EDEM: Explaining Deep ℓEarning Models of Satellite Time Series for the Agritech domain.

Context and Motivation Ensuring sustainable food security and promoting transparent global agricultural markets is a key objective of the European Green Deal [1]. Historically, crop productivity in temperate regions like France used to be relatively stable. Increasingly frequent extreme weather events strongly affect European agriculture, particularly in France, posing a threat to sustainable food production. In this context, an urgent need exists for the development of novel approaches based on **Big Data Analytics** and **Artificial Intelligence (AI)** capable of discovering, collecting, and mining big data sets and leveraging **global satellite data**. Specifically, we note that many modern technologies used in agriculture based **on Deep Learning** (DL) models are often considered black boxes due to their complex and opaque decision-making processes. This lack of transparency is a showstopper in the adoption process of these technologies by end-users. Farmers crave for decision-making-support to adapt agriculture to a changing climate (e.g., monitoring of field and crop damage). Industry, policy makers, and large aggregators such as the European Commission's Joint Research Centre, seek tools to assess crop loss at regional level.

In EDEM we are interested in developing novel Explainable Artificial Intelligence (xAI) techniques applied to DL models adopted in multi-dimensional time series[1] (MTS) analysis. In fact, satellite multispectral images of agricultural parcels recorded over time (aka satellite MTS) are typically used to monitor crop phenology. We will therefore validate our contribution measuring how xAI enhances decision-making processes and optimizes large-scale agricultural monitoring using publicly available benchmark dataset.

Positioning in relation to the state of the art. Despite developments in the use of MTS of satellite data for crop mapping[2] [2] at European scale, an urgent need exists for systems to make improved crop monitoring and mapping [3,4]. Focusing on xAI methods in the geoscience area, few methods have been recently applied to satellite data [5,6,7]. In [6], xAI has been used to facilitate the interpretation of variables extracted from satellite and meteorological data and environmental conditions that lead to yield variability in the Indian Wheat Belt. In [7], the authors test whether the critical dates for crop disambiguation, derived from the attention weights of the transformation model, are aligned with prior knowledge of crop phenology. Outside from the geoscience domain, several forms of explanations studied for unstructured data (e.g., images) have been recently transferred to explain DL models on MTS. The most recurrent one is **Relevance Attribution** [8], where importance scores are assigned for each data feature over time (i.e., the dimensions in an MTS) in **forecasting**, **classification** and **anomaly detection** tasks. Despite the popularity of relevance attribution methods, existing techniques applied to MTS cannot properly explain the causes that drive the model to assign a specific target (e.g., classification, prediction) to a given instance. In fact, they aim to explain the most important variables by exclusively leveraging the correlation dependencies learned over the data, rather than considering the causal relationships among the random variables in play. In this context, we report an **example in Figure 1**, which depicts the feature relevance scores

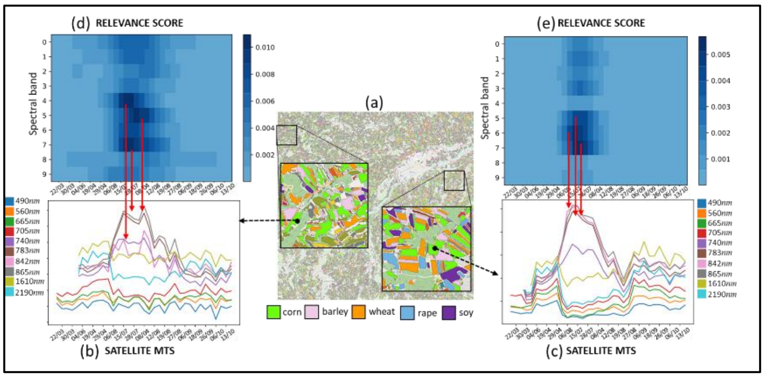

__Figure 1__ Crop type classification in Sentinel-2 optical satellite sensors images. (a) Geographical area (dataset obtained from the Google earth engine tool). (b),(c) MTS reporting the spectral bands wavelength along one year of two distinct pixels (corn cultivations). (d),(e) Relevance Attribution scores that explain the most relevant MTS features for the (correct) Temporal CNN classification.

[1] Multivariate time series are sets of one-dimension variables (a.k.a metrics), namely sequences of real values recorded over time at a fix interval rate. We may also refer to multivariate time series by the generic definition *temporal data.*

[2] Crop mapping (a.k.a. crop classification) is the operation of automatically define the types of agricultural species that grow in a selected region, the size of the territory they cover, and their stage of development at a specified time.

(heatmaps) in the classification of two satellite MTS collected by the Sentinel-2 multispectral optical satellite sensors in order to reveal the type of cultivation (at pixel level). The two MTS (Figure(b),(c)) report the temporal evolution of the central wavelength of ten spectral bands (y-axis) recorded on two distinct pixels along one year of phenological calendar (timestamps on the x-axis). The pixels belong to corn fields, which are correctly classified by a temporal CNN [9,10] trained on data collected in an Austrian agricultural area (Figure(a)). The heatmaps of Figure(d) and (e) show the relevance score of each feature (along time) in the final classification. As we can observe, in Figure 1, the **Explainability scope** is limited to an individual instance, i.e., a specific crop parcel. As a result, we may discover inconsistent relevance scores over different MTS. For example, in Figure 1(b), the features with the highest relevance do not match with those in Figure 1(c) (see the red arrows). The available explanations do not shed any light on which are the causes of this variation, which in this example may be related to the different spatial locations (i.e., different soil type, landscape position, management history, etc.) or extreme weather events which may generate crop yield loss. The same reasoning applies to instances extracted from different years, where the problem is how to explain feature relevance shifts over time.
In our work, we want to develop xAI approaches that provide a *stronger descriptive approach* to DL algorithms, as well as additional information to end-users, hence enhancing data insights. *High-level insights* regarding model predictions aim to assist analysts in understanding the underlying data generation process in various domains, but also to validate model outcomes and debug model internals to improve its performance.

Added value in terms of scientific contribution. EDEM aims to propose a novel Explainability framework containing methods able to capture causal relationships of model outcomes with important events inferred both from the data, and from MTS DL models building blocks. Such contributions have many practical benefits on Deep Domain Adaptation and Generalization [11,12,13] where the DL training process typically faces label scarcity, while the model is required to generalize over multi-source data features that can drastically differ. Domain Adaptation has been widely used in the remote sensing literature to tackle the demanding task of classifying satellite data for large-scale monitoring and mapping [11,12,13]. For the specific task of agriculture, analysts require to understand the heterogeneous characteristics of the geographical study areas located in different countries. Moreover, they need to cope with the varying climate evolution across the years and the different imaging resolutions of the available sensors, which affect the applicability of MTS DL models due to irregular changes in the data distribution. While other data types, such as images, are naturally interpretable signals with high spatial redundancy, MTS incorporate a low semantic level even in presence of consolidate domain knowledge. Hence, experts require more complex form of explanations able to match human intuition and interpret main causes behind the final outcome of a DL model applied to MTS. We note that such causes can have a dual nature. First, specific data variables or instances (e.g., parents' node in a structural causal model, repeated patterns, anomaly) can induce a DL prediction model to assign importance to certain features for obtaining the final decision. Moreover, results can also depend on the model's internal primitives (e.g., convolutional filters, positional encoding, attention heads) that drive the feature learning mechanism, and thus become the root causes of the final decisions.

Methodology, Validation and Project Organization. We organize our project into three parts, referring to three fundamental problems, each requiring a year of dedicated work from a Ph.D student.

Problem 1. The first problem we want to solve is to report **features importance in term of their causal effect** on the outcome of DL classifiers applied to the classification of agricultural areas. We note that several MTS DL models such as Long Short-Term Memory (LSTM), and Transformer [14] obtain state-of-the-art classification performance in crop mapping [4,15,16] and can be moreover adopted to infer Granger causality relations between each pair of MTS dimensions. So far, this important relation that describes the data generation mechanism has not been studied in Explainability methods based on Relevance Attribution. As a starting point, we consider the possibility to compute feature relevance by quantify neuron activations while ablating each potential cause (MTS variable at a given time t), thus obtaining relevance scores that integrate a causal semantic. This option poses different challenges. First, we need to deal with a large investigation space typical of MTS, where testing all the possible combinations of variable/time lags and the resulting network settings quickly becomes prohibitively expensive. Therefore, we need a novel filtering strategy that allows us to infer explanations in a reduced space, while also considering the different DL models available. In addition, we want to study here other types of conditional dependency test (e.g., causal Markov models) [17], that can leverage DL models of temporal data. This will permit us to fulfil the limits of Granger causality, that only allows us to test causality on variable pairs, disregarding more complex relationships (e.g., in the Variables Markov Blanket). Our recent study [18] confirms that leveraging causal structural models allows to provide more effective explanations that can be leveraged to build surrogate models with improved classification performance (compare to the original model). To validate the results obtained, we will leverage the public availability of benchmark datasets, including France-specific data, such as the *BreizhCrops* dataset [14] and the *TimeMatch* dataset [16].

Problem 2. We want to provide explainability solutions that enable automatic tuning of DL architectural elements in regions distinct from the ones used for training the model. While Problem 1 concerns explanation through data driven causes, we want to study how to explain DL models by analysing the effects of neural network internals on the final outcomes.
Exploiting such kind of explanations can be beneficial to understand the reasons why the performance of a DL crop classifier over ***satellite MTS*** considerably drops when changing the domain from both temporal and spatial perspectives. From an operational viewpoint, it is unfeasible to continuously acquire a large number of reference data every year to generate updated maps. In our recent work, we have already started to study explainability over DL architectural elements [19] (i.e., MTS Positional Encoding of Temporal Attention Encoders). Specifically, our solution permits to correct the quality of network input embeddings observing a significant improvement of crop mapping accuracy on domains in which data labels are not available. As a next step, we aim to explain and intervene on DL architecture in a holistic manner. The developed xAI solutions will boost confidence and enhance the quantification of uncertainties in the current DL method for crop type mapping. This will empower end-users to understand and trust the DL models and their outputs. Additionally, these solutions can enhance the generalization capabilities of DL models, leading to improved consistency in generating accurate crop type mapping results on a country or continental scale.

Problem 3. The third problem we want to tackle concerns the **generation of MTS concepts that can drive Explainability by counterfactual hypothesis testing**. Concept-based explanation approach is a very-well studied *interpretability* tool in Machine Learning, as it can express the reasons for a model prediction in terms of concepts that are well-*known* by expert users. Generating concepts out of an imaging dataset represents a task that often matches human intuitions [20,21]. On the contrary, extracting meaningful events in MTS datasets poses several challenges. To that extent, we want to focus on concept-based extraction from MTS, which can support the generation of counterfactuals for causal tests in a DL model. This will provide augmented validation capabilities to the users. In literature, there exists several concepts characterization of univariate time series concepts, such as Shaplets [22] and Prototypes [23], which are data subsequences that identify a certain class. In a similar manner we would like to define concepts, providing a novel model that can describe causes (multivariate sub-instances), which identify a class or an outlier with a certain statistical significance. Such causes can enable several conditional tests on MTS analysis that adopt DL models. One example can be to quantify the sensitivity of the neural network activation with respect to a given concept. In the agricultural domain, we aim to investigate the impact of the crop rotation practice (that may introduce large changes in the prior probabilities of the crop type classes), the satellite data acquisition conditions (e.g., atmospheric condition, illumination, look/view angle, sensor parameters), the climatic conditions (e.g., annual temperature can strongly vary from one year to another) as well as the environmental conditions (e.g., topography, soil moisture). These factors may lead to a strong variation, thus sharply affecting the quality and reliability of the DL model results.

Team. Dr. Michele LINARDI is an assistant professor (maître de conferences) with Cergy Paris University (CYU), currently affiliated with ETIS laboratory (team MIDI). The research topics of Dr. Linardi are mainly about time series analytics and databases, with a great interest in machine learning for temporal data. Among different contributions addressing TS management and analytics problems. Dr. Linardi was the responsible for a project financed by the INS2I/CNRS Institute, whose objective was the study of Explainable MTS models applied to Remote Sensing Analysis. The project was over in November 2022, and it has permitted Dr. Linardi and Pr. Vasillis Christophides (ETIS laboratory) to start collaborating with Dr. Claudia Paris, who will be also actively participating in the EDEM team. Dr. Claudia Paris Assistant Professor at the University of Twente, Faculty ITC, Department of Natural Resources. She has experience in designing advanced and efficient methods for the analysis and the classification of large-scale Earth Observation data for different applications (e.g., forestry, agriculture, urban areas). To perpetuate this research, we will hire a Ph.D. student thanks to the funding of this ANR project. Prof. Vassilis Christophides will be the principal thesis supervisor (Directeur de thèse), and Dr. Linardi will be formally a co-supervisor. Dr. Paris (formally the second Ph.D thesis co-supervisor) will support the successful PhD candidate in the analysis of the time series of satellite data, including the application of deep learning models to extract meaningful phenological crop information. Moreover, she will provide scientific guidance in the interpretation of the crop type mapping results obtained.

Requested Resources. We plan a duration of the project of 42 months. At the beginning of this period, we will reserve some time to audition and find a candidate that will follow a Ph.D. program for 36 months.

REFERENCES

[1] K. Hainschet al., “Energy transition scenarios: What policies, societal attitudes, and technology developments will realize the eu green deal?” Energy , 2022 **[2]** Raphaël d’Andrimont, et al., 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, 2021 **[3]** C.Persello, **Claudia Paris**, et al. “AI4SmallFarms: A Data Set for Crop Field Delineation in Southeast Asian Smallholder Farms”, IEEE Geoscience and Remote Sensing Letters. **[4]** G. Weikmann, **Claudia. Paris**, et al., "Multiyear Mapping of Water Demand at Crop Level: An End-to-End Workflow Based on High-Resolution Crop Type Maps and Meteorological Data," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **[5]** Caroline M Gevaert, “Explainable ai for earth observation: A review including societal and regulatory perspectives,” International Journal of Applied Earth Observation and Geoinformation **[6]** leksandra Wolanin et al., “Estimating and understanding crop yields with explainable deep learning in the indian wheat belt,” Environmental research letters, 2020. **[7]** Ovica Obadic et al., “Exploring self-attention for crop-type classification explainability,” arXiv preprint arXiv:2210.13167, 2022. **[8]** Theissler et al., (2022), Explainable AI for Time Series Classification: A Review, Taxonomy and Research Directions, in IEEE Access **[9]** Paul Boniol, et.al. dcam: Dimension-wise class activation map for explaining multivariate dataseries classification. SIGMOD ’22. **[10]** Kevin Fauvel, et al. XCM: an explainable convolutional neural network for MTS classification. CoRR, abs/2009.04796, 2020. **[11]** Indrajit Kalita, et al. Deep learning-based cross-sensor domain adaptation under active learning for land cover classification. IEEE Geoscience and Remote Sensing Letters, 2022. **[12]** Zhe Liu, et al. Cross-year reuse of historical samples for crop mapping based on environmental similarity. Front. in Plant Science, 2022. **[13]** Ziqiao Wang, et al. Phenology alignment network: A novel framework for cross-regional time series crop classification. In CVPR Workshops 2021. **[14]** Rußwurm, M et al., (2019). Breizhcrops: A time series dataset for crop type mapping. arXiv preprint arXiv:1905.11893. **[15]** Nyborg, Joachim, et al. "Generalized classification of satellite image time series with thermal positional encoding." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. **[16]** Nyborg, J., et al. (2022). TimeMatch: Unsupervised cross-region adaptation by temporal shift estimation. ISPRS **[17]** Raha Moraffah, et al. Causal inference for time series analysis: problems, methods and evaluation. Knowl. Inf. Syst., 2021. **[18]** Etienne Vareille and Adel Abbas and **Michele Linardi** and **Vassilis Christophides**, Evaluating Explanation Methods of Multivariate Time Series Classification through Causal Lenses. IEEE DSAA 2023 **[19]** Adel Abbass and **Michele Linardi** and Etienne Vareille and **Vassilis Christophides** and **Claudia Paris**, Towards Explainable AI4EO: an Explainable Deep Learning Approach for Crop Type Mapping in Satellite Image Time Series. IEEE IGARSS 2023 **[20]** Been Kim, et al. Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (tcav). 2018.**[21]** Mohammad Taha Bahadori and David Heckerman. Debiasing concept-based explanations with causal analysis. In ICLR 2021. **[22]** Lexiang Ye et.al. Time series shapelets: a novel technique that allows accurate, interpretable and fast classification. DAMI, 2011. **[23]** Alan H. Gee, et al. Explaining deep classification of time-series data with learned prototypes. KDH@IJCAI 2019.