Model fusion for multimodal prediction of plant species composition

Warsaw University of Technology

ML in PL CONFERENCE 2024

Kinga Kwoka[1] Mateusz Zembroń[1] Kamil Deja [1,2]

[1]Warsaw University of Technology [2]IDEAS NCBR

#TLDR

- We develop a model to predict plant species presence across space and time.
- We employ a multi-modal neural network combining ViT and ResNet architectures, processing satellite images, climate time series, and environmental maps for species prediction.
- We use **Focal Loss** to focus learning on rare species, addressing the challenge of presence-only data and improving model robustness.

GeoLifeCLEF 2024 Competition

- Prediction of plant species composition across space and time.
- Learning from a small amount of high quality presence-absence multi-label data and a large number of presence-only single-label samples.

Multimodal data

Provided datasets included species distribution as well as geographic and environmental data.

(a) RGB Image (b) NIR Image

Figure 1. Satellite image patches for sample location

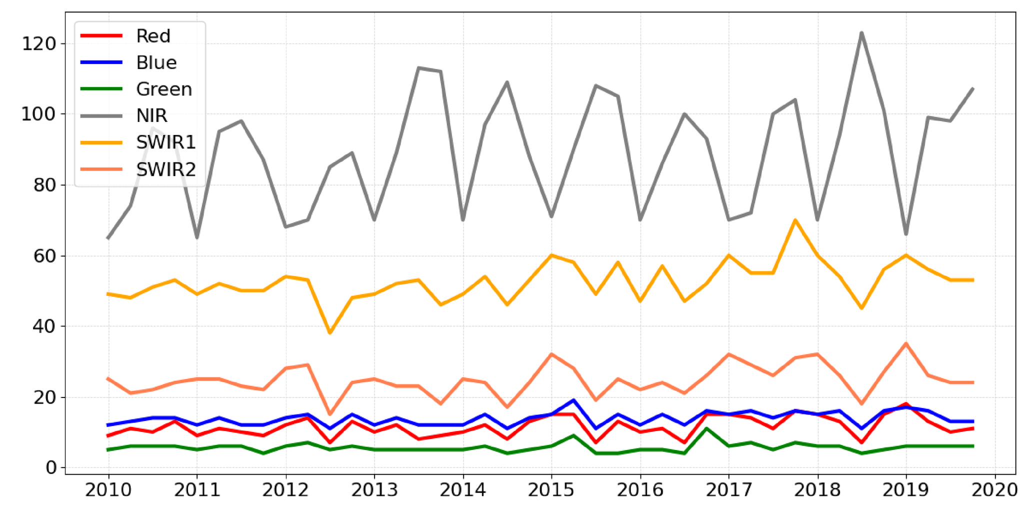


Figure 2. Satellite time series for sample location

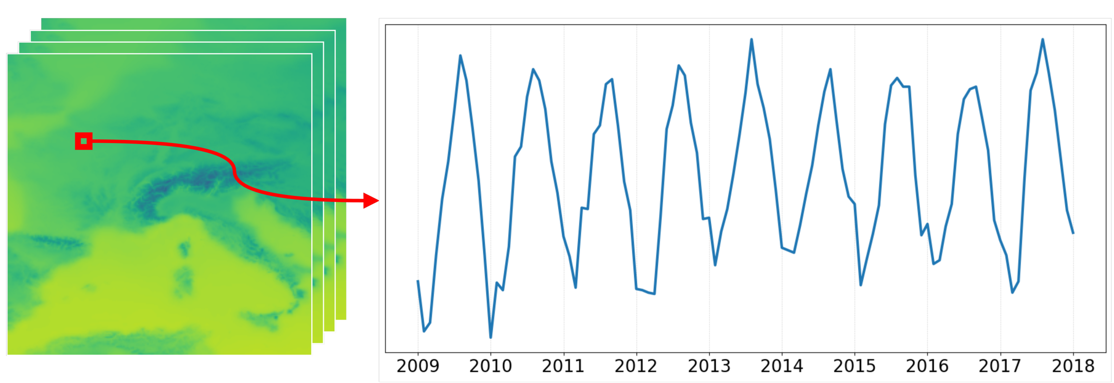


Figure 3. Mean daily temperature series for sample location

Top 25

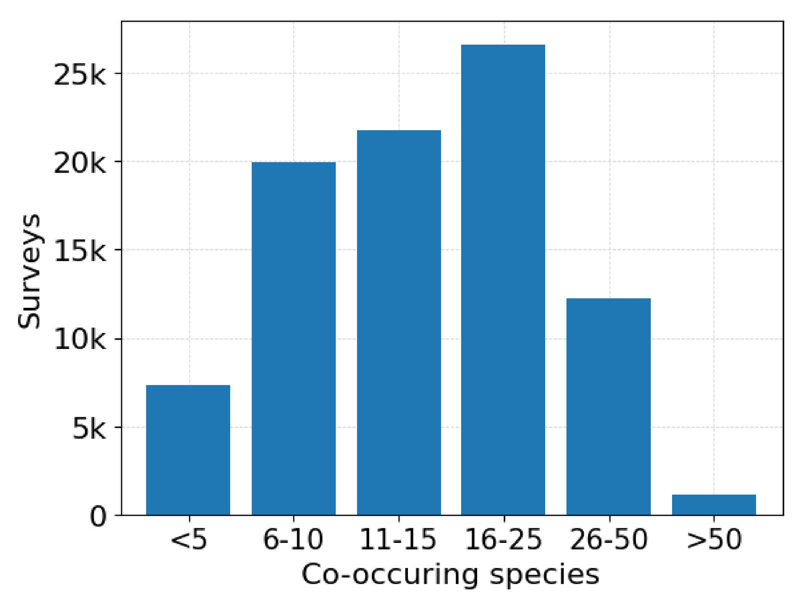
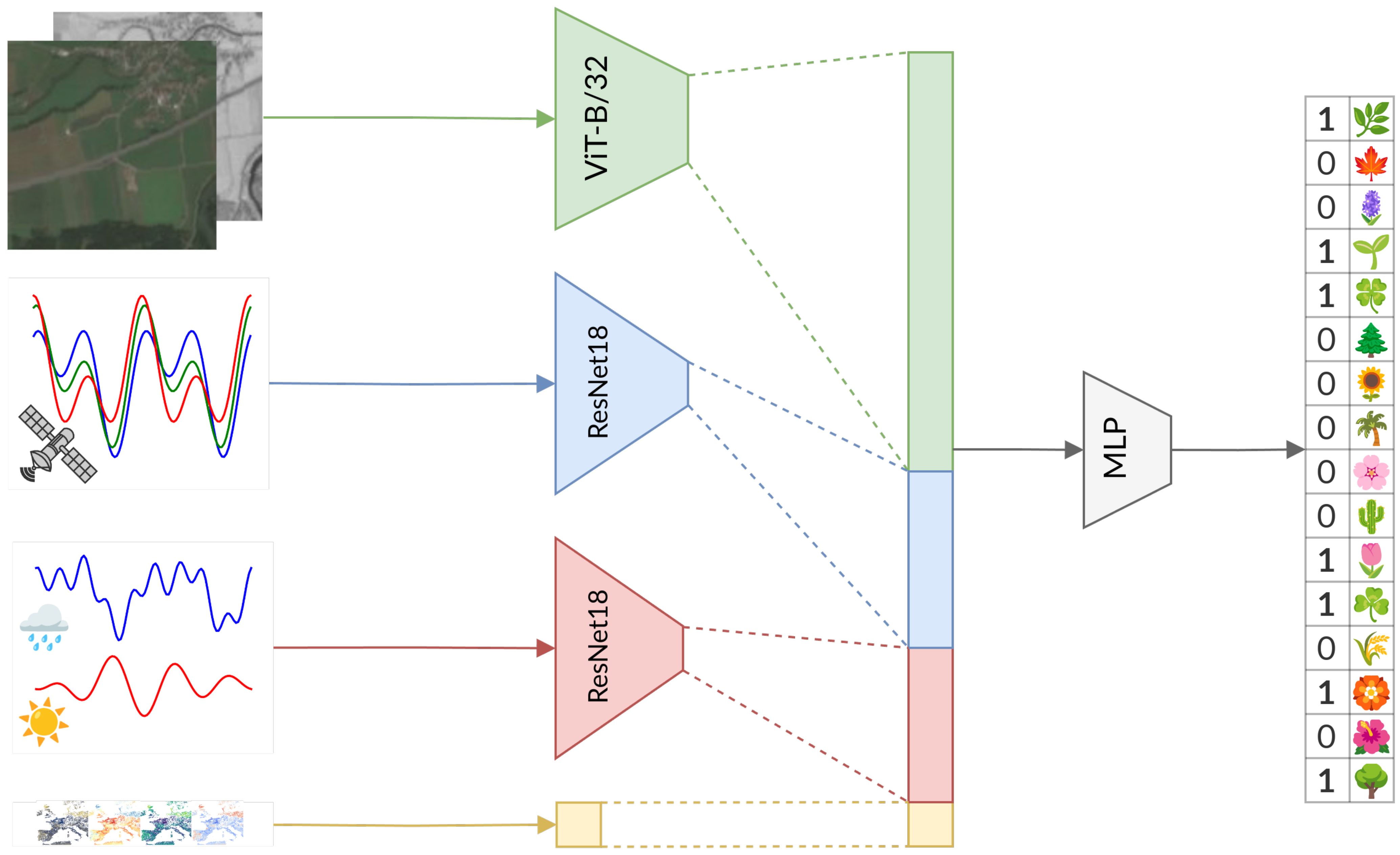


Figure 4. Species distribution across surveys

Multi-modal feature fusion - one to rule them all!

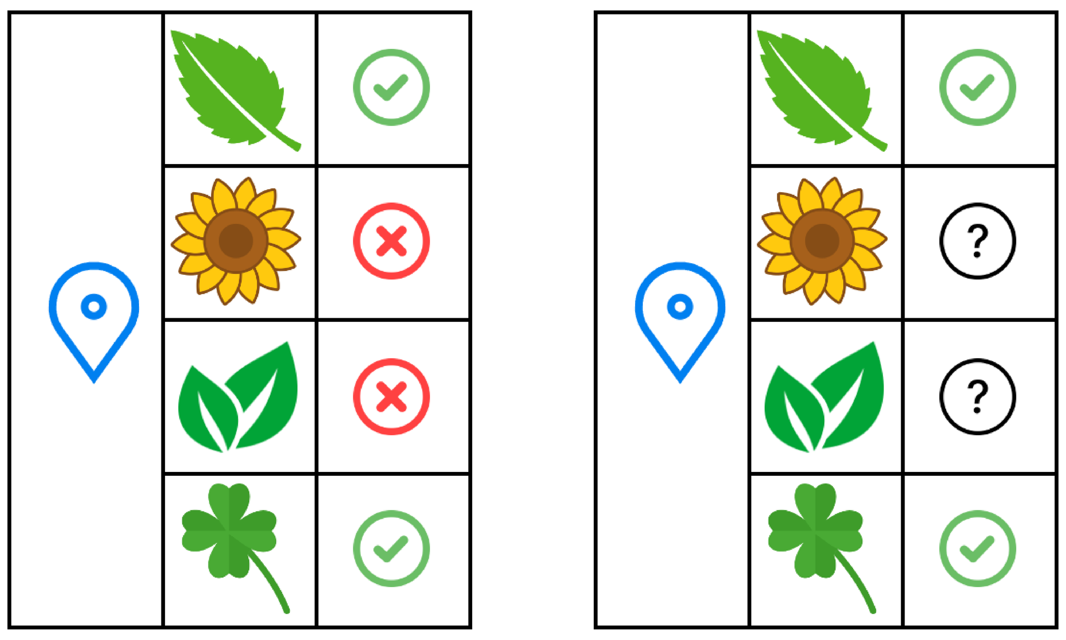
Each modality is processed by separate model:

- Visual Transformer processes RGB and NIR images of the area.
- ResNets extract essential information from lansat and bioclimatic time series.
- Standalone features are fed directly to the head classifier.



Flower, are you there?

Presence-Only (PO): species observations with *unknown* absences; *easier* to obtain.
Presence-Absence (PA): species observations and *confirmed* absences; *harder* to obtain.



(a) Presence Absence (b) Presence Only

Figure 5. Comparison of training datasets

Focus on the positives!

Significant label imbalance - on average:

333x

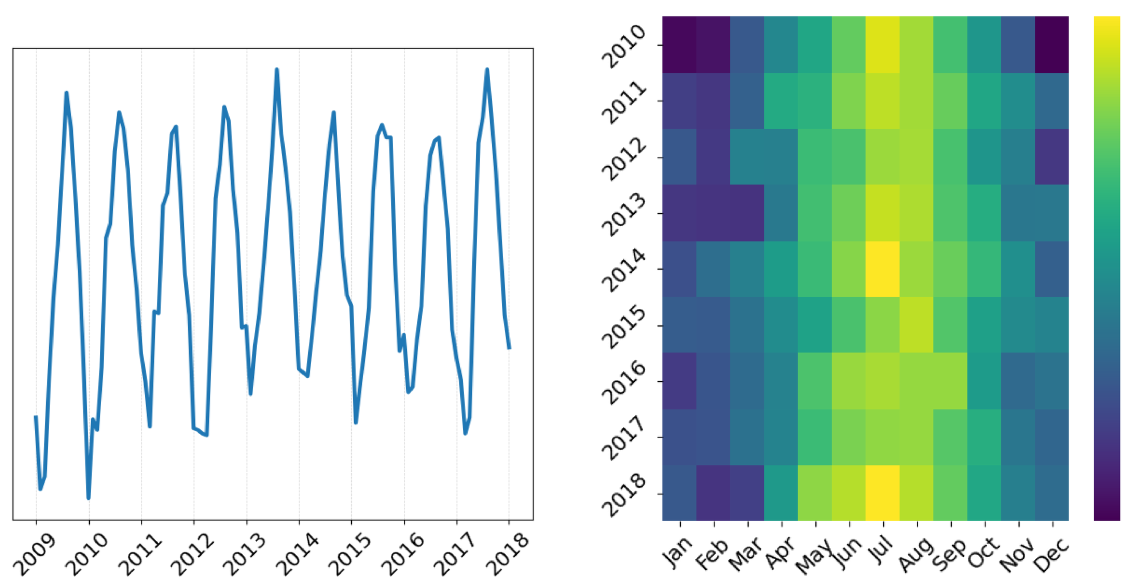
more negative than positive labels per sample location.

$$\mathrm{FL}(p_t) = -\alpha \underbrace{(1-p_t)^\gamma}_{\text{scaling factor}} \underbrace{log(p_t)}_{\text{CE loss}}$$

Focal Loss concentrates on rare, positive labels, thanks to focusing parameter $\gamma \geq 0$:

- misclassfied example ($p_t \longrightarrow 0$) - loss unaffected,
- correct example ($p_t \longrightarrow 1$) - loss downweighted.

Arranging time series into cubes



(a) Time series (b) Cube slice

Figure 6. Transformed mean daily temperature data

- Landsat and bioclimatic time series reshaped into cubes `(variables, quarters/months, years)`.
- Cubes to be treated as image-like tensors for use in ResNet.

Take-Away Points

The final model was verified in the GeoLifeCLEF 2024 CVPR competition, securing 13th place, indicating promising results.

- Combining specialized models significantly improves performance.
- Loss functions need to be well-adjusted to reflect specific challenges.
- Creative data preprocessing provides additional performance boost.

Learn more!

(a) GitHub repo (b) Contest website

This research was supported in part by PL-Grid Infrastructure grant nr PLG/2024/017176.

ML in PL Conference 2024

Contact: kinga.kwoka.stud@pw.edu.pl, mateusz.zembron.stud@pw.edu.pl