

PREDICTING CROP YIELD WITH MACHINE LEARNING: AN EXTENSIVE ANALYSIS OF INPUT MODALITIES AND MODELS ON A FIELD AND SUB-FIELD LEVEL

Deepak Pathak ^{*,1,2}, Miro Miranda ^{*,1,2}, Francisco Mena ^{1,2}, Cristhian Sanchez ^{1,2}, Patrick Helber ³, Benjamin Bischke ³, Peter Habelitz ³, Hiba Najjar ^{1,2}, Jayanth Siddamsetty ², Diego Arenas ², Michaela Vollmer ², Marcela Charfuelan ², Marlon Nuske ², and Andreas Dengel ^{1,2}

¹University of Kaiserslautern-Landau (RPTU), Kaiserslautern, Germany

²German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany

³Vision Impulse GmbH, Kaiserslautern, Germany

ABSTRACT

We introduce a simple yet effective early fusion method for crop yield prediction that handles multiple input modalities with different temporal and spatial resolutions. We use high-resolution crop yield maps as ground truth data to train crop and machine learning model agnostic methods at the sub-field level. We use Sentinel-2 satellite imagery as the primary modality for input data with other complementary modalities, including weather, soil, and DEM data. The proposed method uses input modalities available with global coverage, making the framework globally scalable. We explicitly highlight the importance of input modalities for crop yield prediction and emphasize that the best-performing combination of input modalities depends on region, crop, and chosen model.

Index Terms— Sentinel-2, Multi-modal data, Early fusion, Precision Farming, Yield Maps

1. INTRODUCTION

Yield prediction is an essential task in the agricultural sector. Yet, it is still challenging due to multidimensional factors defining the yield, including environmental factors, management, the genotype, and their interactions. Providing accurate yield prediction not only supports industry and farmers in decision-making such as pest control, fertilization, and harvest time prediction, but also policymakers. In light of changing and fluctuating climate conditions, reliable yield predictions are increasingly challenging and are nowadays addressed from multiple perspectives. Here, machine learning has played an increasingly important role in recent years [1]. With the rise of remote sensing technology, yield prediction can be addressed from a large-scale perspective, as data is available globally with high temporal frequencies. This offers various opportunities for crop monitoring and particularly for

yield prediction [2, 3]. Currently, models are trained on diverse sets of remotely sensed input modalities such as satellite imagery, weather, soil, and Digital Elevation Model (DEM) data. [4, 3, 5]. Although it is known that all the mentioned modalities are good yield predictors, only a subset is included in most studies. It still needs to be determined if including multiple modalities is beneficial for model performance. In addition, most studies focus on a narrow regional level with single crop cultivars and few training years, making models highly susceptible to regional and temporal overfitting. It is, moreover, still an open question if machine learning can predict crop yields consistently over years, regions, and crop types.

In this research, we present an operational approach to multimodal yield prediction at the pixel level in 10m resolution, referred to as sub-field, that is crop and region independent and globally scalable. We analyze the importance of input modalities, such as satellite imagery, and additional modalities, including weather, soil, and DEM data. A simple but effective way of data fusion is proposed to combine data with different temporal and spatial resolutions. Results are evaluated on a large dataset containing different countries, crops, and years at field and sub-field level.

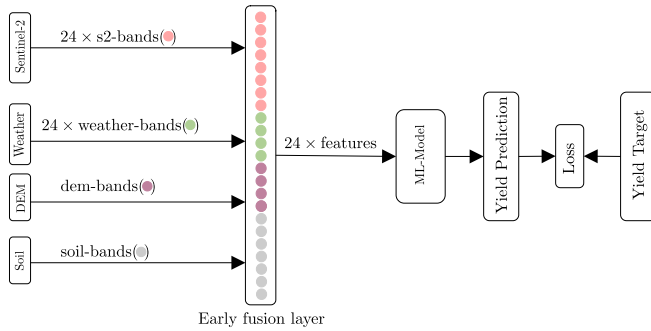
2. MATERIAL & METHODS

We include data over different countries, crop types and years. In detail, we use data coming from *Germany, Argentina, and Uruguay*. For each country, different crop types are available, including *wheat, rapeseed, and soybean*. For yield forecasting, we use gradient boosting and deep learning-based methods.

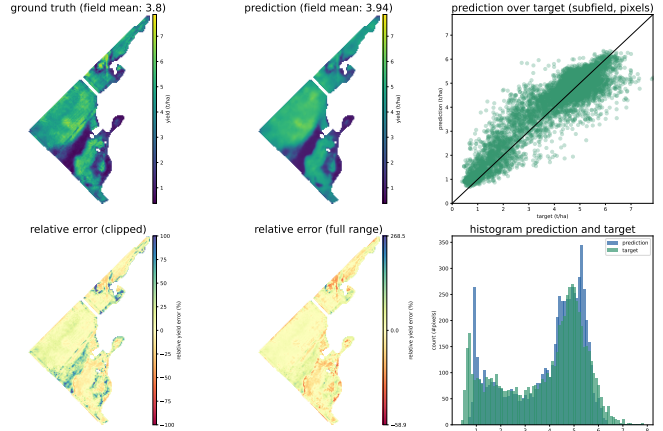
2.1. Data

For training, yield data is used as ground truth, combined with remotely sensed input data available with global coverage as predictive features.

*both authors contributed equally to this work



(a) Early Fusion Model



(b) Prediction, target and yield distribution.

Fig. 1: (a) Framework for multimodal data fusion for yield predictions. Multiple modalities with different spatial and temporal resolutions are fused at the input level. A machine learning model is then trained pixel-wise to produce yield predictions in $10m$ resolution. (b) Performance plots for visual inspection of a single field. Yield data from soybean in Argentina is shown, harvested in 2021. The model was trained on Sentinel-2 and DEM data. Upper left: ground truth yield map, upper middle: pixel-based yield prediction, upper right: scatterplot comparing predictions with ground truth data, lower left: relative prediction clipped at 100%, lower middle: relative prediction error in full range, lower right: distribution plot of predictions against the target.

Yield Data Yield data from combine harvesters on a sub-field level is used as ground truth data. While harvesting, the combine harvester with yield monitors drives through the field, collecting equidistant data points in high spatial resolution. Each data point is characterized by different features such as the geographic coordinate, the amount of yield in t/ha, the yield moisture in %. We use a standardized data preprocessing pipeline to harmonize the raw yield data. This includes reprojecting the coordinate reference system, standardization of feature naming, and removing erroneous values for position, timestamp, yield, moisture, and non-activated harvesters. Zero yield points and biologically infeasible points are removed. In addition, data points are filtered by statistical thresholds, meaning that a yield point must be within three standard deviations. For more details, we refer the reader to [6]. The resulting point vector data is rasterized into $10m$ resolution yield maps aligning with satellite imagery raster data. Tab. 1 gives an overview of the used yield datasets.

Sentinel-2 Data All experiments use cloud-free Sentinel-2 (L2A) images (S2) with $10m$ resolution for model training. Spectral bands with lower resolutions are upsampled to $10m$ resolution, resulting in twelve spectral bands. Images are collected within the growing period, i.e., between each field’s seeding and harvesting date.

Additional Data Modalities In addition to satellite imagery, we select a set of data modalities that are known to play a role in plant development and yield formation. We can

categorize Additional Data Modalities (ADM) into weather data, soil data, and DEM data. Weather data for each field is derived from the ECMWF Reanalysis (ERA5) [7], soil data from SoilGrids in $250m$ resolution [8], and DEM data from NASA’s Shuttle Radar Topography Mission (SRTM)[9] in $30m$ resolution. We prepared the ADM based on the bounds of the field using ground truth data. For soil and DEM data, raster images are created and upsampled to $10m$ resolution using a cubic spline interpolation. For soil, we use all eight available soil properties, i.e. *cec*, *cfvo*, *nitrogen*, *phh2o*, *sand*, *silt*, *soc*, *clay* at depth of 0-5, 5-15, and 15-30 cm. For DEM, we used the RichDEM [10] tool for feature engineering and deriving more features that include *aspect*, *curvature*, *dem*, *slope*, *twi*. Weather data is aggregated for each day at field level for minimum, maximum, and mean temperature and total precipitation.

2.2. Data Preprocessing

For each field, input data is represented as a sequence of 24 timesteps defining two calendar years (a sample for each month) with the harvesting date in the second year. We mask all samples outside the crop season, i.e., before seeding and after harvest. S2 images are used as reference data to create 24 timesteps [11] by selecting the best cloud free S2 image among all images within each time interval, and features from other modalities are concatenated for each timestep with S2 features. Daily weather data is aggregated by summing all values between each time interval based on the dates of the

selected S2 images. Soil and DEM features are vectorized and repeated at each timestep. This preprocessing results in a multivariate time series in which each sample represents a raster pixel of the yield map, with a maximum of 45 features at each timestep, depending on the selected ADM.

Table 1: Yield map (fields) data per country and crop type for different years.

Country	Years	Rapeseed	Wheat	Soybean	Sum
Germany	2016-2022	111	188	0	299
Uruguay	2018-2021	0	0	486	486
Argentina	2017-2022	0	0	192	192
Sum		111	188	678	977

2.3. Methods

We used state-of-the-art machine learning and deep learning models to capture in-field variability, namely *Light Gradient-Boosting Machine (LGBM)* [12] and *Long Short-Term Memory (LSTM)* [13]. An overview of the proposed framework is illustrated in Fig.1a. Following the early fusion method [14], a multivariate time series is created, wherein each timestep represents concatenated features described in sec. 2.2. The time series is further fed to a machine learning model for a regression task, where each sample represents a pixel with 10m resolution based on S2 images. For LGBM, the preprocessed data is vectorized by concatenating all timesteps into one vector. For LSTM, the preprocessed data is used sequentially, one timestep at a time to feed the model. The LGBM model is trained using regression objective with *gbd*t boosting type, learning rate as 0.1, and early stopping round as 10. In the LSTM model, 2 stacked LSTM layers with 128 hidden units are used, followed by two fully-connected layers with 128 and 1 neurons respectively, separated with a ReLU non-linear activation and batch-normalization to output the predicted yield value. The LSTM model is trained using ADAM optimizer with a fixed learning rate of 0.001 and batch size 1024 for 50 epochs. An early stopping method is used to halt training if the model does not improve for 8 consecutive epochs on the validation data. To avoid overfitting, stratified grouped K-fold cross validation is used, grouped with field name and stratified with farm name. Here, a farm represents either a set of fields operated by a farmer or geographically nearby fields in case farmer information is unavailable. We report scores as the average over K-Folds, using 10-Folds in all experiments.

3. RESULTS & EVALUATION

We quantitatively and qualitatively evaluate model performance. For quantitative evaluation at field and sub-field level regression, we use the Mean Absolute Percentage Error (MAPE) and the coefficient of determination, *R-squared*(R^2). For qualitative evaluation, a three point-guideline is used. (1)

In-field variability: the model should capture sub-field differences, (2) low prediction error: low pixel-wise prediction error, (3) distribution match: prediction and target distribution must be close to each other. We consider a model trained on S2 data only as a baseline and investigate the contribution of ADM. Tab. 2 shows the effect of including ADM in addition to S2 on the performance of the LSTM model in Argentina for soybean. We observe that although all ADM improve performance, DEM data in addition to S2 boosts the performance most, i.e. an R^2 of 0.82, resulting in an improvement of 8 percentage points (p.p.) over S2 only. Similarly, in *Germany*, for rapeseed, we observe an R^2 of 0.78 by using S2 and soil data, meaning an improvement of 13 p.p. over S2 data only. Moreover, in Tab. 3, similar experiments are done for all other crops and regions. We present results of the best performing combination of model and different modalities in the context of field and sub-field level performances. Looking at the qualitative evaluation, we see reasonable performances of all models over countries, crops, and years. We note that the presented framework captures in-field variability. In addition, we observe low prediction errors and good distribution match in numerous instances. An example is shown in Fig.1b.

Table 2: Contribution of different modalities in soybean yield prediction for Argentina using the LSTM model.

Modalities	FIELD		SUBFIELD	
	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

Table 3: Results show the best-performing combination of different modalities and ML methods for distinct crops and countries at the field and sub-field level crop yield prediction.

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

4. CONCLUSION & OUTLOOK

State-of-the-art machine learning models are well suited for yield predictions over countries, crops, and years. Surprisingly, we observe regional different feature importance, resulting in the selection of input features being essential for ML-based crop yield prediction. Models trained on multi-modal data outperform models trained on satellite imagery only. Adding additional modalities with low spatial resolution significantly increases field-level performance and, moreover, improves sub-field level performance. In this study, we fo

cused on evaluating early fusion methods. Nevertheless, it is still unclear whether other fusion methods can better extract yield-driving features and thus learn to avoid insignificant modalities for the crop-region combination. Also, it is still to be examined if more data modalities, including expert knowledge, would further contribute to the models' performance.

5. ACKNOWLEDGEMENT

The research results presented are part of a large collaborative project on agricultural yield predictions, which was partly funded through the ESA InCubed Programme (<https://incubed.esa.int/>) as part of the project AI4EO Solution Factory (<https://www.ai4eo-solution-factory.de/>). H.N. and F.M. acknowledge support through a scholarship of the University of Kaiserslautern-Landau.

6. REFERENCES

- [1] Thomas Van Klompenburg, Ayalew Kassahun, and Gatay Catal, "Crop yield prediction using machine learning: A systematic literature review," *Computers and Electronics in Agriculture*, vol. 177, pp. 105709, 2020.
- [2] Merryn L Hunt, George Alan Blackburn, Luis Carrasco, John W Redhead, and Clare S Rowland, "High resolution wheat yield mapping using sentinel-2," *Remote Sensing of Environment*, vol. 233, pp. 111410, 2019.
- [3] Ahmed Kayad, Marco Sozzi, Simone Gatto, Francesco Marinello, and Francesco Pirotti, "Monitoring within-field variability of corn yield using sentinel-2 and machine learning techniques," *Remote Sensing*, vol. 11, no. 23, pp. 2873, 2019.
- [4] D Moravec, J Komárek, J Kumhálová, M Kroulík, J Prošek, P Klápště, et al., "Digital elevation models as predictors of yield: comparison of an UAV and other elevation data sources," *Agronomy Research*, vol. 15, no. 1, pp. 249–255, 2017.
- [5] Raí A Schwalbert, Telmo Amado, Geomar Corassa, Luan Pierre Pott, PV Vara Prasad, and Ignacio A Ciampitti, "Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil," *Agricultural and Forest Meteorology*, vol. 284, pp. 107886, 2020.
- [6] Cristhian Sanchez, Deepak Pathak, Miro Miranda, Patrick Helber, Benjamin Bischke, Peter Habeliz, Hiba Najjar, Francisco Mena, Jayanth Siddamsetty, Diego Arenas, Michaela Vollmer, Marcela Charfuelan, Marlon Nuske, and Andreas Dengel, "Influence of data cleaning techniques on sub-field yield predictions," in *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, 2023.
- [7] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al., "The ERA5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, 2020.
- [8] Tomislav Hengl, Jorge Mendes de Jesus, Gerard B. M. Heuvelink, Maria Ruiperez Gonzalez, Milan Kilibarda, Aleksandar Blagotić, Wei Shangquan, Marvin N. Wright, Xiaoyuan Geng, Bernhard Bauer-Marschallinger, Mario Antonio Guevara, Rodrigo Vargas, Robert A. MacMillan, Niels H. Batjes, Johan G. B. Leenaars, Eloi Ribeiro, Ichsan Wheeler, Stephan Mantel, and Bas Kempen, "Soilgrids250m: Global gridded soil information based on machine learning," *PLOS ONE*, vol. 12, no. 2, pp. 1–40, 02 2017.
- [9] Tom G Farr and Mike Kobrick, "Shuttle radar topography mission produces a wealth of data," *Eos, Transactions American Geophysical Union*, vol. 81, no. 48, pp. 583–585, 2000.
- [10] Richard Barnes, *RichDEM: Terrain Analysis Software*, 2016.
- [11] Patrick Helber, Benjamin Bischke, Peter Habeliz, Cristhian Sanchez, Deepak Pathak, Miro Miranda, Hiba Najjar, Francisco Mena, Jayanth Siddamsetty, Diego Arenas, Michaela Vollmer, Marcela Charfuelan, Marlon Nuske, and Andreas Dengel, "Crop yield prediction: An operational approach to crop yield modeling on field and subfield level with machine learning models," in *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, 2023.
- [12] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu, "Lightgbm: A highly efficient gradient boosting decision tree," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2017, NIPS'17, p. 3149–3157, Curran Associates Inc.
- [13] Sepp Hochreiter and Jürgen Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, nov 1997.
- [14] Pradeep K. Atrey, M. Anwar Hossain, Abdulmotaleb El Saddik, and Mohan S. Kankanhalli, "Multimodal fusion for multimedia analysis: A survey," *Multimedia Syst.*, vol. 16, no. 6, pp. 345–379, nov 2010.