Early Classification for Agricultural Monitoring

from Satellite Time Series

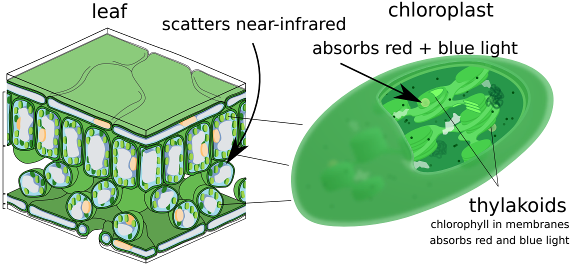
Marc Rußwurm, Romain Tavenard, Sébastien Lefèvre, Marco Körner

July 8, 2021

Photosynthesis: Light → Life

Vegetation

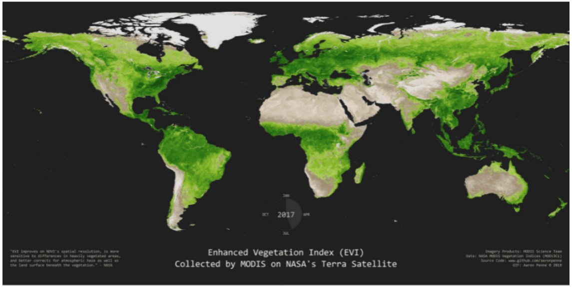
Spectral characteristics



common remote sensing feature: $\text{NDVI} = \frac{\text{NIR}-\text{RED}}{\text{NIR}+\text{RED}}$

images modified from wikipedia cc

Temporal characteristics



visualization Aaron Penne Github

Vegetation follows **seasonal life cycles** (phenology) which can be used to **distinguish categories**.

Why do we want to classify vegetation?

Distinguishing Vegetation

Food security Estimate the expected yield (from cultivated area) to predict food prices, shortages in countries that collect few agronomic statistics.

Precision Agriculture Identifying vegetation is a first step to provide personalized recommendations to farmers on machining practices and location-based fertilizer use.

Subsidy control In Europe: monitor and control crop subsidy payments for European farmers.

Why do we want to classify vegetation?

Distinguishing Vegetation

Food security Estimate the expected yield (from cultivated area) to predict food prices, shortages in countries that collect few agronomic statistics.

Precision Agriculture Identifying vegetation is a first step to provide personalized recommendations to farmers on machining practices and location-based fertilizer use.

Subsidy control In Europe: monitor and control crop subsidy payments for European farmers.

Why do we want to classify vegetation?

Distinguishing Vegetation

Food security Estimate the expected yield (from cultivated area) to predict food prices, shortages in countries that collect few agronomic statistics.

Precision Agriculture Identifying vegetation is a first step to provide personalized recommendations to farmers on machining practices and location-based fertilizer use.

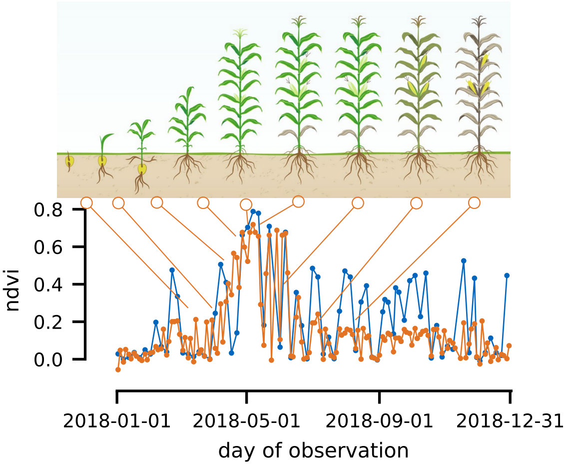
Subsidy control In Europe: monitor and control crop subsidy payments for European farmers.

Earth Observation

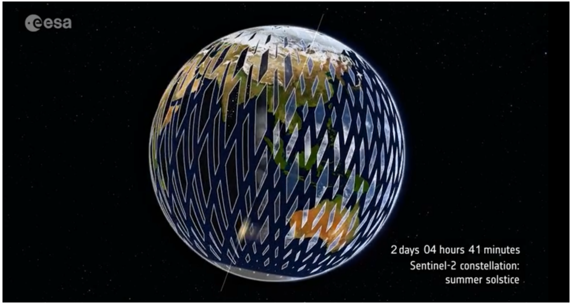
Earth Observation

Satellite Data

growth stages of corn Mimić et al., 2020



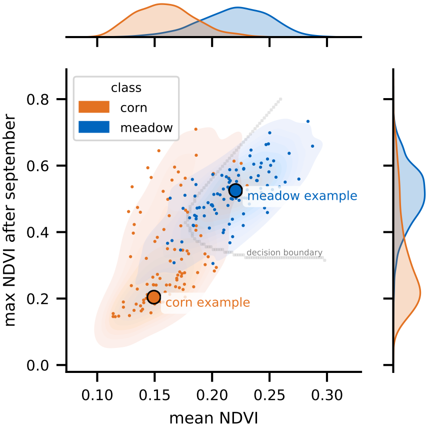
Sentinel 2. every 2-5 days at $10m \times 10m$ resolution.



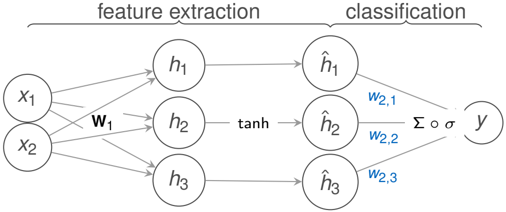
Source ESA

Classifying Vegetation

with hand-defined features

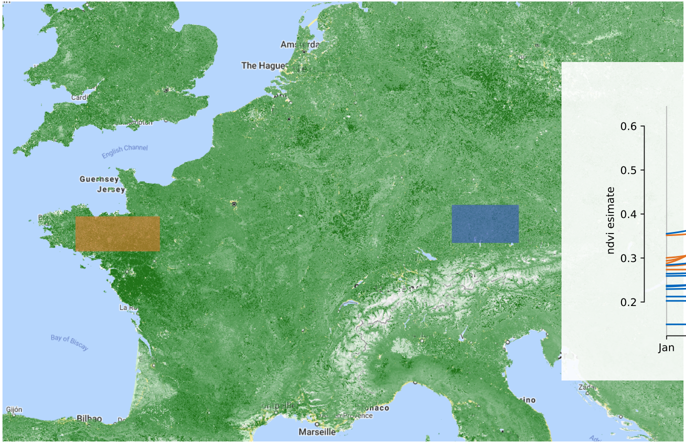


features learned from data

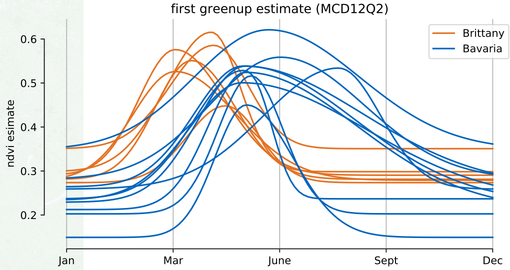


Choosing Observation Period

While we **learn** our **features** from **data**, we use **hand-chosen observation periods** to select the data.



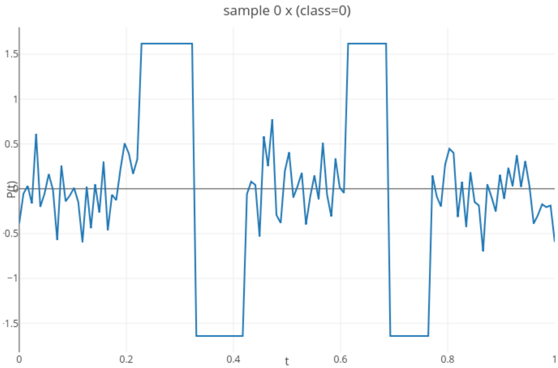
Day of first greenup (Modis product MCD12Q2)



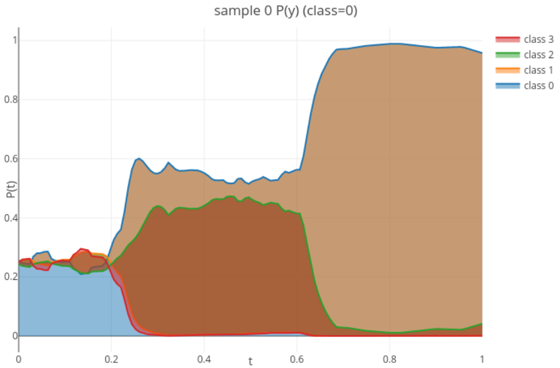
Early Classification

Class Predictions

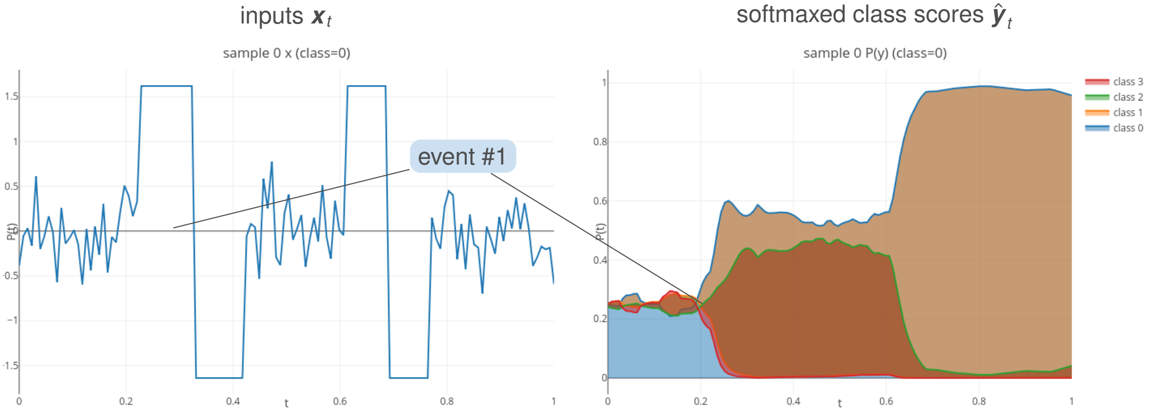
inputs $\boldsymbol{x}_t$



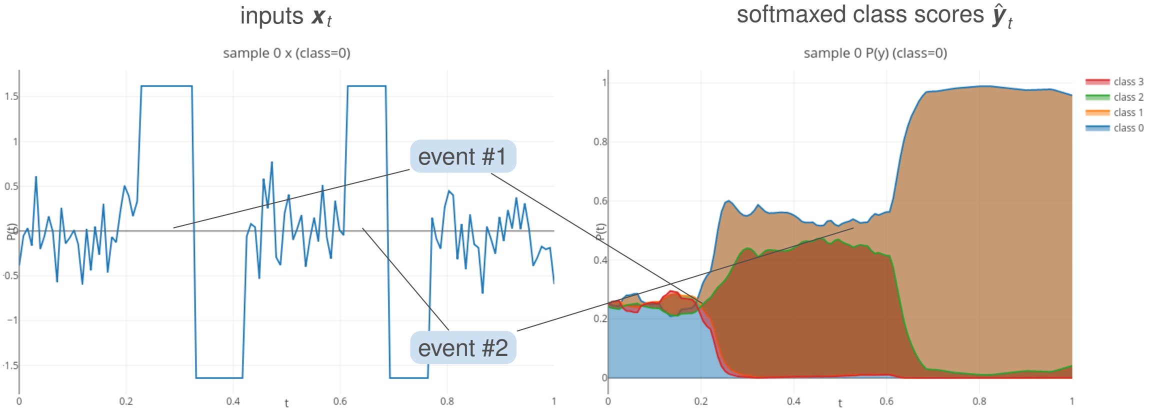
softmaxed class scores $\hat{\boldsymbol{y}}_t$



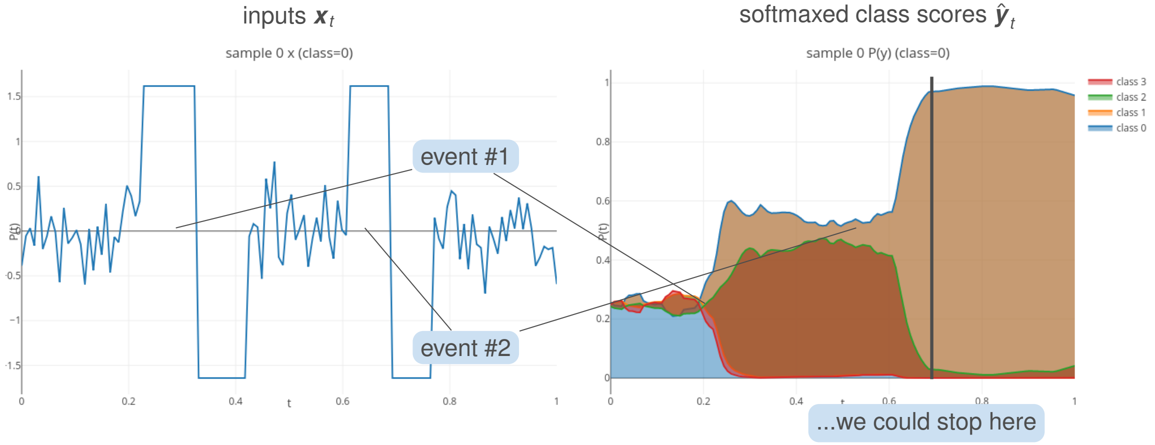
Class Predictions



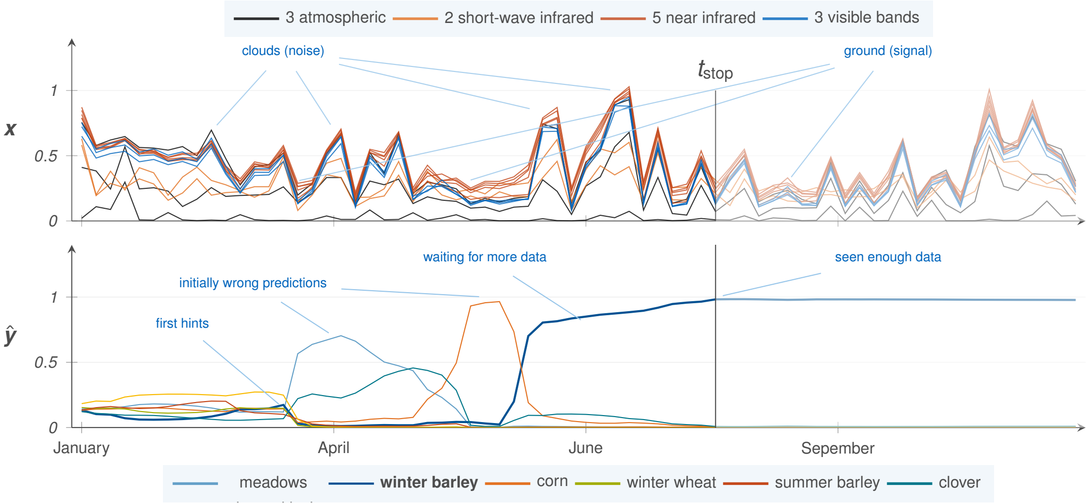
Class Predictions



Class Predictions

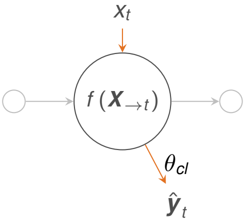


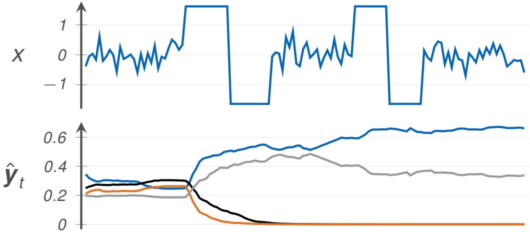
Early Classification on Remote Sensing Data



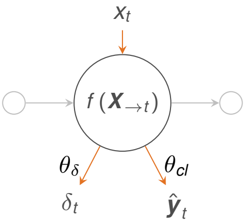
Method

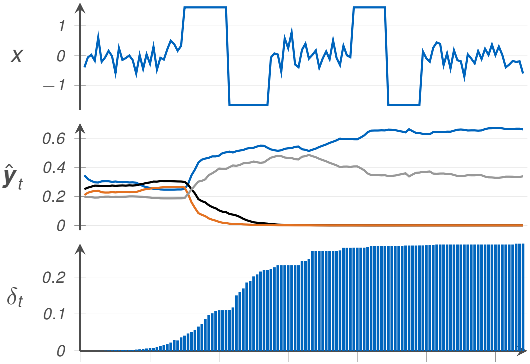
Augmenting Classification Models



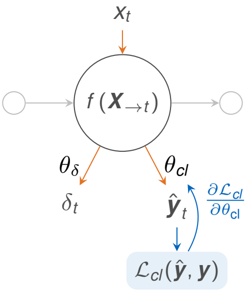


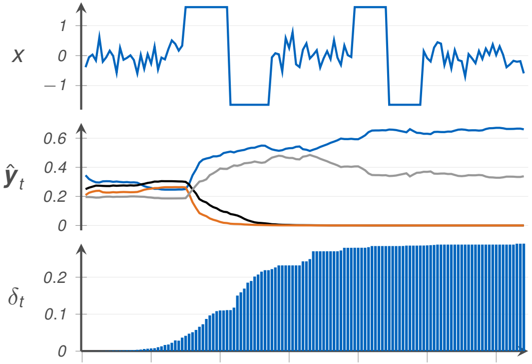
Augmenting Classification Models



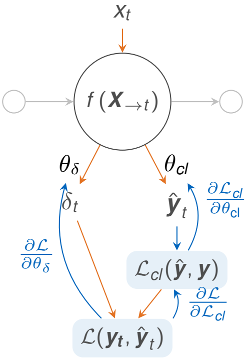


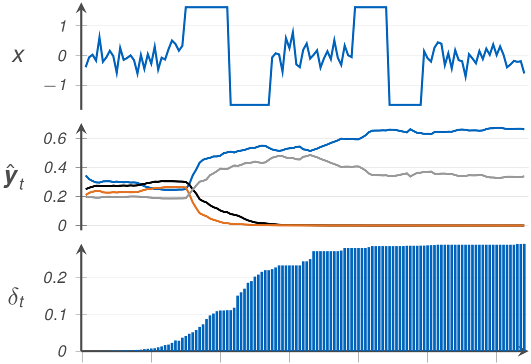
Augmenting Classification Models





Augmenting Classification Models





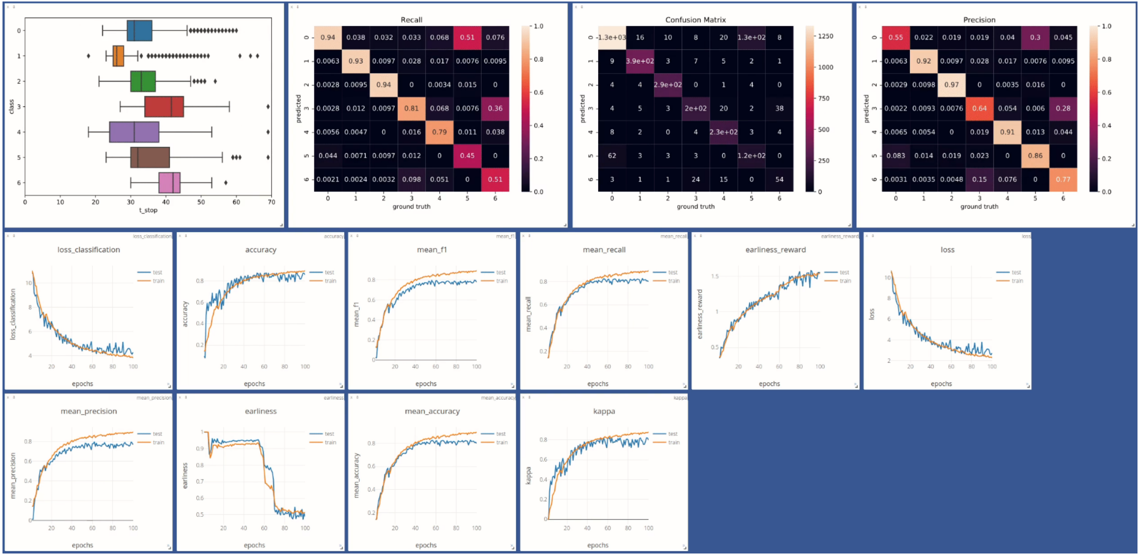
Loss functions

$$\mathcal{L}(\boldsymbol{X}, \boldsymbol{y}) = \sum_{t=1}^{T} P(t) \left(\mathcal{L}_c(\boldsymbol{X}_{\rightarrow t}, \boldsymbol{y}) - \mathcal{R}_e(t, \hat{y}_t^+) \right)$$

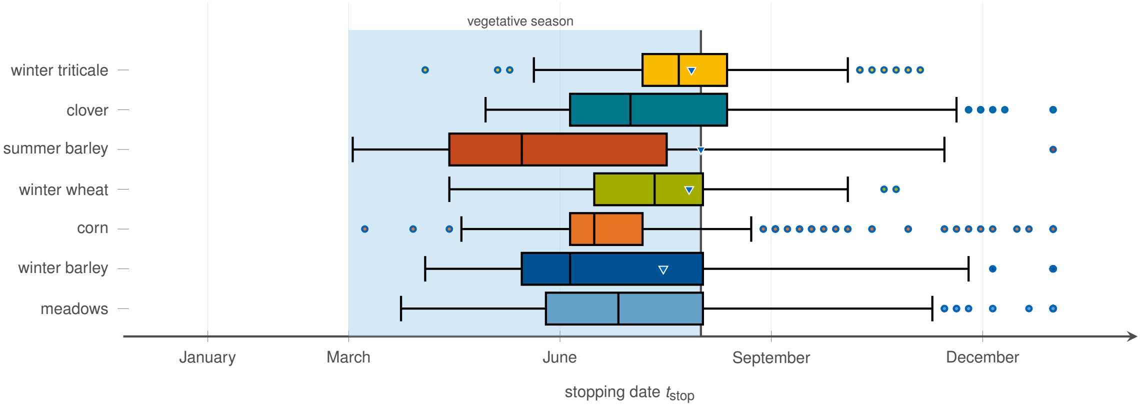
earlyness reward: $\mathcal{R}_e(t, \hat{y}_t^+) = \hat{y}_t^+ \left(1 - \frac{t}{T}\right)$ classificaiton loss: $-\log(\hat{y}_t^+)$ (aka. cross entropy)

- $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T)$: entire time series of observations $\boldsymbol{x}$
- $\boldsymbol{X}_{\rightarrow t} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_t)$: time series until time t
- $\boldsymbol{y} \in \{0, 1\}^C$: one-hot vector of the classes
- $\hat{y}^+ \in [0, 1]$: prediction score of the correct class

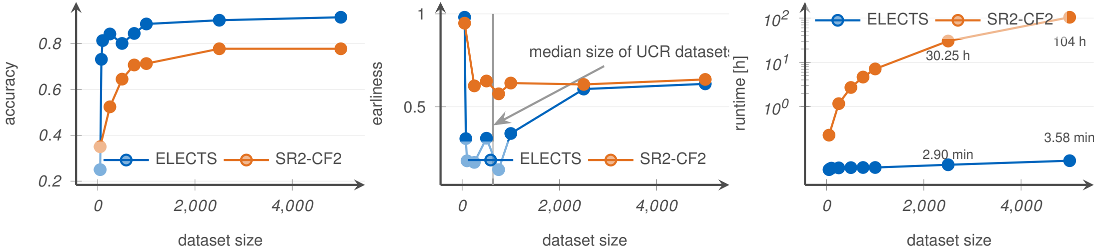
Results


Recall
Confusion Matrix
Precision
class
t_stop
predicted
ground truth
loss_classification
accuracy
mean_f1
mean_recall
earliness_reward
loss
mean_precision
earliness
mean_accuracy
kappa
test
train
epochs

Stopping times per crop Class



Comparison of ELECTS (ours) to SR2-CF2 (Mori et al.)



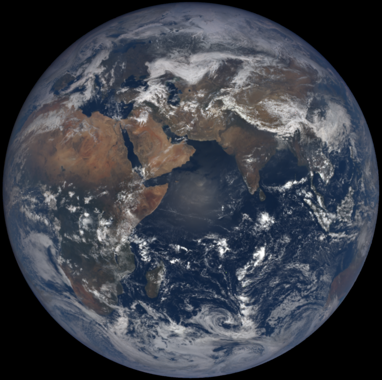
Evaluation of ELECTS (ours) and SR2-CF2 performance on class-balanced subsets of the Remote Sensing dataset.

Conclusion

Remote Sensing **Good** qualitative **results** on a **remote sensing problem** where **little work towards early classification** is done

Time Series Classification **Difficulties** on beating **state-of-the-art consistently** on **UCR datasets**

Runtime ELECTS more scalable (applicable) on large data: 4 minute runtime ELECTS versus 104 hours SR2-CF2



Thank you!

backup slides...

Crop Type Data

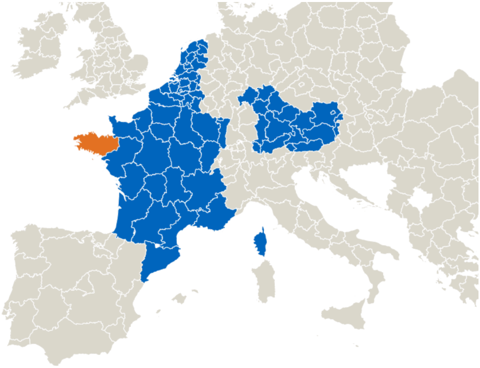
Building Large-Scale Crop Type Mapping Datasets

collected yearly within European **Common Agricultural Policy** (CAP)

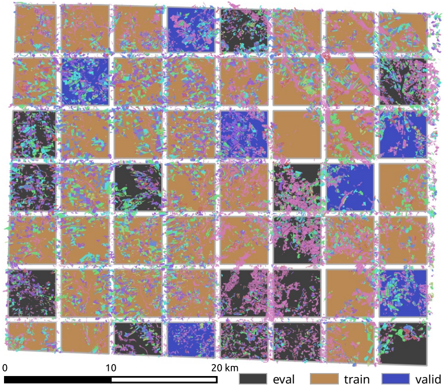
declared by Farmers at **crop subsidy** applications

today slowly made publicly available (on a national basis by French IGN, Bavarian Stmelf, etc.)

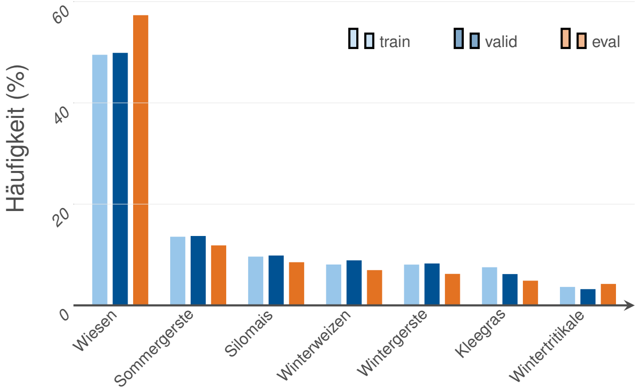
in future further harmonized within **Europe's INSPIRE** directive



Area of Interest for Early Classification



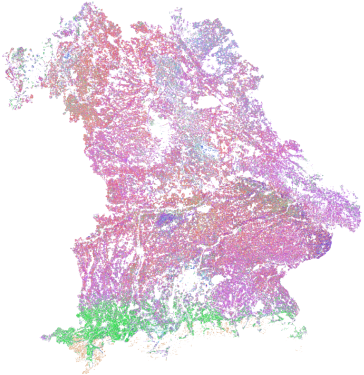
The area of interest and partitioning in blocks of 4.5 km for training validation and evaluation

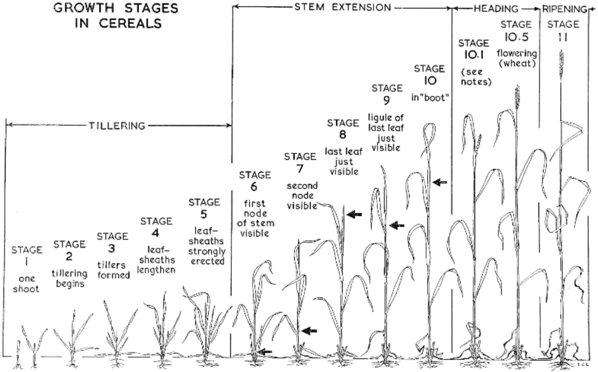


Class distribution in the dataset with block-wise separation of train, valid and evaluation partitions.

Here, we hand-selected 7 classes from a small region of interest...

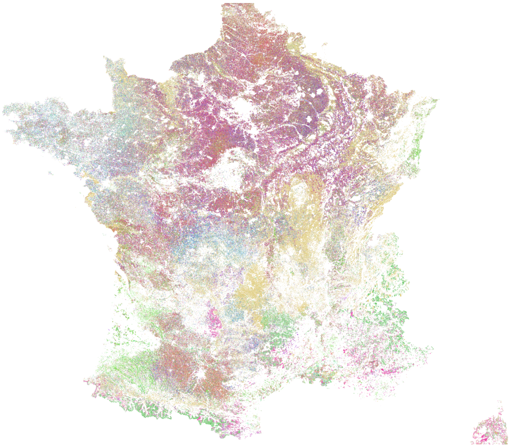
Large Scale Regional Variations





Large, E. C. (1954). Growth stages in cereals illustration of the Feekes scale. Plant pathology, 3(4), 128-129.

Large Scale Regional Variations



Questions:

- how do we learn models on these inter-regional scales?
- the same class label will correspond to different representations in the data.

Outlook

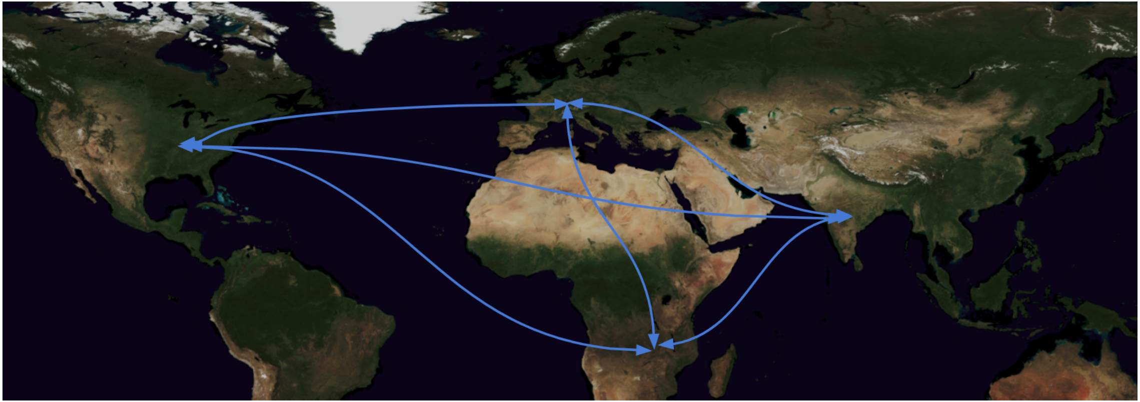
Goals:

- large-scale domain adaptation between regions of multi-temporal vegetation data
- addressing long-tailed class distributions of $>$300 distinct (overlapping) categories with 90% of data in $<$20 classes

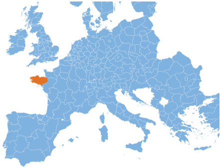
Short-Term Objective:

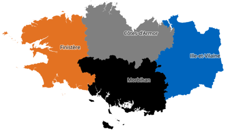
- compile a large-scale inter-regional crop type mapping dataset to be able to evaluate these questions

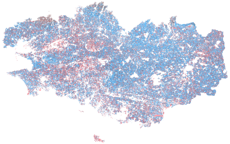
The Goal

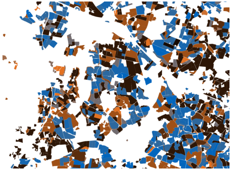


BreizhCrops Dataset (ICML Workshop Submission)

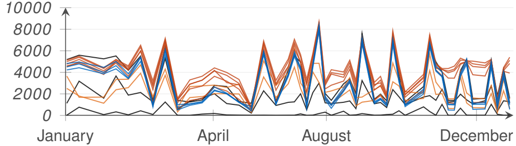




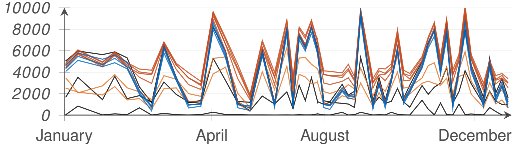




corn grain and silage



temporary meadows



580k samples of Time Series ***X*** and labels ***y***.

`https://github.com/TUM-LMF/BreizhCrops`

Rußwurm M., Lefèvre S., and Körner M (2019). ***BreizhCrops: A Satellite Time Series Dataset for Crop Type Identification****. ICML 2019 Time Series Workshop. arXiv:1905.11893*

Two Baseline Models

Inspired by Models used in NLP, we implemented a **multi-layer LSTM** and a **(minified) Transformer encoder**.

baseline	accuracy	κ	mean f1	mean precision	mean recall
Transformer (Vaswani et al., 2017)	**0.69**	**0.63**	0.57	0.60	0.56
LSTM (Hochreiter and Schmidhuber, 1997)	0.68	0.62	**0.59**	**0.63**	**0.58**

Takeaway:

- models perform quite similar
- potential for improvement
- well-defined classes accurately classified
- broadly defined classes systematically confused

Challenges and Impact

Impact

- large scale **real-world dataset**
- effectively **unlimited data** (spatially and temporally)
- **assessing generalization** over large regions
- extraction for further **vegetation characteristics** in future work (drought indicator, early classification, crop yield)

Challenges and Impact

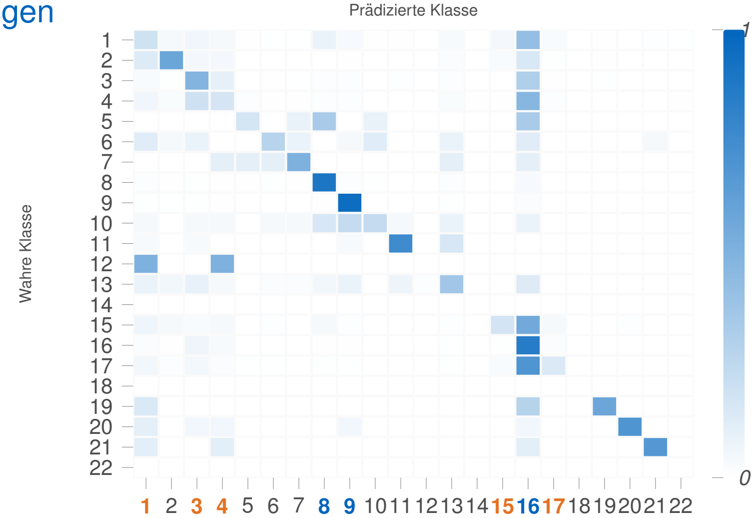
Impact

- large scale **real-world dataset**
- effectively **unlimited data** (spatially and temporally)
- **assessing generalization** over large regions
- extraction for further **vegetation characteristics** in future work (drought indicator, early classification, crop yield)

Challenges

- **imbalanced** class **labels** (¿300 raw classes)
- classes with **similar characteristics**
- non-Gaussian noise induced by **clouds**
- **regional variations** in the class distributions
- **spatial autocorrelation**
- **irregular** temporal **sampling** distance
- **variable sequence** length

Verwechslungen



83 Finale Kategorien

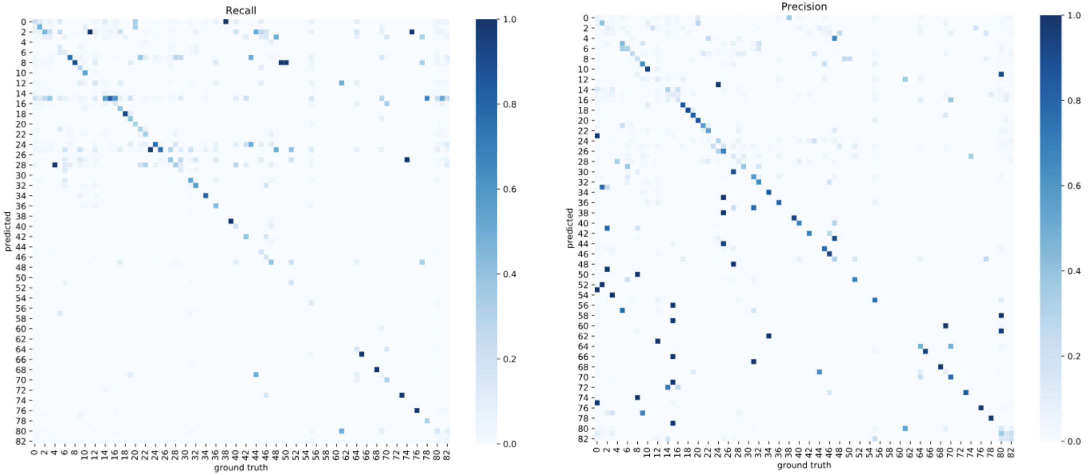


Table: Varying the weighting factor α for the two loss formulations *early reward* (**??**) and *lateness penalty* (**??**).

α	accuracy	$\bar{t}_{\text{stop}}$	precision	recall	f_1	κ
.0	.24 ± .29	.00 ± .00	.04 ± .04	.15 ± .00	.05 ± .05	.00 ± .00
.2	.15 ± .08	.00 ± .00	.10 ± .05	.15 ± .01	.08 ± .03	.01 ± .01
.4	.79 ± .05	.39 ± .03	.67 ± .05	.71 ± .04	.68 ± .04	.69 ± .06
.6	.83 ± .03	.59 ± .17	.71 ± .03	.74 ± .02	.72 ± .02	.74 ± .04
.8	.86 ± .01	.86 ± .12	.74 ± .01	.76 ± .02	.75 ± .02	.79 ± .01
1.0	.86 ± .01	1.00 ± .00	.75 ± .02	.76 ± .02	.75 ± .02	.79 ± .02

(a) two-phase *lateness penalty* loss formulation

α	accuracy	$\bar{t}_{\text{stop}}$	precision	recall	f_1	κ
.0	.25 ± .22	.10 ± .17	.19 ± .20	.25 ± .17	.16 ± .20	.12 ± .19
.2	.81 ± .03	.40 ± .02	.70 ± .01	.74 ± .01	.71 ± .01	.71 ± .04
.4	.80 ± .09	.47 ± .03	.71 ± .02	.74 ± .01	.71 ± .02	.71 ± .10
.6	.85 ± .02	.88 ± .07	.73 ± .04	.74 ± .03	.73 ± .03	.77 ± .03
.8	.84 ± .01	.93 ± .05	.72 ± .02	.75 ± .01	.73 ± .02	.76 ± .02
1.0	.83 ± .03	1.00 ± .00	.72 ± .03	.75 ± .01	.72 ± .03	.75 ± .04

(b) one phase *early-reward* formulation

Table: Quantitative analysis of the ε parameter on different trade-off factors between α earliness and accuracy. The illustrated figures show the mean and standard deviation of three runs with same parameters, but different initial random initialization.

ε	accuracy	$\bar{t}_{stop}$	f_1	precision	recall	κ
0	.10 ± .02	.02 ± .00	.07 ± .01	.13 ± .06	.17 ± .00	.02 ± .00
1	.75 ± .09	.44 ± .06	.65 ± .05	.64 ± .03	.69 ± .03	.64 ± .10
10	.81 ± .03	.40 ± .02	.71 ± .01	.70 ± .01	.74 ± .01	.71 ± .04

(a) $\alpha = 0.2$

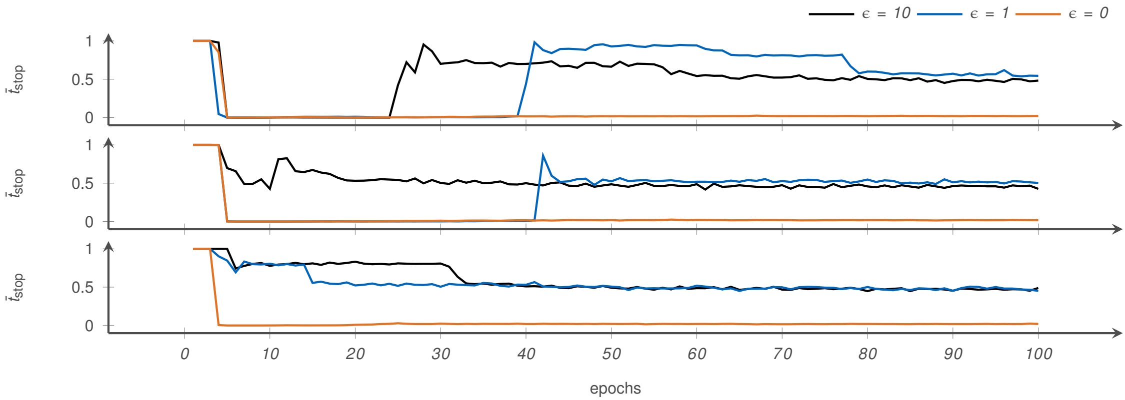
ε	accuracy	$\bar{t}_{stop}$	f_1	precision	recall	κ
0	.21 ± .20	.02 ± .00	.09 ± .04	.18 ± .03	.16 ± .01	.04 ± .03
1	.80 ± .02	.50 ± .05	.68 ± .05	.67 ± .05	.70 ± .07	.70 ± .03
10	.80 ± .09	.47 ± .03	.71 ± .02	.71 ± .02	.74 ± .01	.71 ± .10

(b) $\alpha = 0.4$

ε	accuracy	$\bar{t}_{stop}$	f_1	precision	recall	κ
0	.13 ± .04	.02 ± .00	.08 ± .01	.16 ± .01	.16 ± .01	.02 ± .01
1	.80 ± .05	.85 ± .14	.71 ± .02	.70 ± .02	.74 ± .01	.70 ± .06
10	.85 ± .02	.88 ± .07	.73 ± .03	.73 ± .04	.74 ± .03	.77 ± .03

(c) $\alpha = 0.6$

ε	accuracy	$\bar{t}_{stop}$	f_1	precision	recall	κ



The three runs with different random initialization per ε offset and $\alpha = 0.4$ of Table 4b. The $\varepsilon > 0$ offset factor allows the models to recover from a too early classification, as is visible in the top two plots.