Method title

Freshwater Cultural Ecosystem Services – Large Scale Modelling Framework

Authors

Comalada, Francesc[a,b]*. Acuña, Vicenç[a]. Garcia, Xavier[a]

Affiliations

[a]*Catalan Institute for Water Research (ICRA - CERCA), Carrer Emili Grahit 101, Girona, 17003, Spain*

[b]*University of Girona, Plaça de Sant Domènec 3, 17004 Girona, Spain*

Corresponding author's email address

fcomalada@icra.cat

Keywords

cultural ecosystem services, deep learning, freshwater ecosystems, Flickr

Related research article

Comalada, F., Acuña, V., Garcia, X. Modelling Cultural Ecosystem Services of river landscapes in the Iberian Peninsula using social media data. Under review in Journal of Environmental Management (*under review*).

Abstract

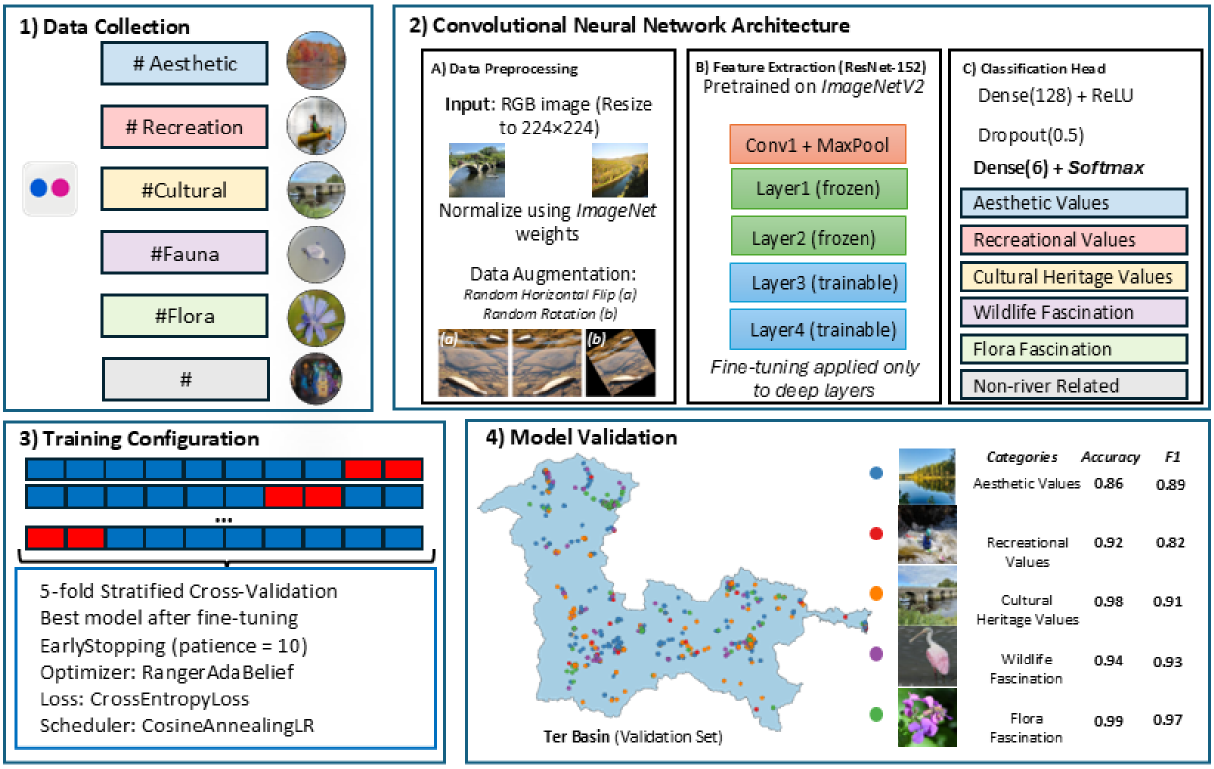
The increasing availability of georeferenced social media images offers new opportunities for mapping Cultural Ecosystem Services (CES) across large spatial scales. However, manually analyzing these datasets remains a significant bottleneck. This method presents *FreshCES-Net*, a scalable deep learning-based tool to classify freshwater-related CES from geotagged Flickr images using a ResNet-152 Convolutional Neural Network (CNN).

The model was trained and validated on over 9,000 manually annotated images representing five CES categories relevant to freshwater landscapes. To guarantee generalization across diverse regions, the training data was collected without geographical restrictions, while an independent test set from the Ter River Basin (northeastern Iberian Peninsula) was used for external validation. The model achieved 0.91 overall accuracy, with macro-averaged F1-scores exceeding 0.90 across categories.

This approach allows researchers and policymakers to automate CES classification from crowdsourced images, supporting evidence-based landscape planning and freshwater conservation efforts.

- Enables large-scale classification of freshwater CES from publicly available social media images.
- Applies a ResNet-152 CNN with transfer learning and cross-validated fine-tuning.
- Demonstrates high performance and generalization using an independent river basin case study.

Graphical abstract



Background

The increasing availability of user-generated, georeferenced images has opened new opportunities for large-scale assessment of Cultural Ecosystem Services (CES) (Langemeyer et al., 2023). While previous studies have demonstrated the utility of social media data to capture human–nature interactions at local scales (Cheng et al., 2019), their greatest potential lies in supporting large-scale assessments that can inform evidence-based landscape management and conservation planning (Ghermandi & Sinclair, 2019).

Despite this potential, the main challenge lies in efficiently extracting meaningful CES information from large image datasets. This includes category overlap, biased user demographics, geolocation inaccuracy (Zielstra & Hochmair, 2013), and visual ambiguity of intangible values (Chan et al., 2012). These issues limit classification precision and generalizability across landscapes. Thus, this methodology was developed to address that challenge by providing a scalable and reproducible framework for classifying freshwater CES images posted on social media using Convolutional Neural Networks (CNNs) (Zhang et al., 2022). The model was selected not for computational efficiency but for its ability to handle conceptual ambiguity and complex image semantics at scale (He et al., 2015).

The motivation is to enable researchers, stakeholders, and policymakers to employ a deep learning-based image classification model in scenarios that require the analysis of large-scale image datasets with minimal effort, time, and cost.

By training and fine-tuning a ResNet-152 CNN (He et al., 2015) on thousands of geotagged Flickr images representing diverse freshwater landscapes, this method supports freshwater CES categorization at large regional and temporal scales. The approach can be adapted to new geographic areas or CES classification schemes, making it broadly applicable (Beery et al., 2018).

Freshwater landscapes were chosen due to their ecological importance (Tickner et al., 2020) and their critical role in providing CES (Comalada et al., 2025). Freshwater landscapes are highly dynamic and support a wide range of non-material interactions, such as recreation, aesthetic enjoyment, and nature appreciation. At the same time, they are under increasing pressure from land-use change, pollution, and climate impacts (Tickner et al., 2020). Mapping CES in freshwater landscapes offers an opportunity to inform spatial planning and environmental management where trade-offs between ecological conservation and human use are most pronounced.

This methodology may be particularly useful for integrating CES into environmental planning in large regions, where traditional survey methods are unfeasible. By identifying landscape preferences, crowdsourced hotspots, or areas with additional CES value, this framework can guide conservation actions, public investment, and stakeholder engagement. This study builds on an evolving body of CES mapping literature using AI (e.g., Cardoso et al., 2022; Zhang et al., 2022) and addresses gaps identified in reviews such as Schirpke et al. (2023), which emphasize the need for scalable and transferable methods

Method details

This method presents a reproducible and scalable framework to classify CES in freshwater landscapes based on georeferenced social media images. The classification pipeline integrates preprocessing, data augmentation, transfer learning, and a five-fold cross-validation training procedure using a ResNet-152 CNN. The complete implementation, including code notebooks and the training and validation image datasets, is publicly available at the following GitHub repository: https://github.com/francesc30/Cultural-Ecosystem-Services-Modelling. Additionally, the repository provides an interactive Jupyter notebook designed to facilitate replication of the methodology for any user-defined region.

The pipeline is designed for researchers, policymakers and stakeholders aiming to integrate crowdsourced photographic data into environmental planning and CES assessment in a cost-effective way.

Step I: Definition of CES Categories Used for Image Classification

Before training CNN, we defined six image classification categories to leverage CES associated with freshwater landscapes (see Table 1).

Five categories correspond to CES, based on previously established conceptual frameworks: Aesthetic Values, Recreational Values, and Cultural Heritage Values, following Cheng et al. (2019); and Wildlife Fascination and Flora Fascination, adapted from Havinga et al. (2023). A sixth category (Non-freshwater CES) was included to exclude images unrelated to freshwater CES. These categories were operationalized for supervised image classification, ensuring conceptual clarity and consistency throughout the annotation and model training processes.

To address inter-category ambiguity (e.g., between Aesthetic and Recreational values), annotation followed a rule-based hierarchy: in scenes combining scenic beauty and visible human activity, 'Recreational' was prioritized.

Table 1: Definitions of the CES categories used to classify the images for model training.

Category	**Description**
Aesthetic values	Beauty or spectacularity of natural aspects of freshwater landscapes.
Recreational values	Sportive, leisure or relaxation activities practiced on freshwater landscapes.
Cultural Heritage values	Appreciation of modern and historical freshwater infrastructures for their functionality, aesthetic appeal, or cultural significance.
Wildlife fascination	Beauty and diversity of wildlife within freshwater ecosystem.
Flora fascination	Beauty and diversity of flora within freshwater ecosystems.
Non-freshwater CES	Photographs that do not relate to any of the defined freshwater CES.

Step II: Training and Validation Data

The ResNet152 CNN architecture (He et al., 2015) was trained to classify photographs into five predefined CES categories, and one extra category to discard those photos non-related to freshwater CES (see Table 2).

Table 2: Tags used for extracting training photos, detailed by category, along with the number of Flickr photos per category included in the training dataset.

Category	Keywords	n
Non-freshwater CES related	*no keywords*	1787
Aesthetic values	*Brooks, Cascade, Creek, FreshwaterBeauty, Gorge, Lagoon, Lake, Pond, Rapids, River, Rivers, Rivulet, Stream, Waterfall, Wetland*	1433
Recreational values	*Boating, Canyoning, Fishing, Kayak, Kayaking, Lake swimming, Rafting, River recreation, River swimming, Riverboarding*	931
Cultural Heritage values	*Aqueduct, Boat, Bridge, Dam, Drawbridge, Dyke, Hydroelectric, River ferry, Riverboat, Steamboat, Watermill, Weir*	608
Wildlife appreciation	*Amphibian, Animal, Ants, Biodiversity, Bird, Birds, Butterfly, Fauna, Frog, Insect, Lizard, Mammal, Reptile, Wetland fauna, Wildlife*	1282
Flora appreciation	*Angiospermae, Floral, Flora, Flower, Flowers, Flowering plants, Freshwater flora, River flora, River flower, Wildflowers*	870

Training data was collected from Flickr using the public API. For generalizability across diverse freshwater landscapes, images were retrieved using category-specific keyword queries without applying geographic filters. This improved the model's capacity to generalize across diverse geographic contexts, as large-scale assessments require (Beery et al., 2018). Only images taken between 2022 and 2024 were included to ensure temporal alignment and visual consistency with the validation dataset. This timeframe corresponds to the widespread availability of high-resolution smartphone cameras, helping to standardize image quality. This consideration is particularly relevant given that ResNet-152 processes input images resized to 224×224 pixels (Dodge & Karam, 2016).

To validate the representativeness and consistency of the classification process, 100 randomly selected images were independently annotated by three researchers. Inter-rater agreement was quantified using Cohen's Kappa coefficient (Cohen, 1960), resulting in a mean score of 0.959. This value indicates "almost perfect agreement" (McHugh, 2012), confirming the reliability of the manual annotation. Thus, a single researcher then labeled the complete training dataset, following the same category definitions used for the validation subset (See Table 2).

All images were subsequently pre-processed for CNN input. This involved resizing to 224×224 pixels, the required input dimension for the ResNet152 model pre-trained on the ImageNet1K dataset (Deng et al., 2009) (See Fig.1). Standard normalization was applied using ImageNet mean and standard deviation values after scaling pixel values to the [0,1] range. To prevent overfitting, data augmentation included RandomHorizontalFlip and RandomRotation (Shorten & Khoshgoftaar, 2019).

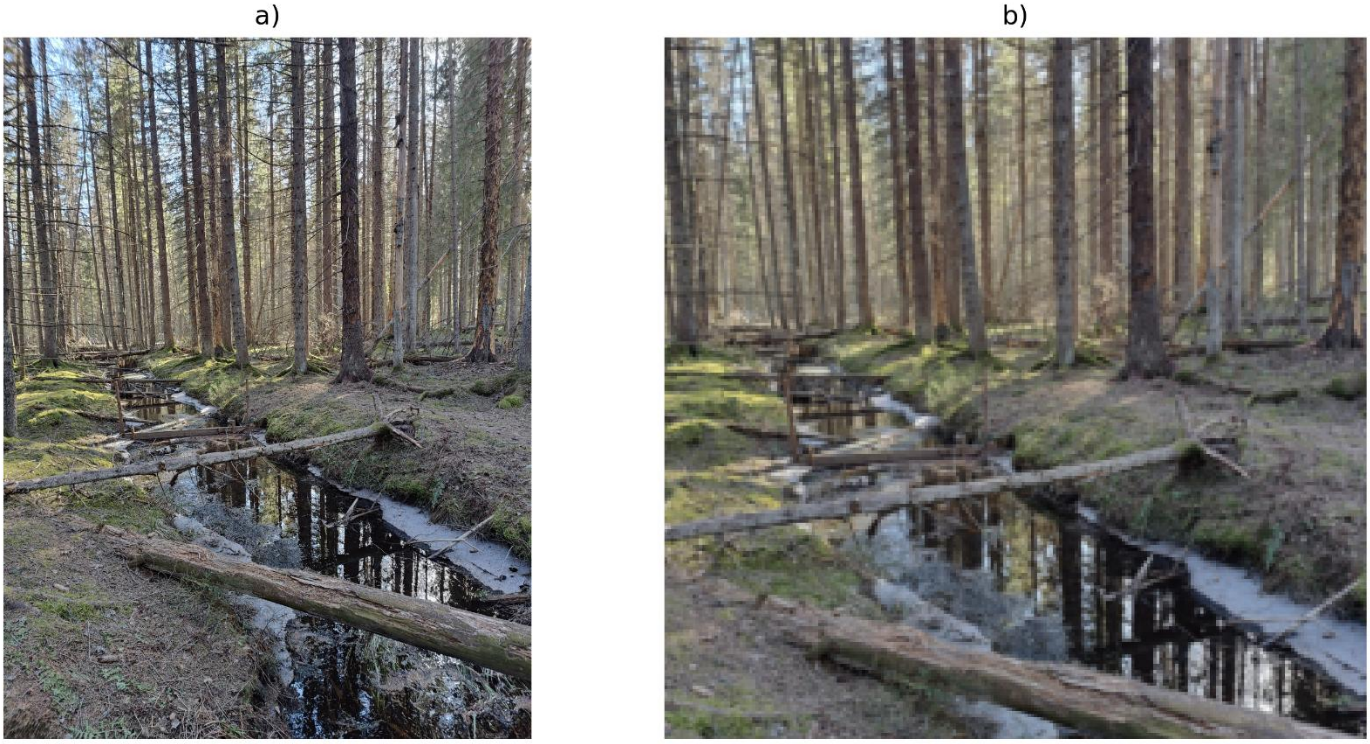


Figure 1. *Example of image quality variation used in model training. (a) Original high-resolution image. (b) Low-resolution variant generated by applying Gaussian blur and downsampling prior to resizing to 224×224 pixels for CNN input.*

Additionally, the Mixup regularization technique (Zhang et al., 2018) was used during training, synthetically generating intermediate images by linearly combining input samples and their corresponding labels. This technique is particularly suitable for image classification tasks involving conceptual and aesthetic categories, such as CES classes, where class boundaries are often visually ambiguous. Mixup encourages the model to learn smoother and more generalizable decision surfaces, improving robustness to label noise and atypical compositions.

In total, the final training dataset consisted of 7911 manually classified images distributed across the six CES categories. These were divided into five folds for cross-validation, with each fold containing a stratified representation of all classes.

An independent test set comprising 1351 additional Flickr images from the Ter River basin was used for final performance evaluation, as He et al. (2015) recommend (see Table 3).

Step III: Convolutional Neural Network (CNN) Training and Architecture

ResNet-152 architecture (He et al., 2015), a CNN designed to mitigate gradient issues through residual connections, was trained for CES classification.

This architecture was selected for its high performance in image recognition tasks and its proven suitability for transfer learning, particularly in environmental image classification contexts. The model was initialized with pre-trained ImageNet1K V2 weights (Deng et al., 2009), following evidence from Cardoso et al. (2022), who demonstrated that this configuration yields optimal performance in CES classification tasks.

To optimize training, the RangerAdaBelief optimizer (Zhuang et al., 2020) was used. This optimizer combines the benefits of Rectified Adam (RAdam) and Lookahead mechanisms and dynamically adapts learning rates based on the "belief" in gradient directions, making it particularly effective for transfer learning scenarios. Early stopping was implemented with a patience threshold of 10 epochs to minimize overfitting by halting training once the validation loss failed to improve consistently.

The loss function used was CrossEntropyLoss, the standard criterion for multi-class classification tasks, as it penalizes the divergence between predicted class probabilities and ground truth labels. All model development and training were conducted using the PyTorch framework (Paszke et al., 2019), within a Python 3.10.3 virtual environment configured for compatibility with PyTorch, TensorFlow (Abadi et al., 2016), and auxiliary libraries including RangerAdaBelief.

The model was trained using a batch size of 32 and a fixed initial learning rate of 0.001. The CosineAnnealingLR scheduler was configured with T_max = 10 and eta_min = 1e-6. Mixup augmentation was applied with an alpha parameter of 0.4. All random number generators (NumPy, PyTorch) were seeded with a fixed value of 42 to guarantee reproducibility across runs.

Step IV: Transfer Learning and Fine-tuning Strategy

To adapt the ResNet-152 model to the specific requirements of the CES classification task, we implemented a combination of transfer learning and fine-tuning. The initial layers of the network, specifically layer1 and layer2, were frozen to retain their pre-trained feature extraction capabilities. These layers are responsible for detecting low- and mid-level features, such as edges, textures, and geometric structures, which are generalizable across visual domains and particularly relevant for capturing recurring patterns in freshwater landscapes (Z. Zhang et al., 2021).

In contrast, the deeper layers of the network were fine-tuned using our labeled image set (see Table 1) to allow the model to learn high-level, domain-specific visual patterns. To adapt the network to our multi-class classification task, the original fully connected layer of ResNet-152 was removed and replaced by a new classification head. This custom head consisted of a dense layer with 128 hidden units, followed by a ReLU activation function, which introduces non-linearity and enables the model to learn complex relationships between features. To mitigate overfitting, we included a dropout layer with a dropout rate of 0.5, randomly deactivating a subset of neurons during each training iteration. Finally, an output layer was appended, with six units corresponding to the six CES categories, producing probability distributions over classes using softmax activation.

The model was trained using 5-fold cross-validation, implemented via the StratifiedKFold function from the scikit-learn library (Pedregosa et al., 2018), which guarantees that the class distribution is preserved across all folds. For each fold, 80% of the data was used for training and 20% for validation. Training proceeded for a maximum of 100 epochs per fold. However, to prevent overfitting and optimize training time, we applied early stopping (Prechelt, 1998), which interrupted training if no improvement in validation loss was observed over 10 consecutive epochs. After completing all folds, the model corresponding to the fold with the best validation performance was selected for final evaluation and deployment.

All training and evaluation processes were executed on a workstation equipped with an NVIDIA RTX 3090 GPU (24 GB VRAM), 128 GB RAM, and an AMD Ryzen Threadripper 3970X CPU. Training time per fold varied between 15 and 35 minutes, depending on early stopping, with a total training time of approximately 2.5 hours for all five folds. The final model had approximately 60.2 million trainable parameters in the fine-tuned layers.

Step V: Model Evaluation and Performance

Model performance was assessed using 5-fold stratified cross-validation. The mean validation accuracy across folds was 97.19% (±0.23), and the mean validation loss was 0.139 (±0.014). The model consistently performed well across all classes, indicating strong generalization capacity. The best-performing model, corresponding to Fold 1, was selected based on minimum validation loss (0.1145).

Step VI: Method Adaptability

The proposed framework is designed to be fully adaptable and extensible. Researchers can retrain or fine-tune the model using their own image datasets. This flexibility allows for the incorporation of alternative classification schemes beyond the six CES categories, or for the adaptability to other landscapes, such as coastal, alpine, and others.

For optimal compatibility with the pretrained visual features extracted from ImageNet, we recommend using high-resolution images taken from 2020 onward, as older content may suffer from reduced quality, resolution, or outdated photographic styles (Dodge & Karam, 2016).

The Supplementary material provides the code to retrain the model, and the GitHub repository the templates and guidance to facilitate this process.

Method validation

Step I: Validation dataset

The Validation area used for external model evaluation is the Ter River Basin, located in the northeastern

Iberian Peninsula. This basin covers a surface area of approximately 3,010 km² and is part of the internal basins of Catalonia, an administrative water management region under the jurisdiction of the Catalan Water Agency (ACA). The Ter River originates in the Pyrenees at an elevation of 2,900 meters above mean sea level and flows southeast into the Mediterranean Sea (see Fig. 3).

The basin exhibits a complex hydrological and climatic regime, shaped by both rainfall and snowfall. It is subject to seasonal peak flows, typically during spring and autumn, and displays frequent intermittent flows in tributaries as well as episodic floods, particularly in response to convective precipitation events. Four distinct climate zones converge within the basin: Polar, cold, temperate and arid (Beck et al., 2018). This climatological heterogeneity, combined with sharp altitudinal gradients, contributes to significant landscape and ecological diversity.

Despite this natural richness, the basin is also heavily impacted by human activity. With a resident population of approximately 187,000 inhabitants (62 inhabitants/km²), the Ter River is subject to extensive hydromorphological alterations for hydropower production and water supply, most notably through the *Sau*, *Susqueda*, and *Pasteral* reservoirs.

Given its combination of natural and anthropogenic features, the Ter River basin offers an ideal case study for evaluating CES-related image classification models in a real-world, heterogeneous freshwater landscape.

For this reason, we used unseen Flickr images captured within the Ter Basin between 2022 and 2024 to test the generalization capacity of the trained CNN model, following the external validation protocol recommended by He et al. (2015).

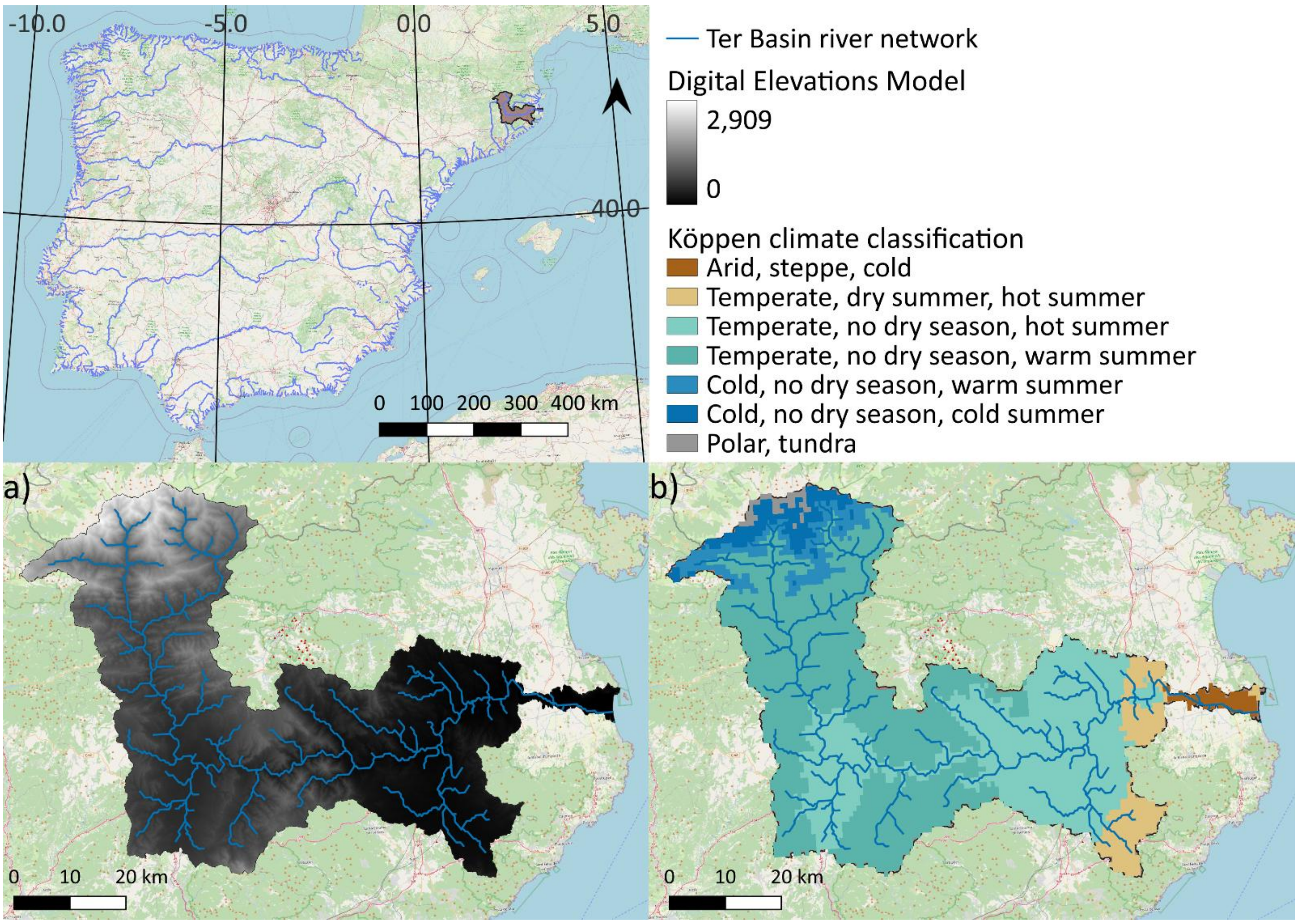


Figure 3. Geographical context and spatial datasets used for the Ter River Basin in northeastern Iberia. (a) Digital Elevation Model (DEM) with delineated river network. (b) Köppen–Geiger climate classification zones across the basin.Base map: OpenStreetMap. Climate classification from Beck et al. (2018).

Step II: Model Performance on Independent Test Set

The performance of the ResNet-152 model was evaluated on an independent test set consisting of 1,351 previously unseen Flickr images taken between 2022 and 2024 from freshwater landscapes within the Ter

River Basin. The model achieved an overall accuracy of 0.91, defined as the percentage of correctly classified images across all categories. Compared to previous studies using ResNet-152 for CES classification (e.g., Cardoso et al., 2022: F1 = 0.69; Lingua et al., 2022: F1 = 0.83), our model achieved a superior F1 score of 0.91 on an independent test set, evidencing enhanced generalization.

Table 3 provides detailed performance metrics per class, including precision, recall, F1-score, and support.

Table 3: AI performance metrics results by category. These range from 0 to 1 (or 0-100%), where 1 is the best value, and are defined as Precision: Percentage of correct predictions among all instances predicted for a class. Recall: Percentage of positive cases correctly identified by the model. F1-score: Harmonic mean of precision and recall. Support: Number of photographs used for validation.

Category	Precision	Recall	f1-score	Support
No river CES Related	0.87	0.96	0.91	404
Aesthetic Values	0.86	0.93	0.89	248
Recreational Values	0.92	0.74	0.82	31
Cultural Heritage Values	0.98	0.85	0.91	363
Wildlife Fascination	0.94	0.92	0.93	91
Flora Fascination	0.99	0.95	0.97	214
macro average	0.93	0.89	0.91	1351
weighted average	0.92	0.91	0.92	1351

Figure 4 suggests that the Aesthetic Values category demonstrated strong classification performance, with an F1-score of 0.89, driven primarily by a high recall of 0.93. This indicates the model was highly sensitive to detecting true aesthetic cases. However, the precision was slightly lower (0.86), suggesting a moderate rate of false positives, likely arising from the visual overlap between aesthetic and other CES-related images, particularly those featuring shared elements such as vegetation, water bodies, or panoramic views.

The Recreational Values Category exhibited a contrasting pattern, with a high precision of 0.92 but a lower recall of 0.74. This asymmetry suggests that the model adopted a conservative strategy for assigning the recreational label, correctly identifying most predictions as recreational but omitting a proportion of actual instances. A similar imbalance was observed for Cultural Heritage Values, where precision was high (0.98) but recall dropped to 0.85, likely due to visual ambiguity between historical infrastructure and natural or aesthetic scenes.

The model performed particularly well in the Non-freshwater CES-related class, achieving a recall of 0.96, indicating that it correctly filtered out most irrelevant images. The slightly lower precision (0.87) implies occasional misclassification of CES-relevant images into this exclusion class, possibly due to weak CES visual signals or ambiguous framing.

Performance on the Wildlife fascination and Flora fascination categories was consistently high, with both classes exceeding 0.92 on all major metrics. Notably, Flora fascination reached an F1-score of 0.97, reflecting the model's robustness in detecting characteristic features of riparian and aquatic plant life.

Overall, the model yielded a macro-average F1-score of 0.91, confirming its effectiveness across conceptually and visually diverse CES categories.

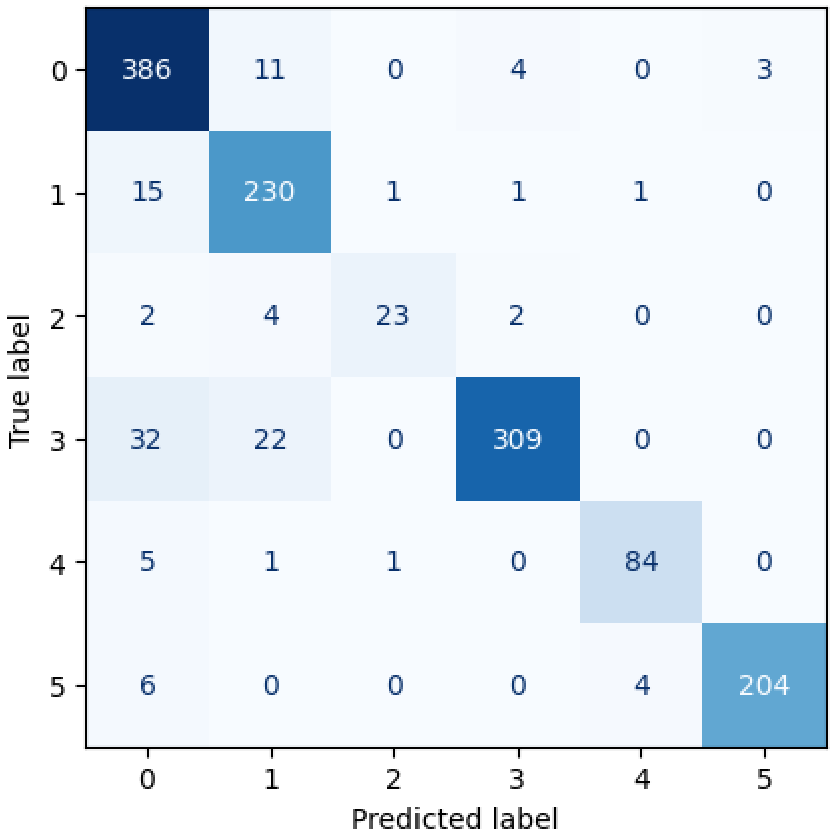


Figure 4. Confusion matrix of the ResNet-152 model on the independent Ter River Basin test set (n = 1,351). Class labels: (0) Non-river CES, (1) Aesthetic values, (2) Recreational values, (3) Cultural Heritage values, (4) Wildlife fascination, (5) Flora fascination. Rows represent the true CES category labels, and columns represent the predicted labels. Diagonal cells indicate correct classifications. Values are absolute counts.

Limitations

The method presented in this study has several limitations that should be considered when interpreting the results or applying the approach to other contexts.

A fundamental limitation concerns the representativeness and quality of the training dataset. While Flickr provides access to a broad and diverse set of user-generated images, this source introduces uncontrolled variability in image composition, resolution, and metadata accuracy. Furthermore, Flickr users represent a demographically and behaviorally biased subset of the population (Leppämäki et al., 2025), and image uploads tend to cluster around visually appealing or touristic freshwater sites (Llanos-Páez & Acuña, 2022). As a result, certain CES types and landscape contexts may be underrepresented. Additionally, the Flickr API can yield inconsistent retrieval results over time, limiting reproducibility (Leppämäki et al., 2025), and manual geotagging can introduce spatial mismatches (& Hochmair, 2013). Also, While images from 2022–2024 ensured resolution consistency (Dodge & Karam, 2016), future should will evaluate model performance decay on older (pre-2020) datasets to assess robustness over time

From a technical standpoint, the performance of the model is tightly coupled to the structure of the training data. The CES categories used, particularly Aesthetic values and Recreational values, are often overlapping and visually similar, leading to semantic ambiguity and misclassifications in hybrid or context-dependent scenes. The use of a single-label classification scheme further limits the ability to represent multifunctional landscapes, where several CES may co-occur in a single image (Fish et al., 2016). Moreover, the dataset suffers from class imbalance, especially for Recreational Values, which may bias the model toward dominant categories even with stratified cross-validation.

Another limitation is the exclusive reliance on visual data. Intangible CES dimensions, such as sense of place, spiritual significance, or symbolic meaning, cannot be adequately captured through images alone (Chan et al., 2012). The method is also restricted to static imagery; it does not account for temporal or experiential aspects of CES perception, which may be critical in fields such as environmental psychology or participatory planning.

Finally, the method is currently tailored to freshwater landscapes. While the underlying CNN architecture is transferable, adapting it to other ecosystem types (e.g., coastal, forest, or agricultural landscapes) would require revisiting the CES classification schema, retraining with domain-specific imagery, and revalidating performance metrics.

Ethics statements

All data used in this study were obtained through Flickr's official public API and consisted exclusively of publicly available photographs. While no interaction with users occurred, and no personally identifiable content was downloaded or stored, certain metadata associated with the images such as geolocation coordinates, user identifiers, and timestamps may be considered sensitive.

To address this, all data were fully anonymized prior to analysis. No results were interpreted or reported with reference to individual users. All interpretations were conducted at the aggregated level, and care was taken to avoid any inference about personal behavior or identity. The use of the data complies with Flickr's terms of service and data redistribution policies. Informed consent was not required as per current ethical guidelines governing the use of public user-generated content for research purposes, but ethical responsibility in data handling and interpretation was explicitly maintained throughout the study.

CRediT author statement

F. Comalada: Writing – review and editing, Writing – original draft, Visualization, Validation, Resources, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **V. Acuña:** Supervision, Resources, Project administration, Investigation, Conceptualization. **X. Garcia:** Writing – review and editing, Visualization, Supervision, Investigation, Conceptualization.

Acknowledgments

This research was funded by the European Union Horizon 2020 project MERLIN (H2020-LC-GD-2020-3: 101036337). Authors acknowledge the support from the Economy and Knowledge Department of the Catalan Government through Consolidated Research Groups (ICRA-ENV 2021 SGR 01282), as well as from the CERCA program.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Supplementary material *and/or* additional information

The complete model script is provided.

Also, the IDs of the Flickr photos used for training and unseen validation, with its corresponding category, stored in an Excel spreadsheet.

Moreover, a GitHub repository contains a reproducible implementation of the framework, including a user-friendly interactive script for model deployment and evaluation

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. arXiv preprint arXiv:1605.08695. https://doi.org/10.48550/arXiv.1605.08695

Beery, S., Van Horn, G., & Perona, P. (2018). Recognition in Terra Incognita. arXiv preprint arXiv:1807.04975. https://doi.org/10.48550/arXiv.1807.04975

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. Scientific Data, 5(1), 180214. https://doi.org/10.1038/sdata.2018.214

Cardoso, A. S., Renna, F., Moreno-Llorca, R., Alcaraz-Segura, D., Tabik, S., Ladle, R. J., & Vaz, A. S. (2022). Classifying the content of social media images to support cultural ecosystem service assessments using deep learning models. Ecosystem Services, 54, 101410. https://doi.org/10.1016/j.ecoser.2022.101410

Chan, K. M. A., Satterfield, T., & Goldstein, J. (2012). Rethinking ecosystem services to better address and navigate cultural values. Ecological Economics, 74, 8–18. https://doi.org/10.1016/j.ecolecon.2011.11.011

Cheng, X., Van Damme, S., Li, L., & Uyttenhove, P. (2019). Evaluation of cultural ecosystem services: A review of methods. Ecosystem Services, 37, 100925. https://doi.org/10.1016/j.ecoser.2019.100925

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37–46. https://doi.org/10.1177/001316446002000104

Comalada, F. (2025). Cultural Ecosystem Services Modelling [GitHub repository]. https://github.com/francesc30/Cultural-Ecosystem-Services-Modelling

Comalada, F., Llorente, O., Acuña, V., Saló, J., & Garcia, X. (2025). Using georeferenced text from social media to map the cultural ecosystem services of freshwater ecosystems. Ecosystem Services, 72, 101702. https://doi.org/10.1016/j.ecoser.2025.101702

Da Mota, V. T., & Pickering, C. (2020). Using social media to assess nature-based tourism: Current research and future trends. Journal of Outdoor Recreation and Tourism, 30, 100295. https://doi.org/10.1016/j.jort.2019.100295

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition (pp. 248–255). IEEE. https://doi.org/10.1109/CVPR.2009.5206848

Dodge, S., & Karam, L. (2016). Understanding How Image Quality Affects Deep Neural Networks (No. arXiv:1604.04004). arXiv. https://doi.org/10.48550/arXiv.1604.04004

Fish, R., Church, A., & Winter, M. (2016). Conceptualising cultural ecosystem services: A novel framework for research and critical engagement. Ecosystem Services, 21, 208–217. https://doi.org/10.1016/j.ecoser.2016.09.002

Ghermandi, A., & Sinclair, M. (2019). Passive crowdsourcing of social media in environmental research: A systematic map. Global Environmental Change, 55, 36–47. https://doi.org/10.1016/j.gloenvcha.2019.02.003

Havinga, I., Marcos, D., Bogaart, P., Massimino, D., Hein, L., & Tuia, D. (2023). Social media and deep learning reveal specific cultural preferences for biodiversity. People and Nature, 5(3), 981–998. https://doi.org/10.1002/pan3.10466

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385. https://doi.org/10.48550/arXiv.1512.03385

Langemeyer, J., Ghermandi, A., Keeler, B., & van Berkel, D. (2023). The future of crowd-sourced cultural ecosystem services assessments. Ecosystem Services, 60, 101518. https://doi.org/10.1016/j.ecoser.2023.101518

Leppämäki, T., Heikinheimo, V., Eklund, J., Hausmann, A., & Toivonen, T. (2025). The rise and fall of the social media platform Flickr: Implications for nature recreation research. Journal of Outdoor Recreation and Tourism, 50, 100880. https://doi.org/10.1016/j.jort.2025.100880

Lingua, F., Coops, N. C., & Griess, V. C. (2022). Valuing cultural ecosystem services combining deep learning and benefit transfer approach. Ecosystem Services, 58, 101487. https://doi.org/10.1016/j.ecoser.2022.101487

Llanos-Páez, O., & Acuña, V. (2022). Analysis of the socio-ecological drivers of the recreational use of temporary streams and rivers. Science of the Total Environment, 807, 150805. https://doi.org/10.1016/j.scitotenv.2021.150805

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. Biochemia Medica, 22(3), 276–282. https://doi.org/10.11613/BM.2012.031

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. arXiv preprint arXiv:1912.01703. https://doi.org/10.48550/arXiv.1912.01703

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2018). Scikit-learn: Machine learning in Python. arXiv preprint arXiv:1201.0490. https://doi.org/10.48550/arXiv.1201.0490

Prechelt, L. (1998). Early stopping—But when? In G. B. Orr & K.-R. Müller (Eds.), Neural Networks: Tricks of the Trade (pp. 55–69). Springer. https://doi.org/10.1007/3-540-49430-8_3

Schirpke, U., Ghermandi, A., Sinclair, M., Van Berkel, D., Fox, N., Vargas, L., & Willemen, L. (2023). Emerging technologies for assessing ecosystem services: A synthesis of opportunities and challenges. Ecosystem Services, 63, 101558. https://doi.org/10.1016/j.ecoser.2023.101558

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6, 60. https://doi.org/10.1186/s40537-019-0197-0

Tickner, D., Opperman, J. J., Abell, R., Acreman, M., Arthington, A. H., Bunn, S. E., ... & Young, L. (2020). Bending the curve of global freshwater biodiversity loss: An emergency recovery plan. BioScience, 70(4), 330–342. https://doi.org/10.1093/biosci/biaa002

Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2018). Mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412. https://doi.org/10.48550/arXiv.1710.09412

Zhang, H., Huang, R., Zhang, Y., & Buhalis, D. (2022). Cultural ecosystem services evaluation using geolocated social media data: A review. Tourism Geographies, 24(4–5), 646–668. https://doi.org/10.1080/14616688.2020.1801828

Zhang, Z., Lu, M., Ji, S., Yu, H., & Nie, C. (2021). *Rich CNN Features for Water-Body Segmentation from Very High Resolution Aerial and Satellite Imagery. Remote Sensing, 13(10), Article 10.* https://doi.org/10.3390/rs13101912

Zhuang, J., Tang, T., Ding, Y., Tatikonda, S. C., Dvornek, N., Papademetris, X., & Duncan, J. S. (2020). AdaBelief Optimizer: Adapting stepsizes by the belief in observed gradients. Advances in Neural Information Processing Systems, 33, 18795–18806.

Zielstra, D., & Hochmair, H. H. (2013). Positional accuracy analysis of Flickr and Panoramio images for selected world regions. Journal of Spatial Science, 58(2), 251–273. https://doi.org/10.1080/14498596.2013.801331