PREDICTING CROP YIELD WITH MACHINE LEARNING

AN EXTENSIVE ANALYSIS OF INPUT MODALITIES AND MODELS ON A FIELD AND SUB-FIELD LEVEL

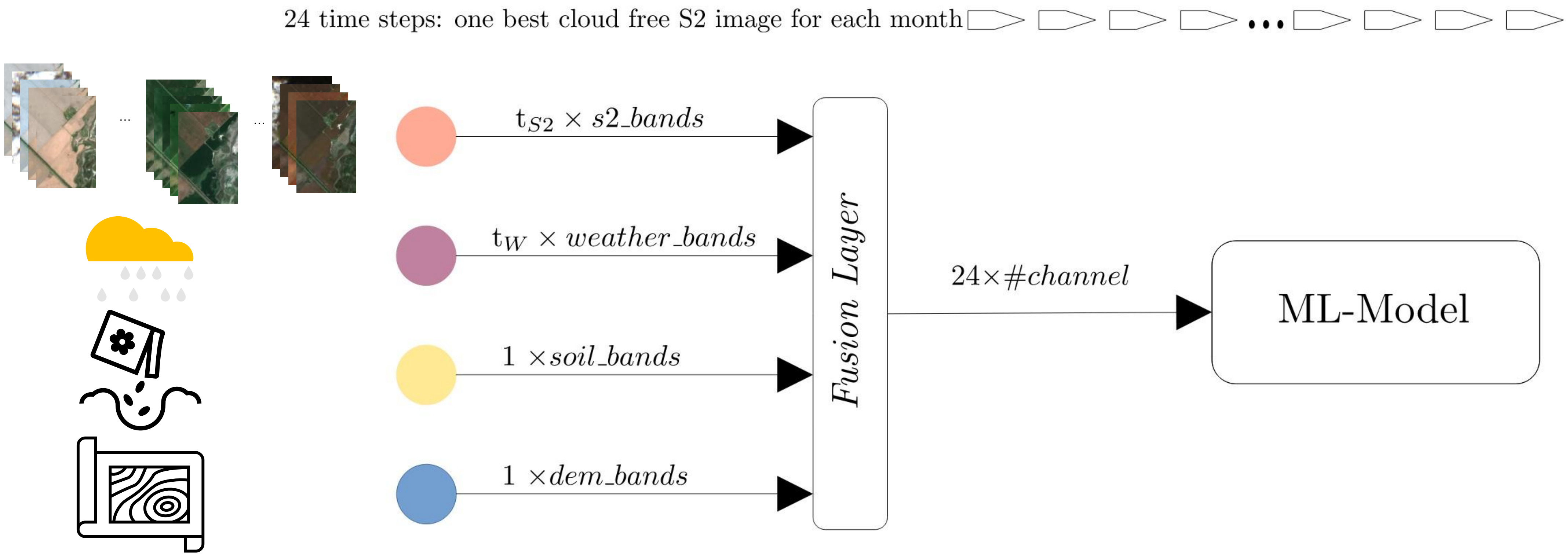
Deepak Pathak and **Miro Miranda**
German Research Center for Artificial Intelligence, SDS and University of Kaiserslautern-Landau (RPTU) ,
Kaiserslautern, Germany
Contact: {deepak_kumar.pathak, miro.miranda_lorenz} @dfki.de

Paper number: 4433

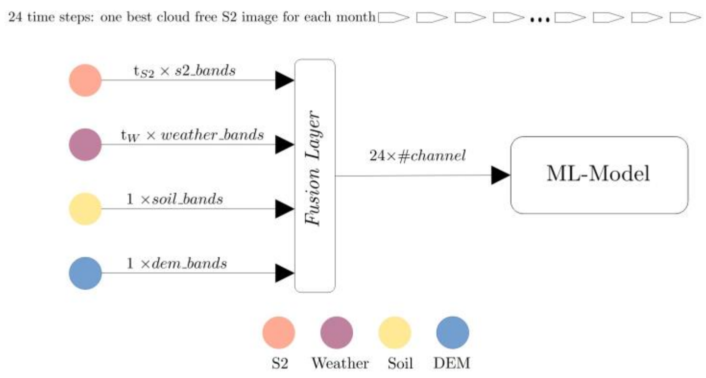
German Research Center for Artificial Intelligence

RPTU Rheinland-Pfälzische Technische Universität Kaiserslautern Landau

Introduction



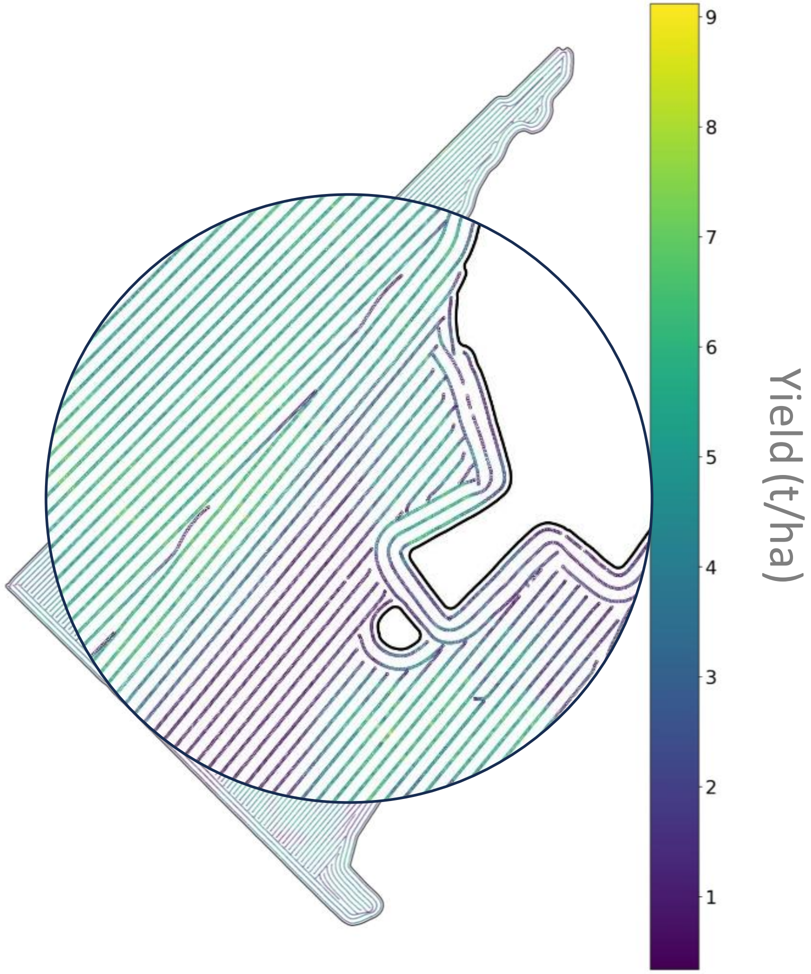
Introduction



Best Performing Combination of Modality

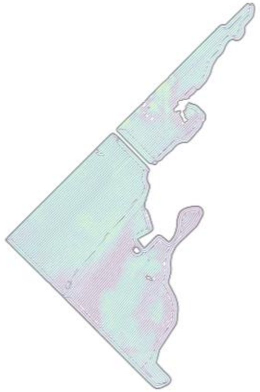
Evaluation				R2	
Model	**Modalities**	**Crop**	**Country**	**Field**	**Sub-Field**
LSTM	S2-DEM	Soybean	Argentina	0.82	0.65
LSTM	S2-Soil	Rapeseed	Germany	0.78	0.45
LGBM	S2-Weather-Soil-DEM	Soybean	Uruguay	0.77	0.42
LGBM	S2-Weather-Soil-DEM	Wheat	Germany	0.68	0.37

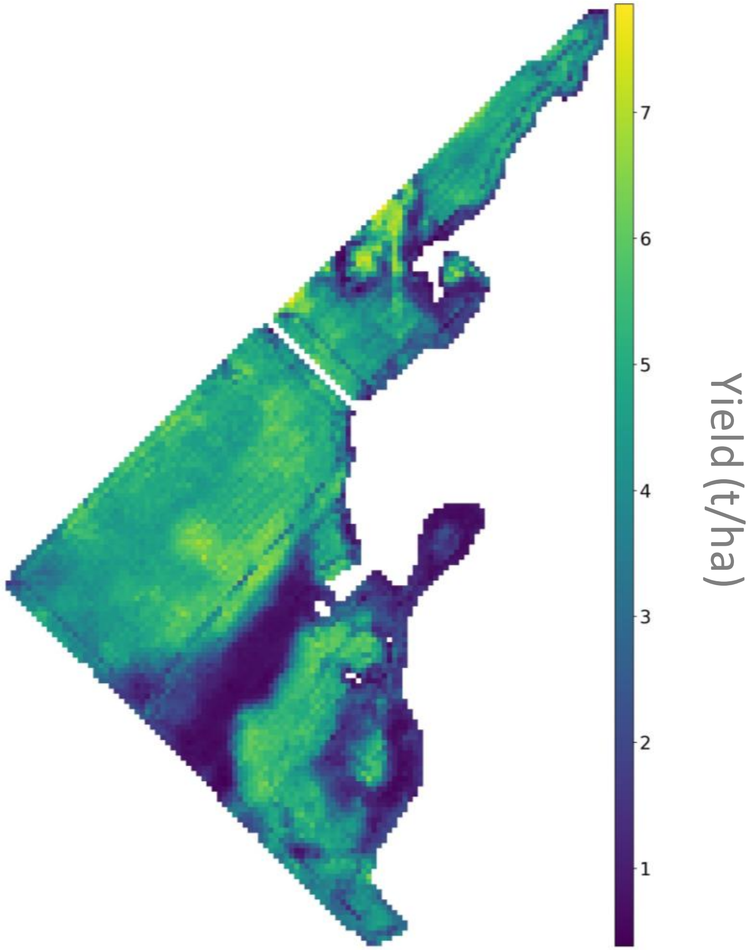
Ground Truth – Yield Map



- Yield values as point vector data.
- Point data resolution
  - along harvesting track: ~1-2 m
  - across harvesting track: ~10-15 m

Ground Truth – Yield Map





- Yield values as point vector data.
- Point data resolution
  - along harvesting track: ~1-2 m
  - across harvesting track: ~10-15 m
- Remove outliers
- Rasterize at 10 m pixel resolution

Input Modalities

- Sentinel-2 L2A Images
  - Spatial resolution: 10 m
  - 12 Spectral Bands: ["B01", "B02", "B03", "B04", "B05", "B06", "B07", "B08", "B8A", "B09", "B11", "B12"]
  - Create 24-time steps representing two full calendar years and select S2 image based on best cloud free S2 images.

Input Modalities

- Sentinel-2 L2A Images
- Weather Data – ECMWF Reanalysis (ERA5)
  - Temperature: minimum, maximum, mean of a day
  - *Total precipitation* in a day
  - Prepared at field level
  - Aggregated features as sum between each time-step, and concatenated along S2 features.

Input Modalities

- Sentinel-2 L2A Images
- Weather Data – ECMWF Reanalysis (ERA5)
- Digital Elevation Model (DEM) data - SRTM
  - Spatial resolution: 30 m → 10 m
  - DEM properties: *dem, slope, aspect, curvature, and topographic wetness index*
  - vectorized and concatenated along S2 features for each time step.

Input Modalities

- Sentinel-2 L2A Images
- Weather Data – ECMWF Reanalysis (ERA5)
- Digital Elevation Model (DEM) data – SRTM
- Soil data – SoilGrid 250m
  - Spatial resolution: 250 m → 10 m
  - Soil properties: *soil organic carbon, nitrogen, cation exchange capacity, pH, coarse fragment content, sand, silt, clay*
  - Depths: three depths from 0-30 cm
  - vectorized and concatenated along S2 features for each time step.

Input Modalities

- Sentinel-2 L2A Images
- Weather Data – ECMWF Reanalysis (ERA5)
- Digital Elevation Model (DEM) data – SRTM
- Soil data – SoilGrid 250m

S2 Bands + # weather features + # dem features + #soil features

$24 \times \#channels$

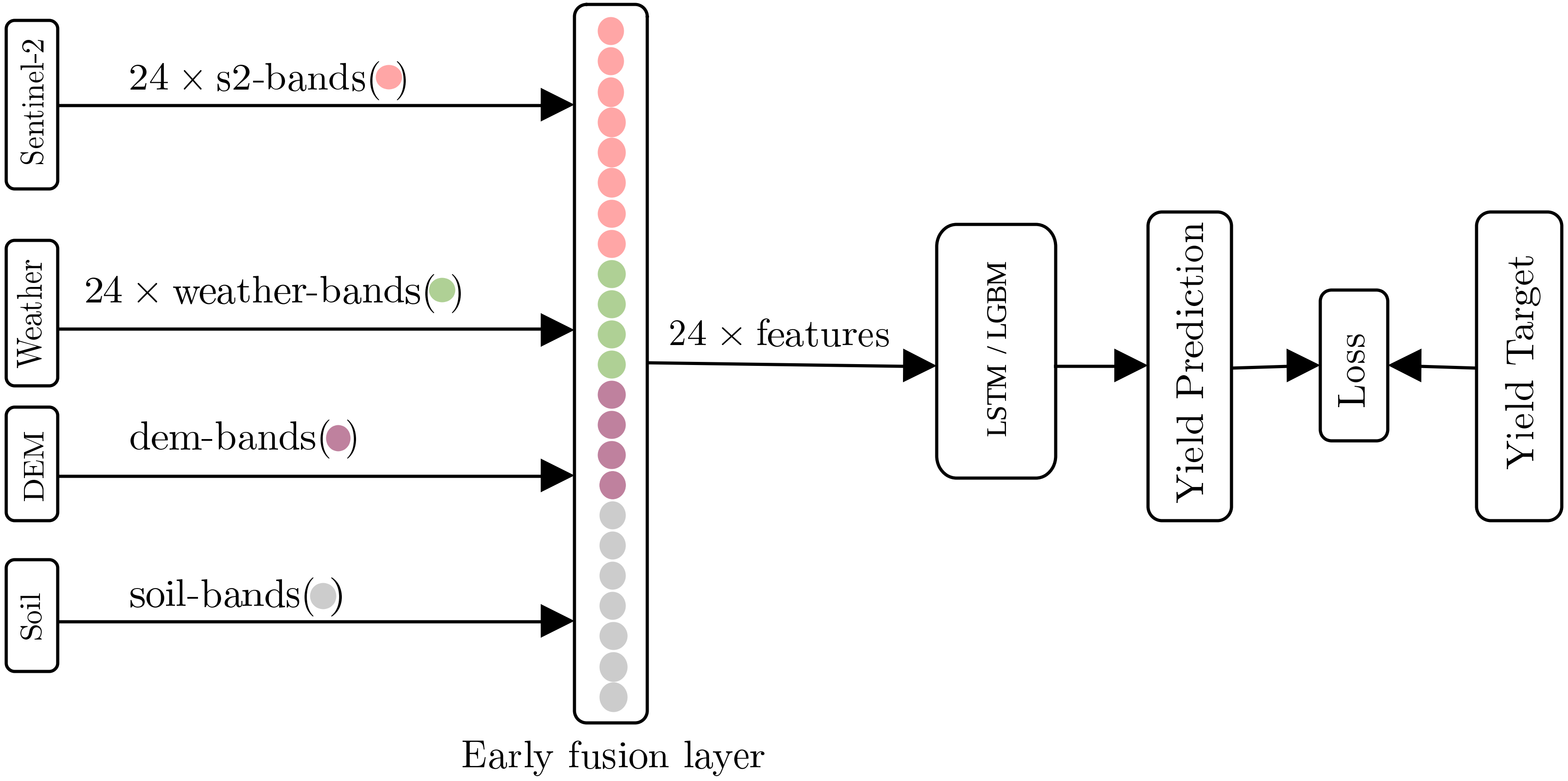
Time step 1
YYYY.01

Time step X
(seeding date)

Time step Y
(harvesting date)

Time step 24
YYYY+1.12

Method – Early Fusion



Dataset and Evaluation

Dataset

Country	Crop	# Fields
Germany	Rapeseed	111
Germany	Wheat	188
Uruguay	Soybean	486
Argentina	Soybean	192

Evaluation

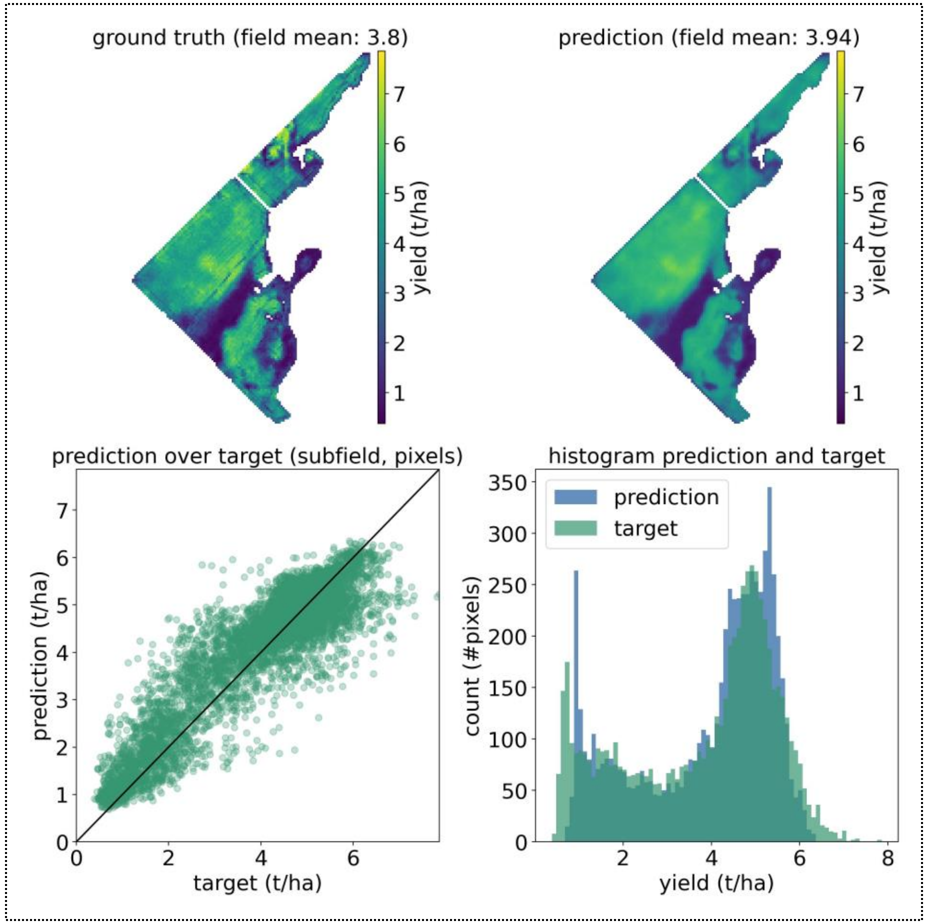
- Train separately for each dataset
- 10-fold cross validation
- Metrics:
  - Mean Absolute Percentage Error
  - R2 (coefficient of determination)
- Qualitative evaluation of predictions

Results

Comparing results with different combination of modalities for Soybean in Argentina using LSTM

Modalities	Field		Sub-field	
	MAPE	**R2**	**MAPE**	**R2**
S2-Weather-Soil-DEM	0.11	0.76	0.24	0.63
S2-DEM	0.09	0.82	0.24	0.65
S2-Soil	0.1	0.76	0.25	0.61
S2-Weather	0.11	0.78	0.25	0.63
S2	0.11	0.74	0.25	0.61

Visual Evaluation for an example field



Results

Best Performing Combination of Modality

Evaluation				Field		Sub-field	
Model	**Modalities**	**Crop**	**Country**	**MAPE**	**R2**	**MAPE**	**R2**
LSTM	S2-DEM	Soybean	Argentina	0.09	0.82	0.24	0.65
LSTM	S2-Soil	Rapeseed	Germany	0.15	0.78	0.39	0.45
LGBM	S2-Weather-Soil-DEM	Soybean	Uruguay	0.2	0.77	1.02	0.42
LGBM	S2-Weather-Soil-DEM	Wheat	Germany	0.09	0.68	0.29	0.37

Conclusion

- We showed good results for pixel-based yield prediction with machine learning.
- Early fusion can well capture infield yield variability.
- In addition to Sentinel-2, additional modalities help to improve the performance.
- Different regions and crops need different combination of input modalities.

PREDICTING CROP YIELD WITH MACHINE LEARNING

AN EXTENSIVE ANALYSIS OF INPUT MODALITIES AND MODELS ON A FIELD AND SUB-FIELD LEVEL

Deepak Pathak and **Miro Miranda**
German Research Center for Artificial Intelligence, SDS and
University of Kaiserslautern-Landau (RPTU) ,
Kaiserslautern, Germany
Contact: {deepak_kumar.pathak, miro.miranda_lorenz} @dfki.de

Paper number: 4433

German Research Center for Artificial Intelligence

RPTU Rheinland-Pfälzische Technische Universität Kaiserslautern Landau