overcome these restrictions. Platforms such as Pl@ntNet and iNaturalist have been widely documented in the literature, showing how citizen science can contribute to generating large volumes of data and to identifying individual species [4, 11]. Nonetheless, most of these approaches focus on isolated images of plants, which limits their applicability in complex ecological contexts.

The PlantCLEF 2024 and 2025 overviews show a consistent pattern in which competitive approaches tend to combine strong visual extractors—often pretrained Transformers—region-based inference mechanisms to capture local signals in large images, and ensembles to increase the coverage of the predicted species set. In particular, tiling-based solutions—splitting the image into tiles and aggregating predictions—appear as a direct response to the scene-dense nature of the test set, where small species may occupy a minimal fraction of the full image. This idea is reflected both in general descriptions of the challenge and in technical reports from participants [1, 2, 3].

In this context, adopting self-supervised backbones is a natural choice because they provide more transferable visual representations under distribution shifts. A widely used architecture is DINOv2, designed to produce robust “all-purpose” features and, therefore, used as a foundation for adapting to new tasks with less dependence on exhaustive labeling. In PlantCLEF, there are proposals that use DINOv2 to extract embeddings and then perform multilabel classification with tile-based inference and aggregation, reporting that this combination helps scale processing and recover species that do not visually dominate the frame. Complementarily, multi-head architectures have also been explored to incorporate taxonomic hierarchy and auxiliary signals to stabilize prediction when the visual evidence per species is weak or partial, which is frequent in quadrats with occlusion and co-occurrence [5, 14].

Recent evidence suggests that multi-species identification in vegetation plots requires integrating global and local signals to mitigate domain shift and the strong scale variation inherent to dense scenes. In this setting, tile-based aggregation and the use of self-supervised representations emerge as key components for expanding species coverage, while ensembles make it possible to exploit complementarity across models and reduce omissions in the predicted set. Likewise, multi-view schemes are justified as pragmatic strategies to increase recall by capturing evidence that may be missed depending on framing or the spatial distribution of species. This line of work motivates the methodological approach of the present study, aimed at evaluating in a controlled manner how the combination of robust representations, local inference, and prediction fusion contributes to improving multilabel identification in the PlantCLEF context.

3. Methodology

This section details the training set used, which consists of individual observations organized by species, as well as the test set comprising multi-species vegetation plot images captured in various floristic contexts. Regarding the architecture, described the process of image partition, data preparation for analysis, and the mechanism used to obtain and filter the final predictions.

3.1. Dataset Description

PlantCLEF 2026 is formulated as a multi-label prediction task at the *quadrat* level, in which each plot image may contain multiple plant species. This setting requires moving from a training set mainly based on individual plant observations to an evaluation scenario composed of complex plant communities. Consequently, the dataset requires careful organization of images, metadata, plot identifiers, and the taxonomic catalogue in order to ensure coherent and valid predictions for each *quadrat_id*.

The dataset used integrates PlantCLEF training and validation images, taxonomic metadata, and a catalogue of 7806 plant species. In the developed implementation, the data were organized through unified *manifests*, path validation, species identifier normalization, and taxonomic consistency checking. This stage made it possible to build a coherent representation of the images, their metadata, and the *quadrat_id* identifiers required to generate valid plot-level predictions.

3.1.1. Training Set

The training set consists of images of individual plants from the PlantCLEF/Pl@ntNet corpus, focused on the flora of southwestern Europe, as shown in Figure 1. According to the official organization of the challenge, this set contains approximately 1.4 million images and covers around 7.8 thousand plant species, complemented with reliably labelled images aggregated from GBIF for less represented species.

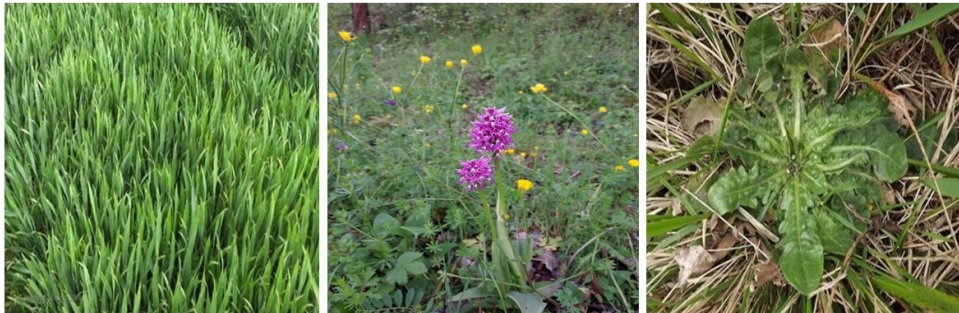


Figure 1: Examples of images of plant plots from the training set containing multiple species in different floristic contexts and capture conditions.

In this work, the training set was used as a direct source for classification, and as the basis for constructing a visual reference bank. From the training and validation images, deep representations were extracted to generate visual prototypes for each species. This methodological decision made it possible to transform the training corpus into a global visual memory capable of supporting the retrieval of visually plausible species during the inference stage.

In addition, the taxonomic metadata were harmonized through a species table that includes *species_id*, scientific name, genus, and family. This structure made it possible to maintain the correspondence between the model outputs, the visual prototypes, and the official submission format. Consequently, the training set was leveraged both as a repository of labelled examples and as a structured source for modelling visual similarity and taxonomic coherence.

3.1.2. Test Set

The test set corresponds to images of vegetation plots produced in different floristic contexts, including Pyrenean and Mediterranean floras, as shown in Figure 2. These images represent real plant communities, where several species may coexist within the same plot. Therefore, the prediction unit is not the individual image or an isolated region, but the plot identifier or `quadrat_id`.



Figure 2: Examples of images of plant plots from the test set containing multiple species in different floristic contexts and capture conditions.

In the developed implementation, the clean *manifest* of the test set included 2,105 valid images, each associated with a unique `quadrat_id`. The absence of duplicates, missing paths, and identifier inconsistencies was verified. This validation was essential to ensure that each row in the output file corresponded exactly to an evaluable plot.

Given that the test images contain multiple species, different spatial scales, and possible regions with partial information, each plot was processed using a multi-view strategy. For each image, full-image views, central crops, and overlapping tiles were generated. This decomposition made it possible to capture both the global structure of the plant community and local signals associated with small, partially occluded species or species visually dominated by the surrounding environment.

3.1.3. Architecture

The proposed architecture is conceived as a hybrid pipeline for visual inference, prototype-based retrieval, and ecological consensus, as shown in Figure 3. Instead of relying exclusively on direct classification of the full image, the method integrates multiscale visual evidence, similarity to visual prototypes, and ecological context derived from plot identifiers. This integration explicitly addresses the domain shift between individual training images and multi-species evaluation plots, where several taxa may appear at different spatial scales within the same quadrat.

Multiview preprocessing. Each test image was decomposed into complementary views in order to preserve both global community-level context and local visual evidence. The generated views included the resized full image, a central crop, and overlapping local tiles. All views were resized to 518×518 pixels, matching the input resolution of the visual backbone. Local tiles were extracted using a tile size of 518×518 pixels and a stride of 259 pixels, corresponding to an approximate 50% overlap. This configuration allowed the system to capture small or partially occluded species that could be underrepresented in full-image predictions.

In addition, a lightweight vegetation-ratio estimate was computed for each view using a green-channel dominance heuristic. A pixel was considered vegetation-like when:

$$G > 1.05R, \quad G > 1.05B, \quad G > 35.$$

The vegetation ratio was defined as the proportion of pixels satisfying this condition. This estimate was used as an independent segmentation module and as an auxiliary weighting factor during score aggregation and species selection.

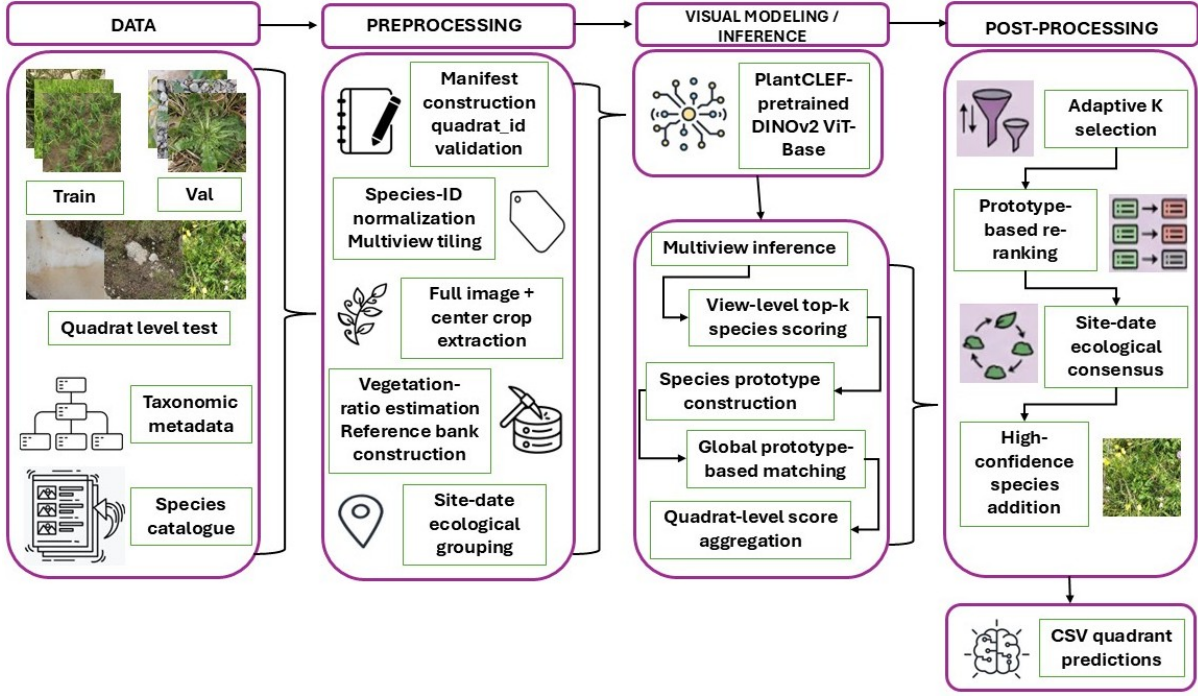


Figure 3: General diagram of the system workflow, showing the steps from image upload to generating and filtering predictions by species.

Visual backbone and view-level inference. The visual core of the system was a DINOv2 ViT-Base model pretrained for PlantCLEF, with an output space of 7806 plant species. For each generated view, the model produced a ranked list of candidate species. The top 40 predictions per view were retained as local visual evidence. These view-level predictions were then grouped by `quadrat_id` and `species_id` to construct a consolidated candidate pool for each plot.

This design avoided depending on a single full-image prediction and allowed localized species evidence to contribute to the final plot-level prediction. Therefore, species that appeared only in specific tiles, central regions, or less dominant parts of the image could still be considered during quadrat-level aggregation.

Quadrat-level aggregation and adaptive species selection. View-level predictions were aggregated at the quadrat level by combining confidence, recurrence across views, maximum evidence, and vegetation-weighted support. For each candidate species s , the visual aggregation score was computed as:

$$S_s = (0.50 \cdot \bar{p}_s^{top} + 0.25 \cdot p_s^{max} + 0.15 \cdot p_s^{veg} + 0.10 \cdot f_s + b_s) \cdot \lambda_s,$$

where $\bar{p}_s^{top}$ represents the mean of the highest view-level scores for species s , p_s^{max} is the maximum score observed across views, p_s^{veg} is the vegetation-weighted score, f_s is a frequency-based support term, and b_s is a contextual bonus assigned when the species was supported by global or central views. The penalty factor λ_s was used to reduce the effect of weak isolated detections; specifically, species appearing in only one view with $p_s^{max} < 0.30$ were penalized.

The initial number of predicted species was determined using an adaptive K strategy. This strategy considered the global vegetation ratio of the quadrat, the relative score with respect to the top-ranked species, the minimum absolute score, the number of supporting views, and the frequency of occurrence across views. The adaptive selection was designed to avoid both underprediction in dense quadrats and overprediction in visually simpler plots.

Prototype-based visual retrieval. To complement the classifier-based ranking, a prototype-based retrieval mechanism was incorporated. DINOv2 embeddings were extracted from training and validation reference images. The embeddings were L2-normalized before prototype construction. For each species, a visual prototype was computed as the mean of its normalized embeddings, followed by a second L2 normalization. Since both test embeddings and species prototypes were normalized, similarity was computed using the dot product, equivalent to cosine similarity.

Two retrieval levels were used. First, a local prototype-based re-ranking stage was applied to species already present in the visual candidate pool. For this stage, up to 24 reference images per candidate species were used, balancing training and validation examples when available. Second, a global prototype recovery stage was constructed over the full catalogue of 7806 species, using up to six reference images per species. This global stage expanded the search space and allowed the system to recover visually plausible species that were not sufficiently ranked by the direct classifier.

For global retrieval, the top 80 prototype matches per view were retained and aggregated at the quadrat level. The global retrieval score for species s was computed as:

$$G_s = 0.40 \cdot \text{sim}_s^{\text{top5}} + 0.25 \cdot \text{sim}_s^{\text{max}} + 0.20 \cdot \text{sim}_s^{\text{veg}} + 0.10 \cdot \text{sim}_s^{\text{mean}} + 0.05 \cdot r_s,$$

where $\text{sim}_s^{\text{top5}}$ is the mean similarity among the best view-level matches, $\text{sim}_s^{\text{max}}$ is the maximum similarity, $\text{sim}_s^{\text{veg}}$ is the vegetation-weighted similarity, $\text{sim}_s^{\text{mean}}$ is the average similarity across supporting views, and r_s is a rank-based term. The top 300 global candidates per quadrat were retained for subsequent re-ranking and contextual refinement.

Site-date ecological consensus. The final post-processing layer incorporated ecological consensus by exploiting the structure of the `quadrat_id` identifiers. When possible, quadrats sharing site-date information were grouped together, under the assumption that ecologically related plots may share part of their species composition. Within each group, recurrent species were reinforced only when they had sufficient visual or prototype-based support.

For each ecological group, a consensus score was computed by combining evidence from previous prediction stages, reciprocal-rank information, visual ranking, and prototype-based retrieval. The ecological consensus score for species s was defined as:

$$E_s = 0.28 \cdot q_s^{\text{pred}} + 0.20 \cdot w_s^{\text{src}} + 0.14 \cdot rr_s + 0.12 \cdot q_s^{\text{base}} + 0.10 \cdot b_s^{\text{score}} + 0.08 \cdot q_s^{\text{global}} + 0.06 \cdot g_s^{\text{score}} + 0.02 \cdot r_s^{\text{global}},$$

where q_s^{pred} represents the frequency of the species in previous predictions within the ecological group, w_s^{src} summarizes weighted support from different prediction sources, rr_s is a reciprocal-rank term, q_s^{base} is the frequency in the base visual candidates, b_s^{score} is the normalized classifier score, q_s^{global} is the frequency in global prototype candidates, g_s^{score} is the normalized global retrieval score, and r_s^{global} is the global rank score.

The scalar weights used in the aggregation scores (S_s), (G_s), and (E_s), as well as the thresholds used in the green-channel vegetation heuristic, were empirically defined through preliminary development runs and qualitative inspection of the prediction behavior. They were not obtained through an exhaustive hyperparameter optimization procedure. The weights were selected to prioritize the strongest visual evidence while keeping prototype-based retrieval and ecological consensus as complementary and conservative sources of support. Similarly, the RGB rule was used as a lightweight green-dominance heuristic to estimate vegetation-like pixels, rather than as a trained segmentation model. Once selected, these values were kept fixed for the final submission.

The consensus mechanism was applied conservatively. Candidate species could be used for replacement or addition only when they satisfied ecological rank, visual support, and multi-source evidence constraints. This avoided uncontrolled propagation of false positives across related plots while allowing recurrent high-confidence species to improve the coherence of the final species lists.

Final prediction validation. The final species list for each quadrat was generated through adaptive K selection, prototype-based re-ranking, controlled ecological replacement, and conservative addition of high-confidence species. Before submission, all predictions were validated against the official PlantCLEF format. The validation process checked that each output row corresponded to a valid `quadrat_id`, that no quadrat was missing or duplicated, and that all predicted `species_id` values belonged to the official taxonomic catalogue. Duplicate species within the same quadrat were also removed.

Thus, the proposed architecture combines deep visual inference, non-parametric visual retrieval, and lightweight contextual modelling to produce coherent multi-label predictions at the plot level while maintaining control over species-list validity and submission consistency.

4. Experimental Results

The official metric used by the organizers was the sample-averaged macro F1 score, which is suitable for multi-label classification problems because it simultaneously considers precision and recall at the level of each *quadrat*. This metric penalizes both the omission of present species and the excessive inclusion of non-relevant species, favoring systems capable of balancing coverage and specificity in the generated lists. The *leaderboard* was reported using public and private partitions, and the final ranking was determined based on the private partition. Table 1 presents the official public and private results of our solution, together with the position obtained among the 103 participating teams. For comparative purposes, the score achieved by the best-performing team is also included.

Table 1
Official PlantCLEF 2026 public and private leaderboard results.

System	Public F1	Public Rank	Private F1	Private Rank
Best-performing team	0.77083	1	0.75737	1
VerbaNexAI Lab	0.33106	28	0.32980	27

The final solution achieved a score of 0.33106 on the *public leaderboard* and 0.32980 on the *private leaderboard*, ranking 28th and 27th, respectively, among 103 participants. These results place the proposal within the top third of the participating teams, indicating a reasonably competitive capacity in a highly complex scenario characterized by multi-species plot images, ecological variability, and a broad taxonomic prediction space. Although the gap with the leading teams was considerable, the result obtained is methodologically relevant, since it was not based solely on direct classification of the full image. Instead, the system integrated multiple sources of evidence: multiview visual predictions, global prototype-based retrieval, and ecological consensus derived from the site-date structure of the plot identifiers. This combination made it possible to explicitly address the domain shift between individual training images and community-level evaluation scenes.

The experimental evolution showed that the largest gains did not come from isolated modifications to the visual model, but from the progressive incorporation of aggregation and contextual mechanisms. Multiview inference with adaptive aggregation provided an initially competitive baseline. Subsequently, prototype-based re-ranking made it possible to retrieve visually plausible species within the complete taxonomic catalogue. However, the most significant improvement was obtained by incorporating site-date ecological consensus and a conservative strategy for adding high-confidence species. This behavior suggests that, in plot-level prediction tasks, contextual and community-level information can effectively complement local visual evidence.

Table 2 summarizes the incremental contribution of the main components of the pipeline. The results indicate that strategies based exclusively on visual evidence produced moderate improvements, whereas the incorporation of ecological structure led to a more marked increase. In particular, site-date consensus contributed to correcting species lists based on co-occurrence patterns among related plots, and the conservative addition of species increased *recall* without aggressively altering already consolidated predictions.

Table 2

Progressive contribution of the main methodological stages.

Methodological stage	Observed leaderboard score
Multiview inference with adaptive species selection	0.31540
Local prototype-based re-ranking	0.31649
Global prototype-based recovery	0.31660
Ecological consensus with capped replacement	0.32234
Ecological high-confidence species addition	0.33087

The main contribution of the approach lies in shifting prediction from a purely visual scheme toward context-aware ecological inference. The solution does not assume that each test image is independent; rather, it exploits implicit relationships among plots with similar spatial and temporal conditions. This decision was especially important in PlantCLEF 2026, where the evaluation unit corresponds to plant communities rather than isolated botanical observations.

In terms of robustness, the final solution maintained explicit control over species-list length, the validity of taxonomic identifiers, and the format required by the official evaluation. In addition, more aggressive variants were discarded when they increased the risk of introducing false positives, prioritizing a final configuration based on controlled ecological replacements and high-confidence additions. This strategy made it possible to balance precision and coverage in a highly imbalanced, large-scale taxonomic multi-label scenario. Although the final score remains below that of the leading teams, the performance obtained validates the relevance of combining multiscale visual inference, prototype-based retrieval, and ecological consensus as a robust alternative for plot-level plant species prediction.

5. Conclusions

This work presented a hybrid pipeline for PlantCLEF 2026 aimed at multi-label plant species prediction at the *quadrat* level. The proposal integrated multiview visual inference with DINOv2, global prototype-based retrieval, and site-date ecological consensus, enabling a shift from isolated visual classification toward predictions that are more coherent with the community structure of the plots.

The official results show that the system achieved an F1 score of 0.33106 on the public set and 0.32980 on the private set, ranking 28th and 27th, respectively, among 103 participants. Although the performance remained below that of the leading team, the experimental evolution showed that the most relevant improvements came from the proposed aggregation, prototype-based retrieval, and ecological post-processing modules, rather than from direct adjustments to the visual backbone.

The main methodological contribution lies in combining local and global visual evidence with contextual information derived from plot identifiers. In particular, ecological consensus and the conservative addition of high-confidence species proved to be effective mechanisms for improving coverage without indiscriminately increasing the number of predictions. These findings suggest that community-level plant recognition tasks require strategies that incorporate not only visual similarity, but also ecological relationships among samples.

As future work, more robust mechanisms for ecological modelling, hierarchical taxonomic integration, and multi-label list calibration will be explored in order to reduce false positives and improve generalization in scenarios of high plant diversity.

Code availability statement

The source code used for the implementation, training, and evaluation of the proposed models is available in the GitHub repository¹.

¹<https://github.com/VerbaNexAI/CLEF2026>

Generative AI Declaration

During the preparation of this work, ChatGPT was used to review translation, grammar, and spelling. After using this tool, the content, coherence, and cohesion were reviewed and edited as necessary, and full responsibility for the content of the publication was ultimately assumed.

Acknowledgments

The authors express their gratitude to the Call 933 "Training in National Doctorates with a Territorial, Ethnic and Gender Focus in the Framework of the Mission Policy — 2023" of the Ministry of Science, Technology and Innovation (Minciencia). In addition, we thank the team of the Artificial Intelligence Laboratory VerbaNex ², affiliated with the UTB, for their contributions to this project.

References

- [1] G. Martellucci, H. Goëau, P. Bonnet, F. Vinatier, A. Joly, Overview of PlantCLEF 2025: Multi-species plant identification in vegetation quadrat images, arXiv preprint, 2025. URL: <https://arxiv.org/pdf/2509.17602>.
- [2] A. Menco-Tovar, J. Serrano, J. Martínez-Santos, E. Puertas, Tiles-wise inference with vision transformers for multispecies identification in vegetation images, in: Conference and Labs of the Evaluation Forum, 2025. doi:10.1007/978-3-032-08206-0_28.
- [3] PlantCLEF, PlantCLEF 2025 | imageclef / lifeclef - multimedia retrieval in clef, Website, 2025. URL: <https://www.imageclef.org/PlantCLEF2025>, accessed: 2026-02-27.
- [4] A. S. Adebayo, AI driven species recognition and digital systematics: Applying artificial intelligence for automated organism classification in ecological and environmental monitoring, International Journal of Research Publication and Reviews 6 (2025) 31–49. doi:10.55248/gengpi.6.0225.0703.
- [5] M. Gustineli, A. Miyaguchi, I. Stalter, Multi-label plant species classification with self-supervised vision transformers, in: CEUR Workshop Proceedings, volume 3740, 2024, pp. 2093–2102. URL: <http://arxiv.org/abs/2407.06298>.
- [6] A. Lapkovskis, N. Nefedova, A. Beikmohammadi, Automatic fused multimodal deep learning for plant identification, Frontiers in Plant Science 16 (2025) 1616020. doi:10.3389/fpls.2025.1616020.
- [7] B. M. Mason, T. Mesaglio, J. Barratt Heitmann, M. Chandler, S. Chowdhury, S. B. Z. Gorta, F. Grattarola, Q. Groom, C. Hitchcock, L. Hoskins, S. K. Lowe, M. Marquis, N. Pernat, V. Shirey, S. Baasanmunkh, C. T. Callaghan, inaturalist accelerates biodiversity research, BioScience 75 (2025) 953–965. doi:10.1093/biosci/biaf104.
- [8] Pl@ntNet, Read more - pl@ntnet, Website, 2025. URL: <https://plantnet.org>, accessed: 2026-02-27.
- [9] A. Abdulhakim, Remote sensing and AI – opportunities for biodiversity conservation and nature management? — abdulhakim abdi, Blog post, 2025. URL: <https://hakimabdi.com>.
- [10] B. Janga, G. P. Asamani, Z. Sun, N. Cristea, A review of practical AI for remote sensing in earth sciences, Remote Sensing 15 (2023). doi:10.3390/rs15164112.
- [11] F. Ullah, S. Saqib, Y. C. Xiong, Integrating artificial intelligence in biodiversity conservation: bridging classical and modern approaches, Biodiversity and Conservation 34 (2024) 45–65. doi:10.1007/s10531-024-02977-9.
- [12] L. Picek, L. Adam, S. Kahl, R. Bossy, L. Chrobak, H. Goëau, K. Papafitsoros, H. Klinck, W.-P. Vellinga, 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 (CLEF), Springer, 2026.

²<https://github.com/VerbaNexAI/CLEF2026>

- [13] G. Martellucci, I. Moummad, H. Go"eau, 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.
- [14] 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, P. Bojanowski, DINOv2: Learning robust visual features without supervision, Transactions on Machine Learning Research (2024). URL: <http://arxiv.org/abs/2304.07193>.