École Polytechnique Fédérale de Lausanne

Image Processing for Earth Observation
ENV-540

Mapping swiss ecosystems from aerial images and environmental variables

Dany Montandon 357827
Loïc Trochen 364251

Teacher
Devis Tuia

Wednesday 14th January, 2026

1 Introduction

Mapping and monitoring ecosystems is a central task in environmental science, providing essential information for biodiversity conservation, land-use planning, and ecological research. In Switzerland, ecosystem classification plays a key role in understanding spatial patterns of habitats and in supporting evidence-based management decisions. Recent advances in remote sensing and machine learning now enable large-scale, data-driven ecosystem mapping by leveraging both high-resolution aerial imagery and heterogeneous environmental datasets.

This project focuses on predicting ecosystem categories across Switzerland by jointly exploiting aerial orthophotos from the swisstopo *swissIMAGE* product and a comprehensive suite of 48 environmental variables from the SWECO25 database. The provided dataset contains 16,925 georeferenced locations, each associated with a 100 m × 100 m RGB aerial image, standardized environmental variables describing climatic, edaphic, geological, hydrological, and land-use properties, and an ecosystem label based on the EUNIS classification system comprising 17 categories [swisstopo, 2020, Kulling et al., 2024, Chytrý et al., 2020].

The overarching goal of the project is to design a deep learning framework capable of predicting ecosystem classes from these multimodal inputs. Beyond building an accurate classifier, the study aims to disentangle the respective contributions of remote sensing imagery and environmental variables, assess the added value of each data modality, and identify which environmental factors encode information that is not directly observable from aerial images alone. This involves developing an appropriate training pipeline, selecting and tuning model architectures, evaluating performance quantitatively on a geographically separated test set, and interpreting the resulting predictions in an ecologically meaningful way.

Ultimately, this work contributes to understanding how modern machine learning techniques can support automated, scalable, and interpretable ecosystem mapping. Which is a critical component in the sustainable management of natural environments.

2 Method

2.1 Dataset and Study Area

The study is based on a dataset of 16,925 georeferenced locations distributed across Switzerland. Each sample is associated with three types of information: an RGB aerial image, a set of environmental variables, and an ecosystem label.

The aerial images come from the swissIMAGE product and cover an area of 100 m × 100 m at a spatial resolution of 50 cm. Only the RGB bands are used. Environmental information is provided by the SWECO25 database and consists of 48 numerical variables describing climatic, edaphic (soil), vegetation, land use and land cover, geology, hydrology, and population or infrastructure characteristics. All environmental variables are standardized to zero mean and unit variance. Ecosystem labels follow the EUNIS classification system and include 17 mutually exclusive ecosystem classes.

A predefined geographic split is used to divide the dataset into training (60%), validation (10%), and test (30%) sets. This split is respected throughout the study to avoid spatial autocorrelation between training and evaluation data. Model selection and hyperparameter tuning are performed on the validation set, while final performance is reported on the test set only.

2.2 Data Preprocessing

2.2.1 Image Preprocessing

All aerial images are resized to a fixed spatial resolution of 224 × 224 pixels to ensure compatibility with standard convolutional neural network (CNN) backbones. Images are converted to tensors and normalized using ImageNet mean and standard deviation values. Although aerial images differ from natural images, this normalization strategy is commonly used in transfer learning and provides stable training behavior when using pretrained models.

Data augmentation is applied only to the training images to reduce overfitting and improve model robustness. The augmentation strategy includes random resized cropping, horizontal and vertical flipping, and color jittering.

These transformations introduce variability in scale, orientation, and illumination while preserving the semantic content of the images. Validation and test images are processed using deterministic transformations without augmentation.

2.2.2 Tabular Data Preprocessing

The tabular input consists of the 48 SWECO environmental variables. Non-feature columns such as spatial coordinates, split identifiers, and label information are excluded from the input features. No additional normalization is applied, as the variables are already standardized in the original dataset. A systematic check confirms that no missing values are present across the environmental variables.

2.3 Problem Formulation

The task is formulated as a supervised multiclass classification problem. Given an input sample consisting of an aerial image, an environmental variable vector, or both, the goal is to predict the corresponding EUNIS ecosystem class among 17 possible categories. Models are trained to output a probability distribution over classes, and the predicted class corresponds to the maximum probability.

2.4 Image-Based Models

Image-based ecosystem classification is performed using convolutional neural networks pretrained on ImageNet. Transfer learning is employed to take advantage of robust visual features learned from large-scale image datasets, which is particularly beneficial given the moderate size of the available training data.

All image-based models rely on a ResNet-18 backbone as a common feature extractor. ResNet-18 is a widely used architecture with residual connections that facilitate stable training. Several image-based architectures are explored, including a standard fine-tuned ResNet model and variants that modify how image features are aggregated and processed before classification. Using a common backbone allows for a controlled comparison of different design choices while keeping the overall model capacity comparable.

2.5 Tabular-Only Models

To model the environmental variables, several multilayer perceptron (MLP) architectures are evaluated. These models take the 48 SWECO variables as input and consist of fully connected layers with ReLU activations. Batch normalization is applied to stabilize training, and dropout is used to reduce overfitting.

Both deeper and shallower MLP configurations are considered to study the effect of model complexity on performance. In addition, an architecture with residual (skip) connections is tested to improve optimization and gradient flow in deeper networks. All tabular models output a vector of 17 logits corresponding to the ecosystem classes.

2.6 Multimodal Model

To jointly exploit information from aerial imagery and environmental variables, a multimodal neural network architecture is implemented. The model consists of two parallel branches: an image branch and a tabular branch.

The image branch extracts visual features using a CNN based on a ResNet-18 backbone, while the tabular branch processes environmental variables using an MLP. Each branch produces a latent feature representation. These representations are concatenated to form a joint feature vector, which is then passed through a fully connected classification head to predict the ecosystem class.

This feature-level fusion strategy allows the model to combine spatial patterns visible in aerial images with ecological context provided by environmental variables, enabling a direct comparison between unimodal and multimodal approaches.

2.7 Loss Function and Optimization

All models are trained using categorical cross-entropy loss, which is appropriate for multiclass classification with mutually exclusive classes. When training on subsets of the data, class-weighted cross-entropy can be used to account for potential class imbalance.

Model parameters are optimized using the Adam optimizer, with optional weight decay for regularization. Learning rates and other optimization hyperparameters are selected based on validation performance.

2.8 Experimental Design

Three main model families are compared: image-only models, tabular-only models using all SWECO variables, and multimodal models combining both data sources. This comparison allows the contribution of each data modality to ecosystem classification to be assessed.

To analyze the importance of environmental variables, a permutation-based ablation study is performed on the tabular models. Individual variables are randomly permuted, and the resulting decrease in performance is measured. Variables are also grouped into thematic categories, and average importance is computed at the group level to identify the most influential ecological drivers.

Hyperparameters such as learning rate, batch size, dropout rate, and weight decay are explored using a limited set of configurations. The best model is selected based on validation macro F1-score, which accounts for class imbalance. Early stopping and regularization techniques are applied to reduce overfitting and improve generalization.

3 Results and Discussion

This section presents and discusses the quantitative and qualitative results obtained for ecosystem classification in Switzerland. We first compare the different model architectures and training behaviors, then evaluate their performance on the test set. Based on these results, we select the best-performing model for a more detailed analysis of class-level performance, confusion patterns, environmental variable importance, and spatial prediction behavior.

3.1 Comparison of Model Architectures

Figure 1 compares validation accuracy and macro-averaged F1-score across all tested architectures and hyperparameter configurations. Clear differences are observed between tabular-only, image-only, and combined tabular–image models.

Among the tabular-only approaches, the *TabularStandard* model achieves the highest validation performance, with a validation accuracy of approximately 0.57 and a macro F1-score close to 0.56. It also maintains relatively low training time per epoch, making it both effective and computationally efficient. Other tabular variants, such as *TabularSkip* and *TabularShallow*, show slightly lower performance, with the shallow architecture being the fastest to train.

In contrast, image-only models based on RGB aerial imagery perform substantially worse than tabular-based approaches, with validation macro F1-scores remaining below 0.47 across all tested architectures. Despite the use of more advanced designs such as hypercolumn feature extraction, these models struggle to capture sufficient discriminative information and require significantly longer training times due to the convolutional backbone.

The combined tabular–image models achieve competitive performance compared to image-only models but do not surpass the tabular-only baseline. For instance, the *Combined_Resnet18* model reaches a validation accuracy of around 0.55, but all combined architectures incur a much higher computational cost, with average training times per epoch exceeding 35 seconds. This increased cost is mainly driven by image feature extraction, while the resulting performance gain remains limited.

Overall, these results indicate that tabular environmental variables carry the strongest predictive signal for ecosystem classification in this setting. Incorporating aerial imagery increases model complexity and training time without providing a clear improvement in validation performance. The figure also shows that tabular models exhibit consistent performance across different hyperparameter configurations, whereas image-based and combined models are more sensitive to hyperparameter choice.

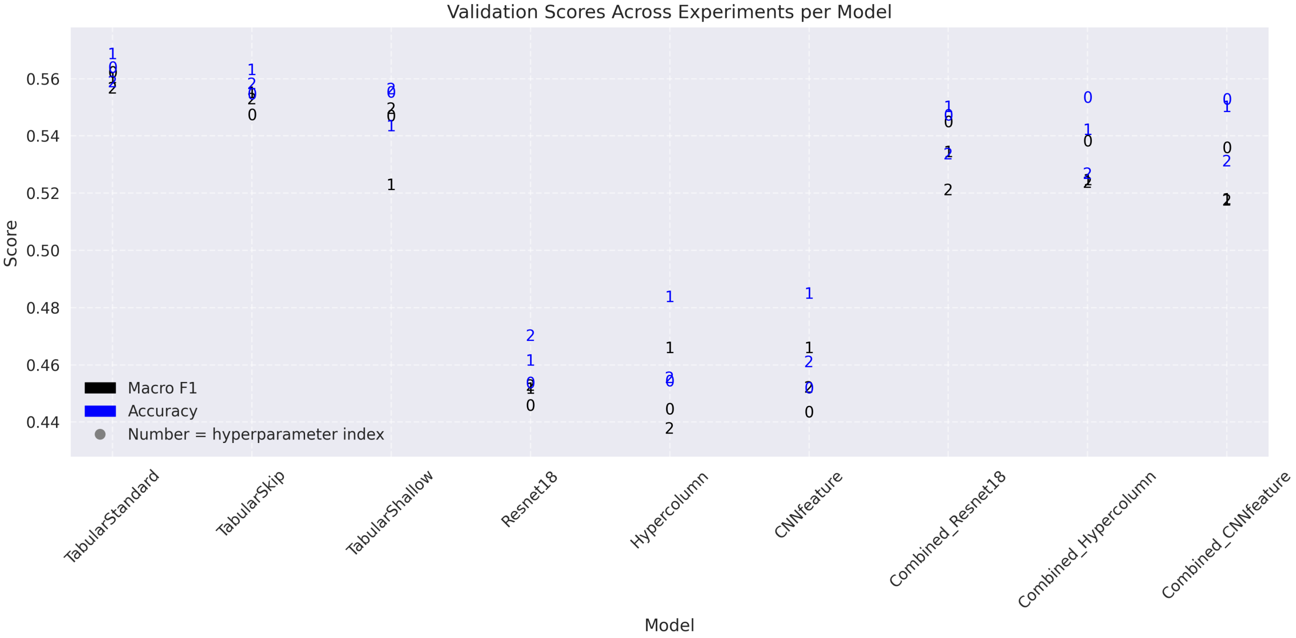


Figure 1: Validation accuracy and macro F1-score across model architectures and hyperparameter configurations.

3.2 Training Dynamics and Model Convergence

Representative training and validation curves for tabular and image-based models are shown in Figures 3, 2, and 4. The figures report both accuracy and macro F1-score, allowing an analysis of convergence behavior and generalization performance.

The tabular model exhibits smooth and stable convergence, as shown in Figure 2. Training and validation accuracy increase rapidly during the first epochs and then stabilize, while the validation macro F1-score remains close to the training macro F1-score throughout training. The small and stable gap between training and validation curves indicates limited overfitting and good generalization. Extending training up to 200 epochs does not degrade validation performance, confirming that multilayer perceptrons are well suited for structured environmental data.

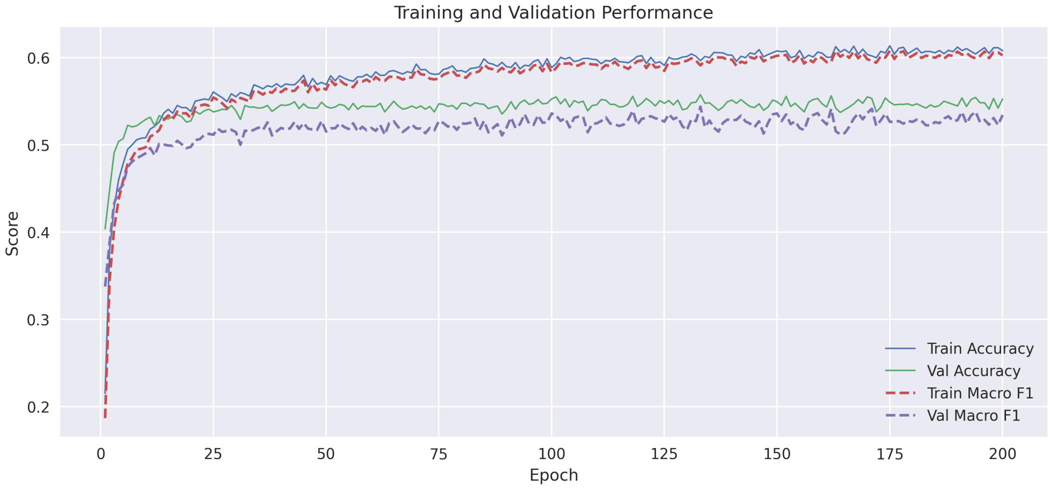


Figure 2: Training and validation accuracy and macro F1-score for the shallow tabular MLP model.

In contrast, image-only models show a markedly different training behavior. For the CNN-based image model (Figure 3), training performance increases steadily, while validation performance peaks early and then decreases, indicating overfitting.

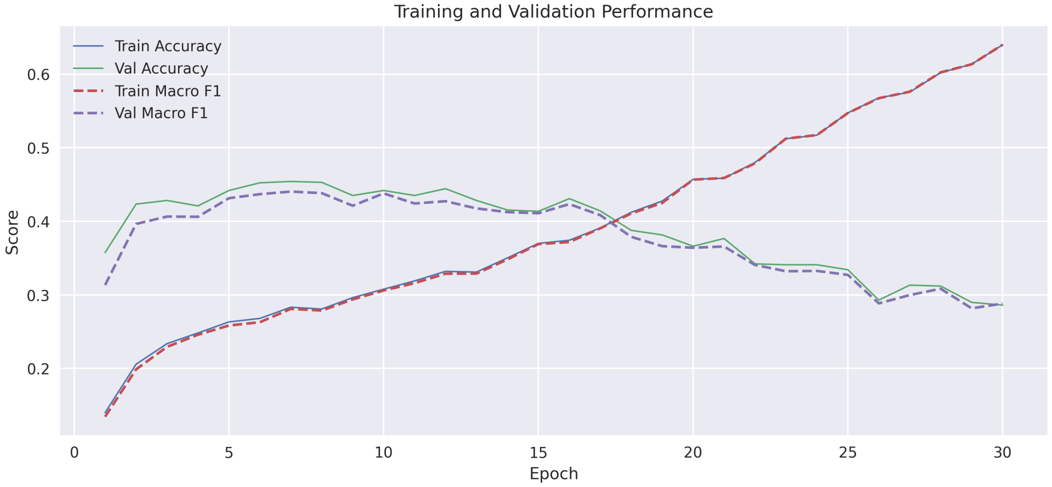


Figure 3: Training and validation accuracy and macro F1-score for the CNN-based image model.

A similar pattern is observed for the hypercolumn-based image model (Figure 4), where validation metrics fluctuate after early epochs and do not show sustained improvement.

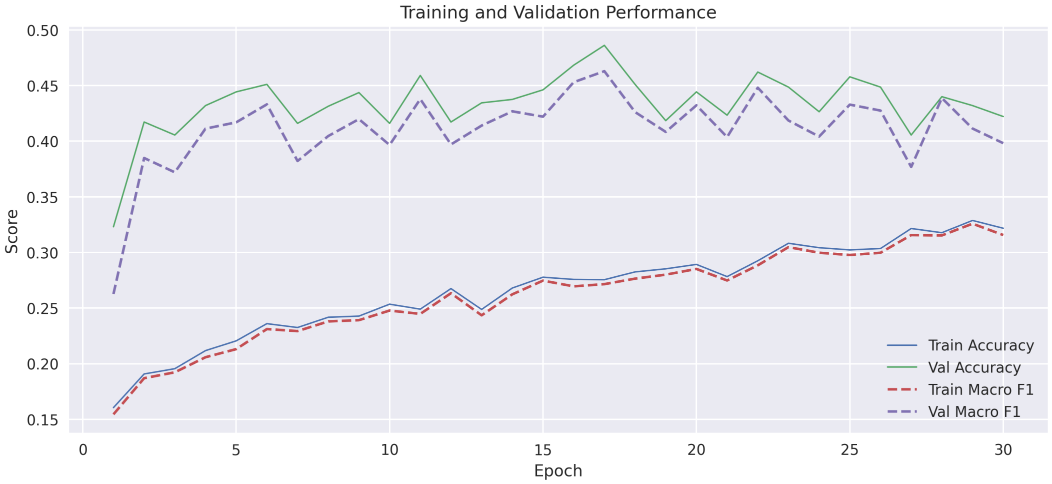


Figure 4: Training and validation accuracy and macro F1-score for the hypercolumn-based image model.

Multimodal models show training dynamics similar to image-only models and were trained on a reduced subset of 4,000 samples due to computational constraints. This reduced data availability likely limited their ability to fully exploit image information and explains why combined models did not surpass the tabular-only baseline.

Based on these observations, training of image-based and multimodal models was limited to 30 epochs, with early stopping used to avoid overfitting.

3.3 Performance on the Test Set

The best-performing model of each architecture type is evaluated on the held-out test set using overall accuracy, macro-averaged F1-score, per-class F1-score, and inference time per sample.

The *TabularStandard* model achieves the best overall test performance, with the highest accuracy and macro F1-score, while also being extremely efficient at inference time. The combined *ResNet18* model performs slightly worse, followed by the image-only *Hypercolumn* model, which shows the lowest performance.

Interestingly, both the tabular and combined models achieve slightly better results on the test set than on validation, suggesting good generalization to unseen data. Inference times remain very low for all models, indicating that they are suitable for large-scale ecosystem mapping.

Based on these results, the *TabularStandard* model is selected as the primary reference for the detailed analyses presented in the following subsections.

3.4 Per-Class Performance and Confusion Analysis

Per-class F1-scores reveal substantial variation across ecosystem classes, highlighting differences in classification difficulty. Figure 5 shows the per-class F1-scores for the *TabularStandard* model on the test set.

High F1-scores are achieved for classes such as surface waters, forests, arable land, and urban environments, which are characterized by clear environmental signatures. In contrast, lower performance is consistently observed for classes such as mesic grasslands, alpine and subalpine grasslands, and other heterogeneous ecosystems.

Image-only and combined models show similar class-wise trends, indicating that the same ecosystem types are challenging across architectures. This suggests that these difficulties are driven by intrinsic ecological similarity rather than by model-specific limitations.

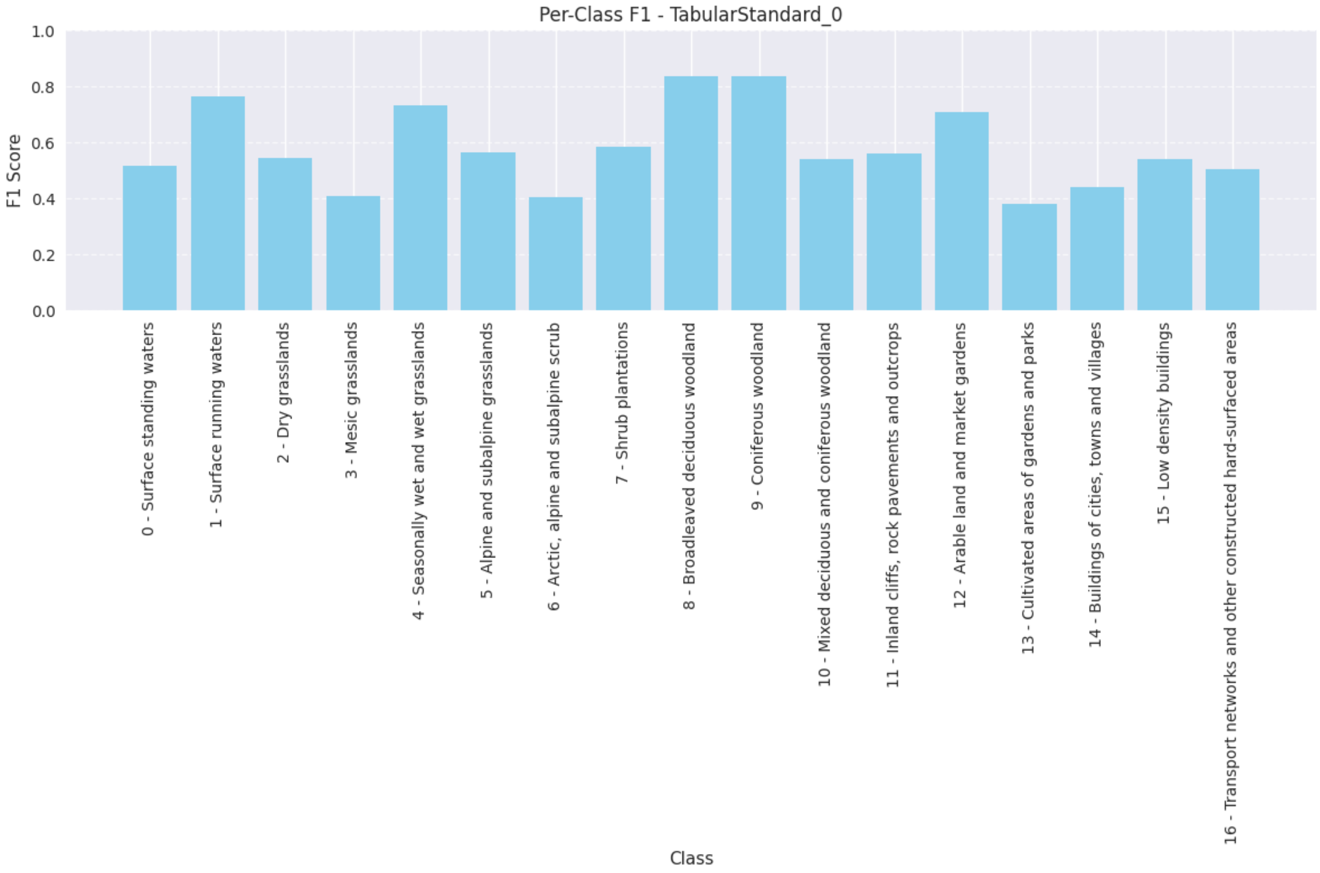


Figure 5: Per-class F1-scores of the TabularStandard model evaluated on the test set.

The confusion matrix for the *TabularStandard* model (Figure 6) confirms these observations. Most misclassifications occur between ecologically related classes, such as different grassland types or between mixed land-use categories. These errors are structured and ecologically plausible, reflecting genuine overlap in environmental conditions.

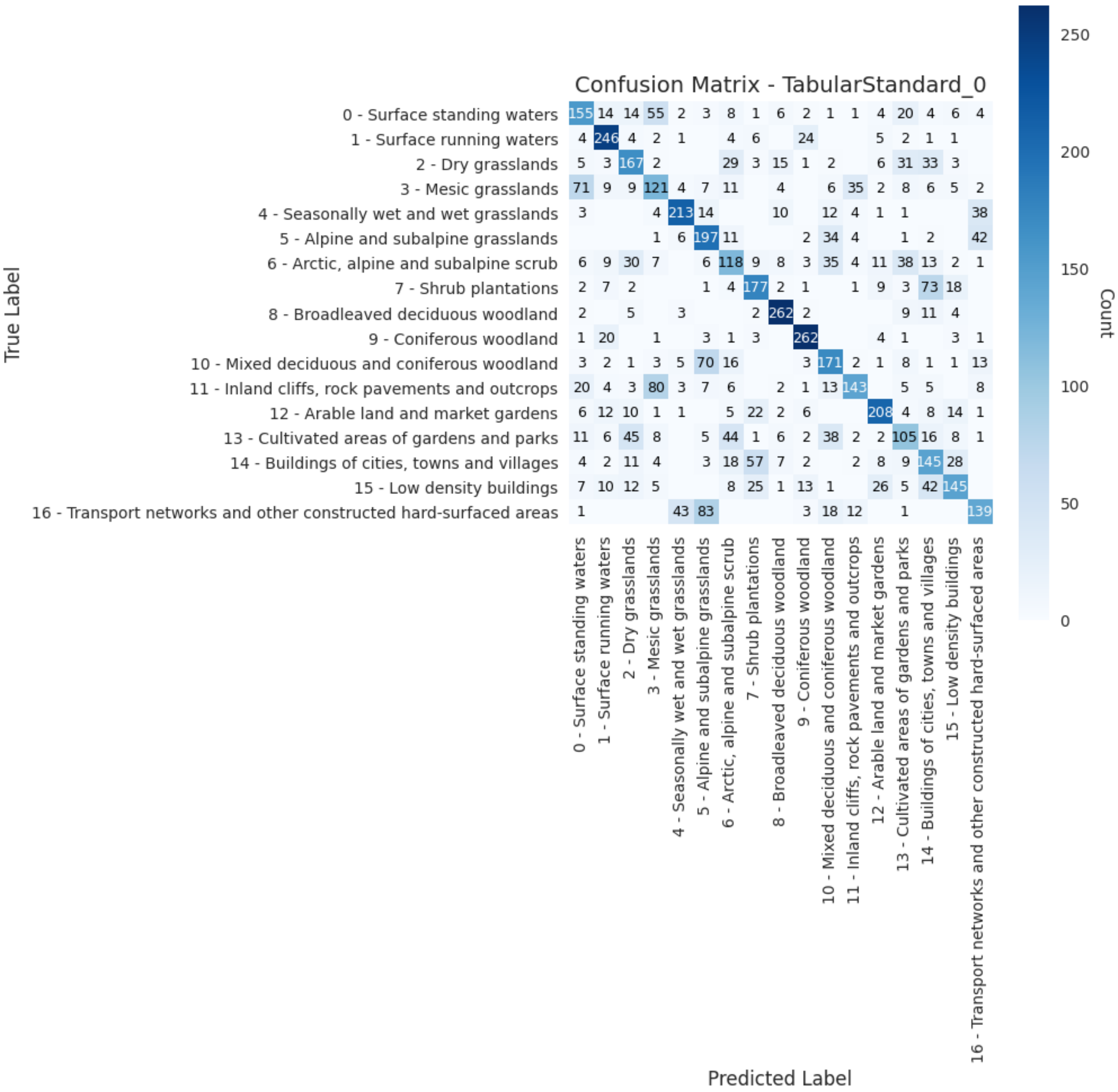


Figure 6: Confusion matrix of the tabular-only model evaluated on the test set.

3.5 Importance of Environmental Variables

Permutation importance aggregated by environmental variable group is shown in Figure 7. Vegetation-related variables are the most influential, followed by land use and land cover indicators. Bioclimatic and edaphic variables also contribute substantially, highlighting the importance of climate and soil conditions.

At the individual feature level, a small number of land-cover variables dominate the ranking, particularly those related to grasslands and meadows. This indicates that ecosystem prediction in the current setup relies strongly on a limited subset of highly informative environmental variables.

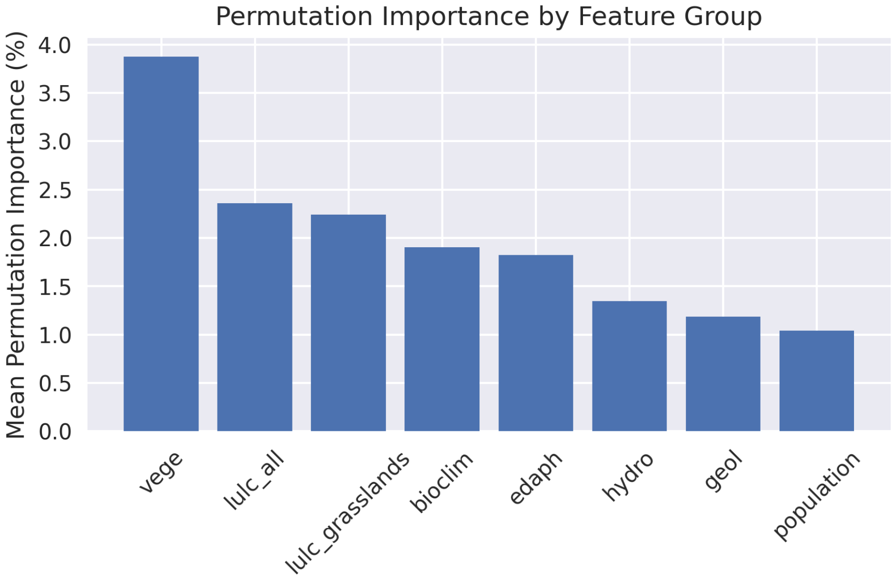
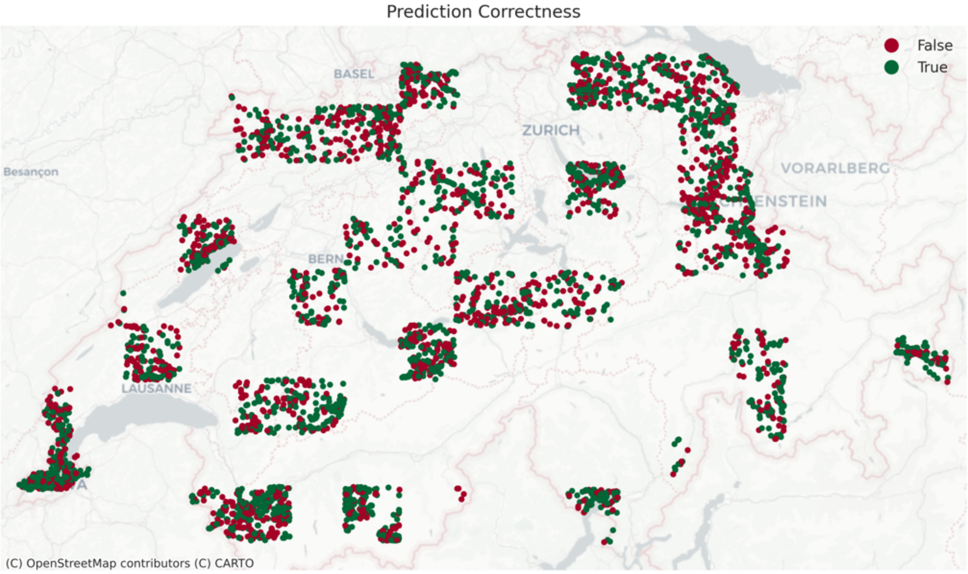


Figure 7: Mean permutation importance by environmental variable group for the tabular model.

3.6 Qualitative Inference and Spatial Patterns

Figure 8 shows the spatial distribution of prediction correctness for the *TabularStandard* model. Correct predictions are well distributed across Switzerland, with no strong spatial clustering of errors.

Misclassifications tend to occur in regions with heterogeneous land use or complex ecological transitions, consistent with the lower per-class performance observed for these ecosystems. This spatial analysis confirms that most errors are driven by local ecological complexity rather than by geographic bias.



(C) OpenStreetMap contributors (C) CARTO

Figure 8: Spatial distribution of prediction correctness for the tabular model on the test set.

3.7 Limitations and Perspectives

This study has several limitations. RGB imagery provides limited spectral information, which constrains image-based models. Some ecosystem classes are underrepresented, affecting class-level performance, and spatial context beyond individual image patches is not explicitly modeled.

Despite these limitations, the results show that tabular environmental variables provide strong predictive power for ecosystem mapping. Future work could explore multi-spectral or temporal imagery, larger image-based training datasets, and explicit spatial context modeling to further improve multimodal approaches.

4 Conclusion

This project investigated the use of deep learning for large-scale ecosystem mapping in Switzerland using environmental variables and high-resolution aerial imagery. Based on a dataset of more than 16,000 georeferenced samples labeled according to the EUNIS ecosystem classification, several model architectures were designed, trained, and evaluated, including tabular-only, image-only, and combined tabular–image approaches.

The results demonstrate that environmental variables from the SWECO25 database provide strong predictive power on their own. The tabular-only model achieved the best overall performance on both validation and test sets, showing stable training behavior, good generalization, and very low inference cost. In contrast, image-only models based on RGB aerial imagery achieved lower performance and exhibited clear signs of overfitting, highlighting the limitations of visual information alone at this spatial resolution. Combined tabular–image models achieved competitive results but did not surpass the tabular-only baseline under the current experimental setup.

Analysis of per-class performance and confusion matrices revealed that ecosystems with clear environmental or land-use characteristics are easier to classify, while heterogeneous and transitional ecosystems, such as grasslands and mixed land-use areas, remain more challenging. Importantly, all models tend to struggle on the same ecosystem classes, suggesting that these difficulties are driven by intrinsic ecological similarities rather than model-specific limitations. Permutation importance analysis further showed that vegetation and land-use variables dominate the predictive signal, followed by climatic and soil-related factors, explaining the strong performance of tabular models.

Spatial visualization of prediction correctness confirmed that classification errors are evenly distributed across Switzerland, indicating the absence of location-specific biases and supporting the robustness of the train–validation–test split. Most errors are associated with local ecological complexity rather than large-scale geographic effects.

Despite some limitations, including the use of RGB imagery only and the absence of explicit spatial context modeling, this work demonstrates that deep learning models based on environmental variables can provide accurate, robust, and interpretable ecosystem predictions at the national scale. While multimodal approaches remain promising, future improvements will likely require richer spectral imagery, larger image-based training datasets, or explicit modeling of spatial and temporal context. The proposed framework offers a solid foundation for scalable ecosystem mapping and can support biodiversity monitoring, land-use planning, and ecological research.

References

[Chytrý et al., 2020] Chytrý, M., Hennekens, S. M., Jiménez-Alfaro, B., Knollová, I., Dengler, J., Jansen, F., Landucci, F., Schaminée, J. H. J., Açık, A., et al. (2020). EUNIS habitat classification: Expert system, characteristic species combinations and distribution maps of european habitats. *Applied Vegetation Science*, 23(4):648–675.

[Kulling et al., 2024] Kulling, N., Abegg, M., Bergamini, A., Bircher, S., Bolli, P., Butsic, V., Gessler, A., Ginzler, C., Graf, R., Guisan, A., Hobi, M. L., Hugentobler, A., Keller, D., Knutti, R., Lehmann, L., Lüscher, G., Moser, B., Nobis, M. P., Obrist, M. K., Pannatier, E., Pellissier, L., Pilla, F., Tinner, W., Verrecchia, E. P., Zappa, M., Zimmermann, N. E., and Karger, D. N. (2024). SWECO25: A cross-thematic raster database for ecological research in switzerland. *Scientific Data*, 11(1):21.

[swisstopo, 2020] swisstopo (2020). swissIMAGE orthophotos. `https://www.swisstopo.admin.ch/en/orthoimage-swissimage-10`. Accessed: 2025-02-10.