

Multimodal Crop Type Classification Fusing Multi-Spectral Satellite Time Series with Farmers Crop Rotations and Local Crop Distribution

Valentin Barriere¹, Martin Claverie¹

¹European Commission's Joint Research Center, Via Fermi, 2749, 21027 Ispra VA, Italy

Abstract

Accurate, detailed, and timely crop type mapping is a very valuable information for the institutions in order to create more accurate policies according to the needs of the citizens. In the last decade, the amount of available data dramatically increased, whether it can come from Remote Sensing (using Copernicus Sentinel-2 data) or directly from the farmers (providing in-situ crop information throughout the years and information on crop rotation). Nevertheless, the majority of the studies are restricted to the use of one modality (Remote Sensing data or crop rotation) and never fuse the Earth Observation data with domain knowledge like crop rotations. Moreover, when they use Earth Observation data they are mainly restrained to one year of data, not taking into account the past years. In this context, we propose to tackle a land use and crop type classification task using three data types, by using a Hierarchical Deep Learning algorithm modeling the crop rotations like a language model, the satellite signals like a speech signal and using the crop distribution as additional context vector. We obtained very promising results compared to classical approaches with significant performances, increasing the Accuracy by 5.1 points in a 28-class setting (.948), and the micro-F1 by 9.6 points in a 10-class setting (.887) using only a set of crop of interests selected by an expert. We finally proposed a data-augmentation technique to allow the model to classify the crop before the end of the season, which works surprisingly well in a multimodal setting.

Keywords

Remote Sensing, Farmer's Rotations, Multimodal System, Hierarchical Model

1. Introduction

Timely and accurate crop type mapping provides valuable information for crop monitoring and productions forecast [1]. In-season crop type mapping can serve not only to better estimate the crop areas, but also to improve the yield forecasting by using crop-type specific models. Crop type mapping is thus a major information of the crop monitoring systems focusing to in-season forecast of the crop production.

The high-spatial resolution time series enables to determine crop type at a sub-parcel level in most agricultural areas. Most of the remote sensing classification systems relies on supervised techniques, requiring in-situ crop identification survey. If the survey data are provided within the season, some systems [2] are designed to predict crop type along the season with a given uncertainty, even if the crop cycle is on-going; such surveys data are expensive because of the need of labels from the current year to train a model, difficult to achieve at large scale and in most cases delivered after the cropping season. There is a high demand for crop type mapping that does not rely on survey data from the on-going season.

Such approaches, as the one proposed in this study, are based on model trained with past seasons and applied on the current one, plus we proposed a data-augmentation method to obtain satisfying results earlier in the season.

Earth Observation-based crop type mapping Machine learning classification methods have been widely tested to derive crop type map from remote sensing data. Among the various methods, Random Forest algorithm has proved its capacity to accurately identify crop type, accounting for large and non parametric data set [3]. Since 2015 and the launch of the first satellite of the Copernicus Sentinel-2 (S2) constellation, the perspective for crop type mapping at large scale has changed. The high spatial and temporal resolution of S2 offers indeed an appropriate data set to distinguish crop type, based on the spectral and temporal signals, at parcel or sub-parcel level in most agricultural region. Taking benefit of this capacity, some operational systems have been expanded [4, 2, 5], combining Earth Observation (EO) data, in situ observations and classifier algorithm to deliver crop type maps at regional, country scale or continental scale [6].

Crop type mapping using Deep-Learning method

The recent progresses in deep-learning benefit the crop type mapping applications. In [7], the authors are classifying crop types at the parcel-level, using the data from the French Brittany during the season 2017. The authors have compared a Transformer-Encoder [8] and a Recurrent Neural Network of type Long-Short-Term-Memory

CDCEO 2022: 2nd Workshop on Complex Data Challenges in Earth Observation, July 25, 2022, Vienna, Austria

✉ valentin.barriere@ec.europa.eu (V. Barriere);

martin.claverie@ec.europa.eu (M. Claverie)

🆔 0000-0002-0877-7063 (V. Barriere); 0000-0002-0877-7063

(M. Claverie)



© 2021 Copyright for this paper by its authors. Use permitted under Creative

Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

(LSTM) [9]. They obtain comparable results between the Transformers and the LSTM, obtaining best accuracy (0.69) for the former and macro-F1 (0.59) for the latter.

In [10], the authors have designed a crop classifier at the parcel-level using S2 and compared several approaches to model the signal, comprising a Transformer and a LSTM. They obtain respective overall accuracies between 0.85 and 0.92 using the LSTM depending on the number of classes considered. A similar approach has been run by [11] on 40k Central Europa parcels using S2. They proposed a new early classification mechanism in order to enhance a classical model with an additional stopping probability based on the previously seen information.

Finally, [12] are using the same technique developed in [13], where they tackle the task of crop classification at the pixel level, i.e. accounting for the spatial variation to detect parcels boundaries. They are using a CNN-LSTM network on S2 images to classify 17 types of crops.

Modeling the crop rotation sequences Crop rotation is a widely-used agronomic technique for sustainable farming, preserving the long term soil quality. Good understanding and design of crop rotation are essentials for sustainability and to mitigate the variability of agricultural productivity induced by climate change. The crop rotation depends on the farmer management decision, but some good practices are shared, enabling to model the crop rotation patterns [14]. They remains nonetheless complex and non stable in time; changes may be related to, e.g. economic consideration (commodities price) or administrative regulation (e.g. subsidies changes). Expert knowledge based models are thus very limited and rarely accurate over large areas and long periods. Alternatively, estimation of the crop sequence probabilities without a priori using survey data and hidden Markov models has been demonstrated in France [?]. However, survey data are not always available. Relying on machine learning techniques, [15] use a Markov Logic model in order to predict the following year's crop in France, with an accuracy of 60%. In [16], the authors focused on deep deep neural networks to reach a maximum accuracy of 88% on a 6-class portion of the US Cropland Data Layer (CDL) dataset over 12 years [17].

Motivation A lot of works are focusing on the use of remote sensing to predict the crop type at pixel or parcel level using only the EO and in-situ observations of the current year. Nevertheless, they consider the signal as independents from a year to another. Other works are using the crop rotations of the parcels in order to tackle a pre-season prediction of the crop type, focusing on a few classes problem. In this case, it is obvious there is too much information missing to reach high performances. As of 2022, we identified a single study combining the

use of crop rotations and satellite time-series data over several years: [18]. They present a methodology to derive near real time Cropland Data Layer over major US agricultural states. The methodology is nonetheless restricted to a limited number of crop types and the use of Random Forest classifier, while the recent progress in deep learning shows tremendous improvements in such data mining problem.

Contributions We propose to model both the crop rotations and the S2 time series signal in a multimodal way using a hierarchical Long-short-term-memory (LSTM). The contribution is unique in term of conception as no work has been proposed fusing the large amount of temporally fine-grained EO data with crop rotation analysis in an advanced deep learning method. The crop rotations and the S2 time series were enhanced by the use of the crop distributions of the neighborhood fields picked from previous year. The crop rotations are modeled over the year as words would be in a language model [19], helped by the S2 time-series data that are modeled as if it was the prosody of the speaker. Finally, the high-level features we add on the last layer of the network could be seen as the distribution of the words used by our speaker. Finally, we also propose a data-augmentation technique for the in-season classification, by randomly cropping the end of the RS time-series data. It allows to learn a model able to classify the type of crop without the whole time-series, hence before the end of the season.

2. Methodology

2.1. Dataset

The study is focused on data acquired over The Netherlands, covers the period 2009-2020 for the crop type labeling and the parcel identification, and the period 2016-2020 for the S2 data.

Crop Type data

The crop type data were obtained from the Dutch Land Parcel Identification System and GeoSpatial Aid Application, named *Basis registratie Percelen* (BRP). Dutch farmers must annually record their field parcel boundaries and associated cultivated crops.¹ The 12 yearly BRP (2009-2020) were merged through geographical polygon intersections. The output polygons correspond to the 12-year intersected areas and there are associated with 386 crop codes. The polygons which areas are lower than half an hectare were discarded. The output product contains 974,000 polygons covering a total of 1,600 Mha.

For the evaluation, we propose 3 granularities of labels, using several aggregations lead by an expert from the domain and yielding to 386, 28 and 12 crop classes.

¹<https://data.overheid.nl/data/dataset/basisregistratie-gewaspercelen-brp>

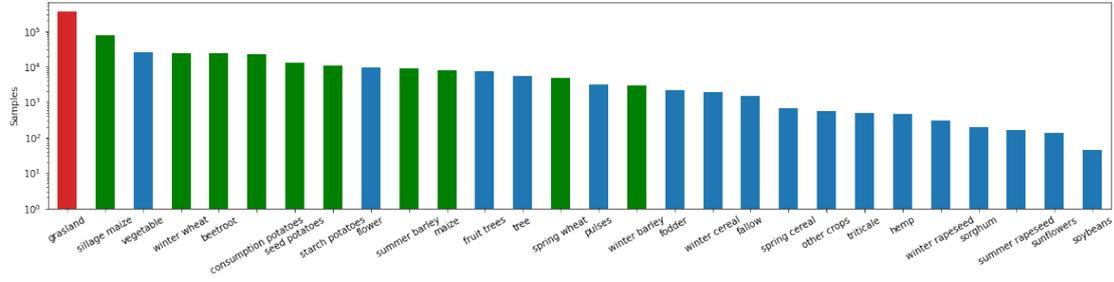


Figure 1: Distributions of the crop types in the dataset. Green crops are the remaining ones for the 10-class evaluation.

Sentinel-2 Data

Data The study relies on the analysis of the optical Copernicus Sentinel-2 (S2) data. S2 constellation provides observations with a minimum revisit of five days over ten land spectral bands of the optical domain (460-2280 nm), with a spatial resolution of 10-20 meters depending on the bands. The data are processed up to surface reflectance (SR) Level 2A accounting for atmospheric corrections and cloud/cloud-shadow screening using sen2cor algorithm [20]. The data are available through the JEODPP platform [21]. Cloud free SR data were processed to 20-m Leaf Area Index (LAI) and Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) using BV-NET [22] and calibration settings of [23]. For each polygon, B4 (red band) SR, B8A (near infrared band) SR, LAI and FAPAR were averaged at polygon level using pixels in a 20-m inner buffer in order to remove parcel edge effects.

Time series Smoothing Despite the cloud and cloud-shadow screening of L2A S2 products, noise remains in the resulting time series [24]. We applied a time series outliers detection based on B4 (for omitted cloud) and B8A (for omitted cloud-shadow) and using the Hampel filter [25]. Filtered data were removed for the four variables. The filtered time series of the four variables were smoothed using the Whittaker algorithm [26] implemented by the World Food Program.² Time series were first resampled and interpolated to a 2-day time step and then the Whittaker algorithm using the V-curve optimization of the smoothing parameter is applied. It yielded to 2-day smoothed time series of each of the four variables, from October N-1 to October N for cropping season of year N.

2.2. Feature Extractions

Crop Types

The crop types labels contains 386 different types of crops over the 12 years of study. We model the crop by a one-hot vector of size $V = 386$ and used it as an input to an embedding layer.

²<https://github.com/WFP-VAM/vam.whittaker>

The crop label categories for 2020, the year used as test set, correspond to a long-tailed class distributions, as shown for the 28-class aggregation in Figure 1.

EO-based Features

We integrate the EO time series spatially by averaging at the parcel-level, then temporally using a sliding window of size 30 days and a step size of 15 days. For each parcel, this yields to 25 windows for the whole year for each of the Remote Sensing (RS) signal, that we integrated temporally using 7 statistical functionals: mean, standard deviation, 1st quartile, median, 3rd quartile, minimum and maximum. In total we obtain $7 \times 4 = 28$ features per window, leading to 700 features per year.

With this configuration we have overlap between the windows and avoiding to loose information by breaking the signal dynamics, at the price of a bit of redundancy in the features. On each window, we integrated each signal using statistical functionals like it would be done for speech data [27].

Spatial Crop Distribution

The spatial crop distribution was derived for the year 2019 (year N-1 as compared to the 2020 validation test set). For each polygon, we compute the sum of the surface for each crops of the data base included in a 10-km circle and turned it to percentage. This a-priori distribution of crops is proven to be relatively stable in time with minor change from year to year [28]. We round the probability at 10^{-4} , leading to some values being 0 when not null.

2.3. Learning Model

This section describes the learning model and the the features' integration as observations.

Unimodal RNN-LSTM Crop Rotations model

We are modeling the crop rotation at the level of a year by using a LSTM that is trained like a language model. Indeed, it is possible to see each crop like a token in a sentence and train a recurrent neural network that will learn to predict the next word regarding the preceding words.

We firstly add an embedding layer to transform the crop type c_t at time t into a vector (see Equation 1).

$$\mathbf{emb}_t = f_e(c_t) \quad (1)$$

Then we feed this vector into the RNN to produce a hidden state h_t at time t (see Equation 2), which will be used to predict the next crop c_{t+1} (see Equation 3).

$$\mathbf{h}_t = LSTM_y(\mathbf{emb}_t | \mathbf{h}_{t-1}) \quad (2)$$

$$P(c_{t+1} | c_t, \dots, c_1) = f_c(\mathbf{h}_t) \quad (3)$$

Features from RS signal

Using only the past rotations to predict the following year's crop is very difficult, hence we chose to add available information from satellite data in order to make the model more robust.

Firstly, we enhance the unimodal LSTM crop model by adding information from RS and aligned it at the year-level before concatenating the unimodal RS vector with the crop embedding. Secondly, we chose to process the RS signal beforehand using another RNN and concatenated this unimodal RS vector obtained with the crop embedding, in a Hierarchical way. Those networks are denoted with a Hier- in their name.

Multimodal model with RS For the first model, we integrated the RS features at the year-level before the LSTM modeling the crop types. We feed the 700 features RS_t into a neural network layer f_{rs} to reduce their size and then concatenate them with the crop embeddings before the LSTM (see Equation 4), using emb_{MM_t} instead of emb_t in Equation 2. This model denoted as LSTM_{MM}

$$\mathbf{emb}_{MM_t} = [\mathbf{emb}_t, f_{rs}(\mathbf{RS}_t)] \quad (4)$$

Bidirectional RNN-LSTM with attention to model the RS time-series The first model presented above does not take into account the sequentiality of the RS signal. We decided to correct this aspect by processing the RS features at the year level with a first RNN before adding their yearly representation into the second neural network modeling the crop types, leading to a hierarchical network [29]. This will give 28 features per window \mathbf{RS}_{t_w} , for a sequence length of 25 per year.

We chose to enhance a simple LSTM with a bidirectional LSTM (biLSTM) with a self-attention mechanism [30] following the assumption that some parts of the year are more important than others to discriminate the crop type. This model denoted as HierbiLSTM_{MM}

The biLSTM is composed of 2 LSTM, one of each read the sequence forward and the other reads it backward. The final hidden states are a concatenation of the

forward and backward hidden states. For a sequence of inputs $[\mathbf{RS}_{t_1}, \dots, \mathbf{RS}_{t_w}]$ it outputs w hidden states $[\mathbf{h}_{RS_{t_1}}, \dots, \mathbf{h}_{RS_{t_w}}]$. The attention layer will compute the scalar weights u_{t_w} for each of the $\mathbf{h}_{RS_{t_w}}$ (see Equation 5) in order to aggregate them to obtain the final state \mathbf{h}_{RS_t} (see Equation 6).

$$u_{t_w} = att(\mathbf{h}_{RS_{t_w}}) \quad (5)$$

$$\mathbf{h}_{RS_t} = \sum_w u_{t_w} \mathbf{h}_{RS_{t_w}} \quad (6)$$

Locally aggregated crop distributions

When classifying at the scale of a whole country, the agricultural practices like the type of crops that are used can change. Typically the distribution of the crop types in a region is a stable value over the years and represent the kind of crops supposed to be found in this part of the world. We integrated this local information by adding a vector representing the distributions over the crop types in an area corresponding to a circle of 10-km centered around the studied parcel.

We chose to add the distribution vector before the last layer because it is a high-level feature regarding the task we are tackling and the deeper you go into the layers the higher-level the representations are w.r.t. the task [31]. We concatenated the hidden state \mathbf{h}_t of the LSTM with the crop distribution vector \mathbf{d} and mixed them using two fully connected layers f_{fc1} and f_{fc2} (see Equation 7). Hence, we obtain \mathbf{h}_{d_t} instead of \mathbf{h}_t before the final fully connected layer f_{fc} from Equation 3. This final model is denoted as Final.

$$\mathbf{h}_{d_t} = f_{fc2}(f_{fc1}([\mathbf{h}_t, \mathbf{d}])) \quad (7)$$

3. Experiments and Results

In this section we will describe the different experiments and results we ran with all the different models. Because of the nature of our predictions, it can be useful to get them before the end of the farming season. In this context, we ran experiments using different setups when predicting, we used an end-of-season configuration and an early-classification configuration. For the end-of-season configuration we feed the neural network with all the RS data of the year while in the early-classification configuration we stop to different date of the year. We compared using LSTM processing the RS data and tagging at the year-level, seeing all the year in an independent way. This year-independent model obtained state-of-the-art results according to [7] and is denoted as LSTM_{YI}.

3.1. Experimental protocol

We trained all the networks via mini-batch stochastic gradient descent using Adam as optimizer [32] with a

Labels Model	#Modal.	386-class				28-class				12-class				10-class			
		P	R	F1	Acc	P	R	F1	Acc	P	R	F1	Acc	P	R	F1	m-F1
LSTM _{Crop}	1 (C)	28.1	22.4	23.1	73.3	45.3	33.4	34.2	76.4	53.1	44.7	43.8	77.2	46.7	39.2	37.1	52.1
LSTM _{YI} [7]	1 (RS)	14.3	9.8	9.9	72.5	53.1	45.1	45.9	88.5	75.0	66.3	67.7	90.4	72.1	62.2	63.7	80.3
LSTM _{RS}	1 (RS)	13.7	11.8	10.8	72.5	49.7	47.0	44.1	87.4	70.6	69.4	65.3	89.0	67.3	65.3	60.7	76.0
HierbiLSTM _{RS}	1 (RS)	10.7	10.0	9.0	78.7	48.0	48.1	44.5	88.7	72.7	71.0	67.7	90.7	69.3	66.6	63.0	79.1
LSTM _{MM}	2 (RS+C)	32.5	26.1	26.2	86.8	63.3	57.4	56.7	91.8	79.9	78.5	78.2	93.2	77.9	75.2	75.4	85.1
HierbiLSTM _{MM}	2 (RS+C)	42.0	33.8	35.1	88.5	68.9	62.8	63.2	93.5	84.1	80.8	81.3	94.5	82.3	77.9	78.8	87.8
Final	3 (All)	41.0	33.3	34.3	89.7	71.4	62.7	63.2	93.8	85.5	81.2	82.6	94.8	84.1	78.3	80.2	88.7

Table 1

Results of the end-of-season classification models with different modalities (Remote Sensing, Crop Rotation, and Spatial Crop Distribution). The metrics shown are Macro precision, recall and F1 score, as well as accuracy and micro-F1 score (m-F1).

learning rate of 10^{-3} and a cross-entropy loss function. The number of neurons for the crop embedding layer, both the RNN internal layers, and the fully connected RS layer f_{rs} as well as the number of stacked LSTM were chosen using hyperparameters search. The sizes of the layers f_{c1} and f_{c2} are the same than the one from the second RNN state \mathbf{h}_t .

We trained our networks as for a sequence classification task, always with ten years of data. The labels from 2018 were used as training set, while the labels from 2019 as development set and the labels from 2020 as test set. All results presented hereafter refer to the analysis of 2020 crop types, which are based on models trained with the period 2009-2019, thus independent from the 2020 crop types observations. We zero-padded when no RS data was available (before 2016).

We proceed to a data-augmentation for the in-season classification model by cropping randomly the end of the timeseries for each batch starting from mid-March. All models were coded using the PyTorch library [33].

3.2. Results

In this Section we will show the results with two different settings: the classical setting where the network sees the whole year of RS signal, and a special early-season setting where the RS signal of the current season stops before the end of the season. In order to deal with unbalanced classes, we used unweighted F1, Precision and Recall as well as the Accuracy. We used also the micro-F1, which is equivalent to Accuracy when having removed classes.

We also present results for 10 classes, which is the 12-class settings without grassland and other crops (see Figure 1).

3.2.1. End-of-Season Classification

The results of the end-of-season classification are available in Table 1. We tested different configurations of networks, using different kind of features. The best results are obtained with our final model using information from the crop rotations, the S2 time series and the crop distribution of the surrounding fields.³

³from 2019

At first glance, we can see that the model using only the crop rotations can still reach an Accuracy of 73.3% for the 386-class problem even if it does not use any information from the current year to make it's prediction.

Our RS models reach high results on the 386-class (up to 78.7% with the HierbiLSTM_{RS} model) due to the fact that, contrary to the main part of the works, they also use RS data from the past years on the same parcel, allowing to model a temporal context. Interestingly, the Hierarchical setup with RS only allows for reaching higher results on the 386-class configuration, going from an accuracy of 72.5 to 78.7, when compared to the LSTM_{YI}.

Finally, the local crop distribution vector allow for a slight improvement, which is more visible in the 10-class configuration. However, it unexpectedly decreases the macro-F1 while increasing the Accuracy for the 386-class configuration. This can be interpreted as the model making more mistakes on non-frequent crops only because it's globally better. An explanation can be that the non frequent crops are not all situated in the same area, hence their distribution probability density is always approximated as 0.

3.2.2. Toward In-Season Classification

We saw earlier that the RS signal has shown pretty good end-of-season results, but it is known that the performances are strongly degraded when classifying during the season[11]. In this case, the crop rotations enhanced modality can help.

For the in-season classification, we simply used our model trained over the whole year with data stopping at a point of the year. In Figure 2, we compared the model using RS signal only with the multimodal model. It is important to notice that we used the same "final" model to adapt our domain to this noisy setup. The missing features, corresponding to unused months, were replaced by zeros. The results are thus preliminary and it is expected to obtained poor performances. A straightforward option could be to train new models for each of the evaluated months of the in-season classification.

The multimodal model always outperforms the RS model which is expected, especially at the beginning of the season when almost no information is available using the RS modality.

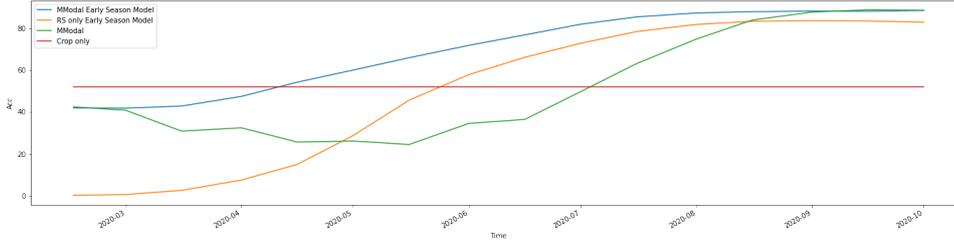


Figure 2: Comparison of Early classification using different modalities, with/out data augmentation (m-F1 with 10 classes).

Labels Thresh	12-class			10-class		
	P	F1	Acc	P	F1	m-F1
none	85.3	82.6	94.8	84.1	80.2	88.7
.9	93.7	88.2	98.2	93.1	86.4	95.8

Table 2

Results with the model using only the examples with a high probability for the predicted class.

4. Analysis

For the sake of clarity, all analyses presented hereafter in this section are limited to a set of crops of interest, corresponding to the 10-class setting. Details on the crops are provided in Figure 1.

High Precision examples

We are presenting the results of our model on a fewer parcels where the precision is better than normal. In the perspective of crop monitoring, this analysis can be very valuable. Even if not 100% of the parcels are aggregated, the output might support crop yield forecasting system, through the analysis of the crop specific RS time series with highest probability.

We are taking the examples that are classified with a probability superior to 0.9 and compute some metrics over them. Those examples represent a big part of the dataset, they are more than 536k for the 12-class dataset and more than 148k for the 10-class dataset, representing respectively 90.0% of all the parcels and 76.5% of the parcels containing crop of interests. The results are shown in Table 2.

In-Season Classification

We compare the vanilla model with the in-season classification model trained with our data-augmentation technique. The vanilla model has only seen during training examples of end-of-season classification, it is normal that they perform worst when used in in-season. This explains the fact that there is a decrease in performance compared to a model only taking into account the crops.

The data-augmentation used for the in-season models surprisingly does not work with RS only model, but allow the multimodal model to overpass the crop-only model in April.

It is also interesting noting that the performances go below the unimodal crop model. This is certainly related that the models may give too much attention to the RS modality compared to the other ones, because the RS data modality has higher impact on the performance as the season progresses. An option to counter this effect would be to use a gate that would discard a noisy modality, as shown in [34, 35].

5. Conclusion and Future works

We presented an innovative study to produce in-season crop mapping without relying on in-situ data of the current season. The approach relies on the analysis of several modalities, including the crop rotation of the previous year as well as the previous year local crop distributions in the neighborhood parcels. A deep learning algorithm was used to model all those modalities at different level using a Hierarchical LSTM model. Firstly, we modeled the RS data with a Bidirectional-LSTM with Attention, using a sliding window on the satellite signals and integrating them using statistical functionals as it can be done for speech. Secondly, we fed the representation into another LSTM network modeling the crops as words and their rotation as sentence as it can be done with a language model. Finally, we added a context vector on the last layer in order to add information about the geographical place of the parcel. The designed methodology was tested over cropland of the Netherlands, benefiting from 12 years of crop rotation data nationwide. More generally, our method outperforms by a great margin the classical state-of-the-art using only a RNN or a Transformer to model the EO data at the level of a year.

Nevertheless, there is still a lot of place for future work. More spectral bands added in the EO data could improve the performances. A better way to model the multimodality, at the level of EO data using multimodal aligned or non-aligned time-series fusion models[36, 37], and at a higher level between static representations [34]. Finally our model impossible to adapt to an unknown place where the crop rotations are not available, a domain adaptation method using few-shot learning could be useful in this case [38].

References

- [1] I. Becker-Reshef, C. Justice, M. Sullivan, E. Vermote, C. Tucker, A. Anyamba, J. Small, E. Pak, E. Masuoka, J. Schmaltz, et al., Monitoring global croplands with coarse resolution earth observations: The global agriculture monitoring (glam) project, *Remote Sensing* 2 (2010) 1589–1609.
- [2] P. Defourny, S. Bontemps, N. Bellemans, C. Cara, G. Dedieu, E. Guzzonato, O. Hagolle, J. Inglada, L. Nicola, T. Rabaute, et al., Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the sen2-agri automated system in various cropping systems around the world, *Remote sensing of environment* 221 (2019) 551–568.
- [3] M. Hansen, R. Dubayah, R. DeFries, Classification trees: an alternative to traditional land cover classifiers, *International journal of remote sensing* 17 (1996) 1075–1081.
- [4] J. Inglada, M. Arias, B. Tardy, O. Hagolle, S. Valero, D. Morin, G. Dedieu, G. Sepulcre, S. Bontemps, P. Defourny, et al., Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery, *Remote Sensing* 7 (2015) 12356–12379.
- [5] D. M. Johnson, R. Mueller, et al., The 2009 cropland data layer, *Photogramm. Eng. Remote Sens* 76 (2010) 1201–1205.
- [6] R. d’Andrimont, A. Verhegghen, G. Lemoine, P. Kempeneers, M. Meroni, M. van der Velde, From parcel to continental scale—a first european crop type map based on sentinel-1 and lucas copernicus in-situ observations, *Remote Sensing of Environment* 266 (2021) 112708.
- [7] M. Rußwurm, S. Lefèvre, M. Körner, Breizhcrops: a satellite time series dataset for crop type identification, *Time Series Workshop of the 36th ICML* (2019).
- [8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention Is All You Need, *arXiv:1706.03762 [cs]* (2017).
- [9] S. Hochreiter, J. Schmidhuber, LONG SHORT-TERM MEMORY, *Neural Computation* 9 (1997) 1735–1780. URL: <http://www7.informatik.tu-muenchen.de/~hochreit/5Cnhttp://www.idsia.ch/~juergen>. doi:10.1162/neco.1997.9.8.1735. arXiv:1206.2944.
- [10] M. Rußwurm, M. Körner, Self-attention for raw optical Satellite Time Series Classification, *ISPRS Journal of Photogrammetry and Remote Sensing* 169 (2020) 421–435. doi:10.1016/j.isprsjprs.2020.06.006. arXiv:1910.10536.
- [11] M. Rußwurm, R. Tavenard, S. Lefèvre, M. Körner, Early Classification for Agricultural Monitoring from Satellite Time Series, *ICML AI4socialgood workshop* (2019). URL: <http://arxiv.org/abs/1908.10283>. arXiv:1908.10283.
- [12] M. Rußwurm, M. Körner, Multi-temporal land cover classification with sequential recurrent encoders, *ISPRS International Journal of Geo-Information* 7 (2018). doi:10.3390/ijgi7040129. arXiv:1802.02080.
- [13] M. Rußwurm, M. Körner, Temporal Vegetation Modelling Using Long Short-Term Memory Networks for Crop Identification from Medium-Resolution Multi-spectral Satellite Images, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops 2017-July* (2017) 1496–1504. doi:10.1109/CVPRW.2017.193.
- [14] S. Dogliotti, W. Rossing, M. Van Ittersum, Rotat, a tool for systematically generating crop rotations, *European Journal of Agronomy* 19 (2003) 239–250.
- [15] J. Osman, J. Inglada, J. F. Dejoux, Assessment of a Markov logic model of crop rotations for early crop mapping, *Computers and Electronics in Agriculture* 113 (2015) 234–243. doi:10.1016/j.compag.2015.02.015.
- [16] R. Yamasu, V. Bandaru, K. Pnvr, Pre-season crop type mapping using deep neural networks, *Computers and Electronics in Agriculture* 176 (2020) 105664. URL: <https://doi.org/10.1016/j.compag.2020.105664>. doi:10.1016/j.compag.2020.105664.
- [17] C. Boryan, Z. Yang, R. Mueller, M. Craig, Monitoring US agriculture: The US department of agriculture, national agricultural statistics service, cropland data layer program, *Geocarto International* 26 (2011) 341–358. doi:10.1080/10106049.2011.562309.
- [18] D. M. Johnson, R. Mueller, Pre-and within-season crop type classification trained with archival land cover information, *Remote Sensing of Environment* 264 (2021) 112576.
- [19] T. Mikolov, M. Karafiát, L. Burget, J. Černocký, S. Khudanpur, Recurrent neural network based language model, in: *Eleventh annual conference of the international speech communication association*, 2010.
- [20] J. Louis, V. Debaecker, B. Pflug, M. Main-Knorn, J. Bieniarz, U. Mueller-Wilm, E. Cadau, F. Gascon, Sentinel-2 sen2cor: L2a processor for users, in: *Proceedings Living Planet Symposium 2016*, Spacebooks Online, 2016, pp. 1–8.
- [21] P. Soille, A. Burger, P. Hasenohr, P. Kempeneers, D. Rodriguez Aseretto, V. Syrris, V. Vasilev, D. Marchi, The jrc earth observation data and processing platform, *Big Data From Space*, Toulouse,

- France (2017).
- [22] F. Baret, O. Hagolle, B. Geiger, P. Bicheron, B. Miras, M. Huc, B. Berthelot, F. Niño, M. Weiss, O. Samain, et al., Lai, fapar and fcover cyclopes global products derived from vegetation: Part 1: Principles of the algorithm, *Remote sensing of environment* 110 (2007) 275–286.
- [23] M. Claverie, E. F. Vermote, M. Weiss, F. Baret, O. Hagolle, V. Demarez, Validation of coarse spatial resolution lai and fapar time series over cropland in southwest france, *Remote Sensing of Environment* 139 (2013) 216–230.
- [24] M. Claverie, J. Ju, J. G. Masek, J. L. Dungan, E. F. Vermote, J.-C. Roger, S. V. Skakun, C. Justice, The harmonized landsat and sentinel-2 surface reflectance data set, *Remote Sensing of Environment* 219 (2018) 145–161.
- [25] R. K. Pearson, Outliers in process modeling and identification, *IEEE Transactions on control systems technology* 10 (2002) 55–63.
- [26] P. H. Eilers, V. Pesendorfer, R. Bonifacio, Automatic smoothing of remote sensing data, in: 2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), IEEE, 2017, pp. 1–3.
- [27] B. Schuller, S. Steidl, A. Batliner, J. Hirschberg, J. K. Burgoon, A. Baird, A. Elkins, Y. Zhang, E. Coutinho, K. Evanini, The INTERSPEECH 2016 Computational Paralinguistics Challenge: Deception, Sincerity & Native Language, in: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2016.
- [28] F. A. Merlos, R. J. Hijmans, The scale dependency of spatial crop species diversity and its relation to temporal diversity, *Proceedings of the National Academy of Sciences* 117 (2020) 26176–26182.
- [29] I. V. Serban, A. Sordoni, Y. Bengio, A. Courville, J. Pineau, Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models (2015). URL: <http://arxiv.org/abs/1507.04808>. doi:10.1017/CBO9781107415324.004. arXiv:1507.04808.
- [30] D. Bahdanau, J. Chorowski, D. Serdyuk, P. Brakel, Y. Bengio, End-to-end attention-based large vocabulary speech recognition, *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings 2016-May (2016)* 4945–4949. doi:10.1109/ICASSP.2016.7472618. arXiv:1508.04395.
- [31] V. Sanh, T. Wolf, S. Ruder, H. Court, H. Row, A Hierarchical Multi-task Approach for Learning Embeddings from Semantic Tasks, in: AACL, 2018. arXiv:arXiv:1811.06031v1.
- [32] D. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, *International Conference on Learning Representations (2014)* 1–13. URL: <http://arxiv.org/abs/1412.6980>. doi:<http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.1830503>. arXiv:1412.6980.
- [33] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimeshain, L. Antiga, et al., Pytorch: An imperative style, high-performance deep learning library, *Advances in neural information processing systems* 32 (2019) 8026–8037.
- [34] J. Arevalo, T. Solorio, M. Montes-Y-Gómez, F. A. González, Gated multimodal units for information fusion, in: 5th International Conference on Learning Representations, ICLR 2017 - Workshop Track Proceedings, 2017. arXiv:1702.01992.
- [35] M. Chen, S. Wang, P. P. Liang, T. Baltrušaitis, A. Zadeh, L.-P. Morency, Multimodal sentiment analysis with word-level fusion and reinforcement learning, in: Proceedings of the 19th ACM International Conference on Multimodal Interaction, 2017, pp. 163–171.
- [36] A. Zadeh, P. P. Liang, N. Mazumder, S. Poria, E. Cambria, L.-P. Morency, Memory Fusion Network for Multi-view Sequential Learning, in: AACL, 2018. arXiv:arXiv:1802.00927v1.
- [37] J. Yang, Y. Wang, R. Yi, Y. Zhu, A. Rehman, A. Zadeh, S. Poria, L.-p. Morency, Unaligned Human Multimodal Language Sequences (2020). arXiv:arXiv:2010.11985v1.
- [38] M. Rußwurm, S. Wang, K. Marco, D. Lobell, Meta-Learning for Few-Shot Land Cover Classification, in: IEEE/CVF conference on computer vision and pattern recognition workshops, 2019.