FEW SHOT CROP MAPPING USING TRANSFORMERS AND TRANSFER LEARNING WITH SENTINEL-2 TIME SERIES: CASE OF KAIROUAN TUNISIA

Mohamed Karim Keraani [1], Khalil Mansour [1], Bilel Khlaifia [2], Nesrine Chehata [3,*]

[1] Borj El Amri aviation school,Tunisian Air Force ,Tunisia - (karimkeraani29, khalilmansour500@gmail.com)@gmail.com
[2] ESIP Gafsa, Tunisia - kalifiabillal@gmail.com
[3]EA Géoressources & Environnement, University Michel Montaigne / Bordeaux INP, France - nesrine.chehata@bordeaux-inp.fr

KEY WORDS: Crop mapping, agriculture, sentinel-2, time series, few shots, transformer, transfer learning, classification

ABSTRACT:

In this paper, we present an approach to land cover mapping from Sentinel-2 (S-2) satellite image time series using deep learning methods in the context of few shots in agricultural areas which aims to learn a classifier to recognize unseen classes during training with limited labelled examples. In many countries, there is a lack of Land Parcel Information Systems (LPIS) and thus of agricultural crop type annotations. Annotations are still based on fastidious digitization of parcels and in-field observations that are available in few numbers. Our idea is to transfer learning from pre-trained models on existing LPIS in France and apply them to a different geographical area in Kairouan in Central Tunisia. We build on work employing multi-headed self-attention mechanisms that have contributed to results that outperform other deep learning algorithms such as convolutional neural networks (CNNs), recurrent neural networks (RNNs) in agricultural context using S-2 Time series. We used two transformer-based deep learning models PSE-TAE (Pixel-Set Encoders + Temporal Self-Attention) and PSE-LTAE (Pixel-Set Encoders + Lightweight Temporal Self-Attention). We first studied their generalisation capacity in a few shot context and on different geographical study site. Then, by transferring the knowledge of these models and adapting them to the Tunisian context with the transfer learning techniques we have demonstrated experimentally that the adaptation of these methods is efficient for land cover mapping in agricultural areas with few in-field observations in terms of accuracy with an overall accuracy for both models reaching almost 93% for a detailed classification level with 17 classes.

1. INTRODUCTION

The development of remote sensing has been accompanied by the emergence of processing technologies that have allowed users to analyse satellite images with the help of increasingly automatic processing chains. The availability of so much data opens the door to many high-impact applications for machine learning methods. Among these is the classification of crop types, which is a major challenge for agricultural and environmental policy makers. In addition, machine learning techniques widely used in remote sensing have improved significantly.

The robustness of classical machine learning approaches was often limited by the amount of data available for the learning phase. New techniques have been developed to efficiently use this new large data stream. Recently, the gradual adoption of deep learning methods such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) for learning spatial and temporal attributes has led to significant improvements in classification performance.

In this paper, our work consists in exploiting the potential of existing annotated LPIS in France and efficient deep learning pre-trained models to produce crop type maps in Tunisia from Sentinel-2 satellite image time series (STIS) while having little training data i.e. in a few shot context. Few-shots classification aims to learn a classifier to recognize unseen classes during training with limited labelled examples.

The contribution of this project is to propose innovative solutions based on deep learning to automate producing land cover maps for agricultural landscapes with better accuracy, especially on cultivated areas with a fine nomenclature while reducing in-field observations that are time and cost-consuming. Moreover, this project is a key to a wide range of applications beyond crop monitoring, both for public and private entities, the large-scale monitoring of agricultural parcels is a matter of major political and economic importance and may provide a sustainable promising solution for land cover mapping in countries that lack geographical information systems and especially LPIS.

2. STATE OF THE ART

2.1 Land Cover Mapping

The most used approaches for land cover map production are supervised classifications (Gómez et al., 2016), which provide us with useful methods to have reproducible and automatically produced maps on a global scale. In response to the different expressed needs in terms of land cover and land use mapping (LCLU), developers and researchers have explored and developed a range of Machine Learning and Deep Learning algorithm families dealing with satellite images time series. Among the most performing families of algorithms (SVMs, Decision trees and Random Forests (RF), Neural networks (Pelletier, 2014), (Pelletier et al., 2016), (Bouaziz et al., 2017) and (Gressin, 2014) ...

2.2 Crop type mapping using Satellite Time Series with Random Forest algorithm and decision trees

In some works, binary decision trees have been used for Landsat time series mapping which have a lower average resolution than the Sentinel, and for the generation of MCD-12Q1 (MODIS LAND COVER) maps (Pelletier, 2014). In addition, another example describes the use of SPOT-4 (10 to 20 meters spatial resolution) and Landsat-8 (15 m spatial resolution) STIS to map land cover in southern France (Pelletier et al., 2016). The results showed the large amount of time

[*] Corresponding author

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIII-B3-2022-899-2022 | © Author(s) 2022. CC BY 4.0 License.

needed to build the decision tree, the sensitivity to noise in STIS and that they are not very efficient for both classical and high dimensional problems. In Tunisia, Random Forest classifier was used in agricultural areas for soil texture mapping over a 4-month time series with optical (S-2) and radar (S-1) Sentinel images in Kairouan, Tunisia (Bousbih et al.,2019). The results showed the robustness of RF and SVM for few shot training data.

2.3 Crop type mapping using Satellite Time Series with SVMs

SVM showed a high potential when applied on time series classification for improving the accuracy of land cover classification of Landsat data incorporating time series (MODIS NDVI data)(Gong et al.,2013), (Jia et al., 2014). In addition, in another study, three types of supervised classification were applied to Landsat 8 imagery to map land cover and land use around the Gulf of Gabes, Tunisia (Bouaziz et al., 2017). The results showed the ability of the method to handle large dimensionalities (Gressin, 2014).

2.4 Land cover mapping with deep learning

For the last five years, innovative deep learning methods based on neural networks have been used more and more in Earth observation.

Recently, CNNs have become the established approach for extracting spatial features from images. However, the scientific publications suggest that convolutions may not be as suitable for analysing high-resolution satellite images of agricultural plots such as Sentinel images (Sainte-Fare Garnot et al., 2020).

On the other hand, RNNs have been shown to be effective in encoding sequential information. However, RNNs process sequence elements in a successive manner, they lose long-term sequence memory, and require long learning times (Hélène and Dennis ,2019).

In a self-attention-based approach (PSE + LTAE model), the time input channels are distributed among several compact attention heads operating in parallel (Sainte-Fare Garnot et Landrieu, 2020). Each head extracts highly specialised temporal features which are in turn concatenated into a single representation. This PSE + LTAE approach outperforms other advanced time-series classification algorithms on an open-access satellite image dataset while using far fewer parameters and with reduced computational complexity (table 1)(Sainte-Fare Garnot et Landrieu, 2020).

Table 1. The evaluation of the compared models (Sainte-Fare Garnot et Landrieu, 2020)

MODEL	OA	mIoU
PSE+LTAE	94.2	51.7
PSE+TAE	94.3	50.9
CNN+GRU	93.8	48.1
CNN+tempCNN	93.3	47.5
Trasformer	92.2	42.8
ConvLSTM	92.5	42.1
Random Forest	91.6	32.5

This result was confirmed recently by different independent works using different datasets (Kondmann et al., 2021) and PSE-LTAE is used as a backbone by many works (Schneider and Körner, 2020).

3. METHODS

3.1 Proposed Workflow

Our processing chain as presented in figure 1 begins with the production of the LPIS on the study site. A digitization using very high resolution imagery of January 2021 is processed and then verified by sampling in field. Secondly, a 3-day terrain campaign was dedicated to crop type observations in April 2021.This collection of data was thus subsequently prepared in the appropriate formats. We first used the best architectures to train the models on purely Tunisian data. Then we tested the generalisation of pre-trained (PSE + TAE) and (PSE + LTAE) using French LPIS (RPG - Registre Parcellaire Graphique), on the locally collected data using a restricted nomenclature similar to the one used in pre-training.

The last step consisted in using available LPIS data in France (RPG), pre-trained models on RPG and local data to transfer learning to a context of few shot training in Tunisia. Models are pre-trained in France on thousands of RPG plots. Local data are introduced in the last layers of the architecture in order to predict a more detailed nomenclature, thus integrating new unknown classes for the pre-trained model, essentially the mixed crop classes by using the techniques of transfer learning techniques.

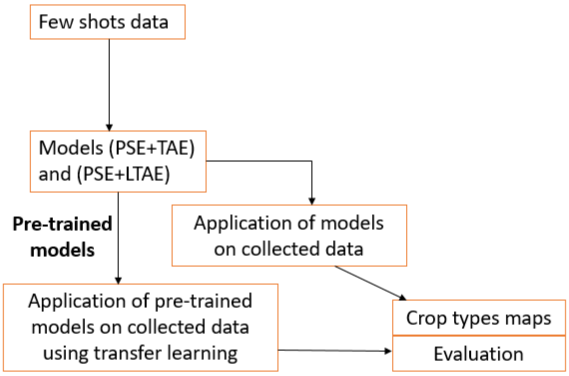

Few shots data
Models (PSE+TAE) and (PSE+LTAE)
Pre-trained models
Application of models on collected data
Application of pre-trained models on collected data using transfer learning
Crop types maps
Evaluation

Figure 1. The proposed workflow

3.2 Pixel Set Encoder and Transformer

In this work, we decided to follow (PSE + TAE) (Sainte-Fare Garnot et al., 2020) and (PSE+LTAE) (Sainte-Fare Garnot et Landrieu, 2020) approaches since they are well suited to classify satellite image time series and map land cover in agricultural environments while using far fewer parameters and with reduced computational complexity (**Figure 2**).

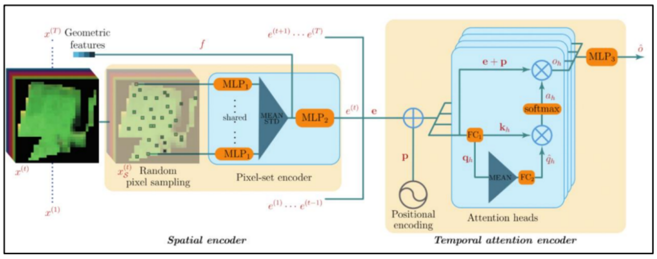


Figure 2. Architecture of the PSE + TAE approach (Sainte-Fare Garnot et al., 2020)

The spatial encoder is inspired by the point-set encoder PointNet and the DeepSet architecture commonly used for 3D point cloud processing. The motivation behind this design is that, instead of textural information (that is not relevant on S-2 imagery), the network computes learned statistical descriptors of the spectral distribution of the parcel's observations (Sainte-Fare Garnot et al., 2020).

The Temporal Attention Encoder is an attention-based network achieving equal or better performance than RNNs, while being completely parallelizable and thus faster.

In 2020, (Sainte-Fare Garnot et al., 2020) presented a new lightweight network for embedding sequences of observations such as satellite time-series. Thanks to a channel grouping strategy and the definition of the master query as a trainable parameter, his proposed approach is more compact and computationally efficient than other attention-based architectures. Evaluated on an open-access satellite dataset, the L-TAE performs better than state-of-the-art approaches, with significantly fewer parameters and a reduced computational load, opening the way for continent-scale automated analysis of Earth observation.

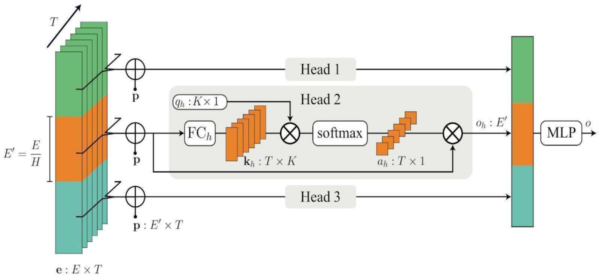


Figure 3. Architecture of the PSE + LTAE approach (Sainte-Fare Garnot et Landrieu, 2020)

In this study, results of both architectures will be compared in a few shots context to test their generalisation capacity.

3.3 Data augmentation

In a context of few shots training, data augmentation allows us to generate new training data from those already available, it is a relatively easy solution to implement, and the classification accuracies can be highly improved. Four augmentations were tested: The addition of Gaussian noise, the change of contrast and brightness, the scale change and finally rotation and translation.

4. STUDY SITE AND DATA

4.1 Study site: Kairouan Tunisia

The study area is located in the governorate of Kairouan which is located in central Tunisia and occupies a strategic position on the regional and national levels. The governorate of Kairouan covers 658 000 ha and is presented in the form of a wide corridor of plains which are limited to the west by mountainous areas and to the east by the depressions constituted by sebkhas. This natural environment is actually formed of physical sets quite contrasted (plains, hills, mountains) offering climatic nuances and different resources that necessarily generate specific uses and modes of occupation (Bouzaine and Lafforgue, 1986). The study area covers 197 km². The figure 4 describes the geographical extent of our study area.

The study area shows a high diversity of present crops and land use with irrigated and rain fed crops (about 16 classes are identified) and besides, it presents priority crops for Tunisia such as olive trees and wheat that are important to map and to monitor.

4.2 Dataset

We used a Sentinel-2 multispectral image sequence at level 2A, in canopy top reflectance. We excluded the atmospheric bands (bands 1, 9, and 10), allowing us to retain C = 10 spectral bands. The six 20m resolution bands are resampled to the maximum spatial resolution of 10m. The Area of Interest (AOI) corresponds to a single tile in the Sentinel-2 tile grid (T32SNE) in central Tunisia. This tile offers a difficult use case with a monitor wide variety of crop types and different terrain conditions. For our theme and depending on the availability of images we have chosen 24 dates from November 2020 to April 2011.

Parcel plots were digitised manually using a high resolution spatial imagery leading to a geo-referenced LPIS with class label for each parcel (figure 4). We have digitised 10111 parcels, covering an area of 197 km². Topological validation was processed. The parcel delimitations were checked in a field mission.

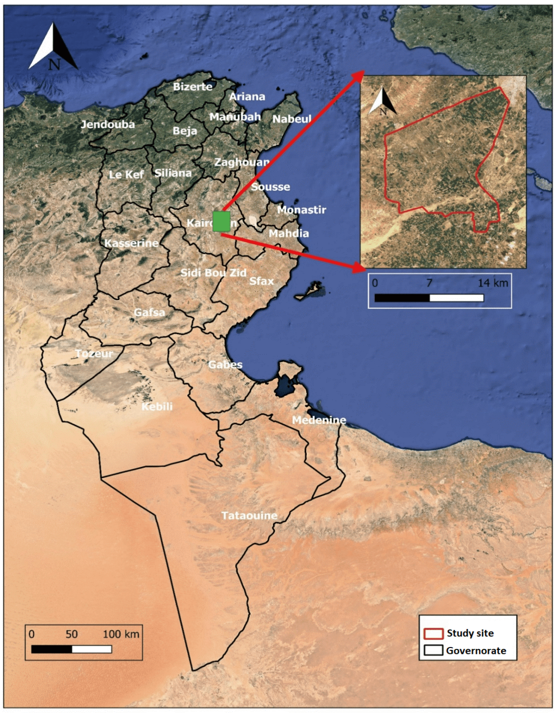


Figure 4. Extent of the study area - Kairouan - Central Tunisia

Besides, we got our ground truth parcels collected from a field mission in Kairouan where finally, we were able to observe the crop types of about 1516 plots located in different areas to ensure the variability of observations and cover the entire study area (figures 6, 7).

A detailed nomenclature of 17 crop type classes was used. The most observed class of parcels is the olive tree with 411 parcels and wheat with 180 parcels. On the other hand there were very minor classes such as vines and prickly pear which leads to a highly imbalanced dataset as shown in the figure 8.Satellite images are cropped using digitised polygons to constitute the image time series.

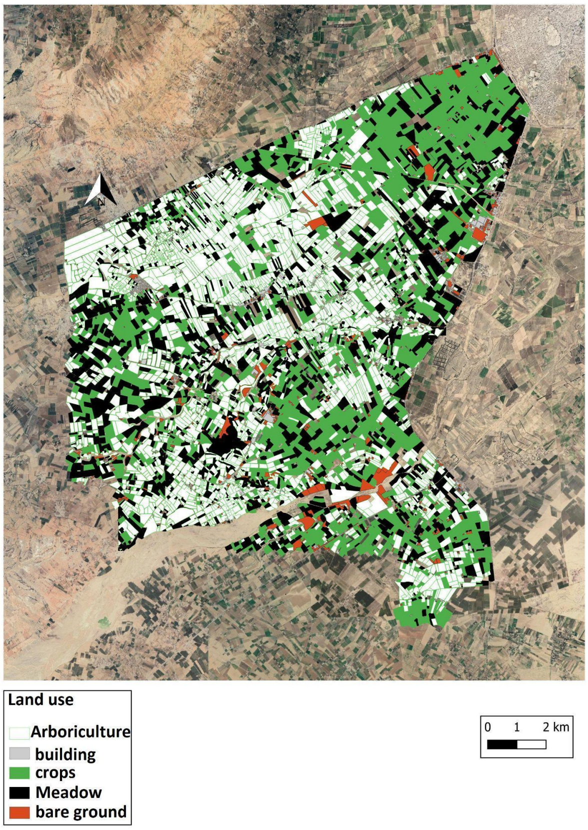


Figure 5. Digitized parcels and photo-interpreted land cover maps using a very high resolution image

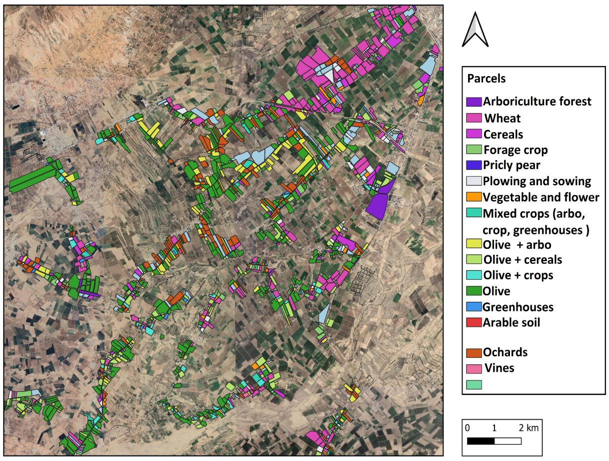


Figure 6. Spatial distribution of observed parcels in field mission - Kairouan site

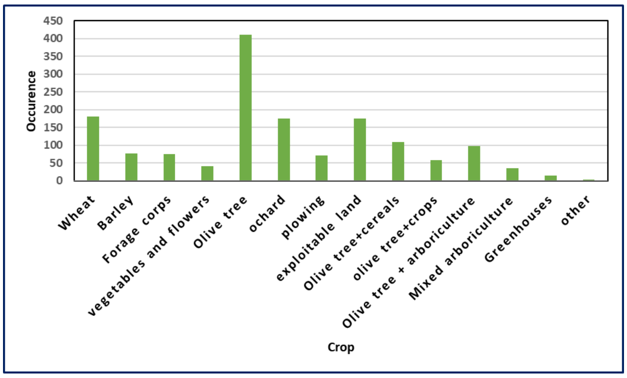


Figure 7. Distribution of observed plots per crop type

The pixels in each parcel are stored in an arbitrary order in an array of size T × C × N with N being the total number of pixels in a given parcel, T being the number of temporal observations, and C being the number of spectral bands because this format does not lose or create any information, regardless of the size of the parcel. Each of these arrays must be stored separately in a numpy file 'unique_id_of_the_sample.npy'. All individual (.npy) files are stored in the same DATA subdirectory. The crop metadata (nomenclatures, dates, geometric characteristics) are generated in the form of Json files to be used in the implementation of the models.

5. RESULTS

5.1 Experiments

We used cloud computing tools to address the concern of large data volumes, and to implement our models we used the Google Colab platform, which is designed for training models in machine learning. In addition, Colab provides us with the opportunity to train and test with a GPU that can be 60 times faster than a classical learning on a CPU. Indeed Colab offers us a Tesla T4 graphics card with 16 GB of RAM. In addition, we used the following tools and libraries: ArcGIS, QGIS, Pytorch, Cuda, Envi, Rasterio.

For data augmentation, we used an efficient Python library *Imgaug* (imgaug Development Team, 2020). Several parameters and criteria were tested in order to choose the most efficient and effective ones.We present below an explanatory table of the models architecture, their modules, hyperparameters and the number of parameters (table 3 and table 2).

Table 2. Configurations of (PSE + TAE)

Modules	Hyperparameters	Number of parameters
	PSE	
S	64	19 936
MPL_1	10 → 32 → 64	
MPL_2	68 → 128	
	TAE	
d_e , d_h , H	128, 32, 4	136 192
FC_1	128 → (32 * 2)	
FC_2	32 → 32	

MLP_3	256 → 128	
Decoder		
MLP_4	128 → 64 →32 → 20	11 180
Total 164116		

For the optimizer, we use Adam (Kingma et al., 2015) with the following values:

- **Lr** = 10^{-3};
- **β** = (0.9, 0,999).

Table 3. Configurations of (PSE + L-TAE)

Modules	Hyperparameters	Number of parameters
PSE		
S	64	19 936
MPL_1	10 → 32→ 64	
MPL_2	132 → 128	
TAE		
d_e , d_h , H	128, 8, 16	116 480
MLP_3	256→ 128	
Decoder		
MLP_4	128 → 64 → 32 → 20	11 180
Total 147 604		

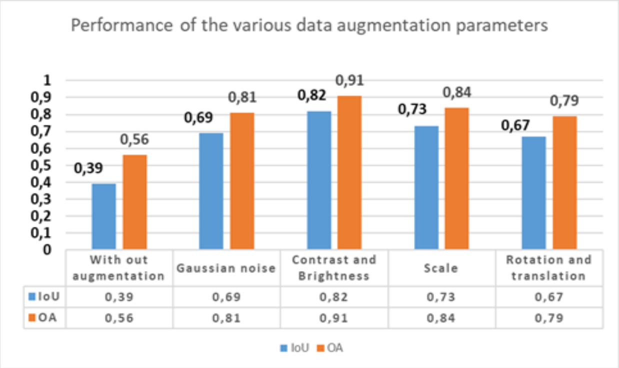
In addition, we train the models with the focal loss (γ = 1) to calculate the losses. We applied the "K-fold" cross-validation with K=5. For each file, the dataset is divided into training, validation and test sets with a ratio of 3:1:1 respectively. The validation step allows us to select the best performing epoch and evaluate it on the test set.

5.2 Data augmentation

We experimented the data augmentation with several parameters and criteria to choose the most efficient and effective ones. The techniques used are:

- **The addition of Gaussian noise**: the addition of noise increases the size of the training data set. Random noise is added to the input variables, making them different each time they are exposed to the model. In this way, adding noise to the input samples is a simple form of data augmentation.
- **Change of contrast and brightness**: For each image, we add a small amount of contrast and brightness (according to 4 intensity levels).
- **Scale change**: We randomly change the scale of the images between 50% and 150%.
- **Rotation and translation:** A random rotation of ±25◦ ensures that this transformation does not change the characteristics of the classes. In addition, a translation was established by moving all images along the x and y axes by 4 pixels.

We mention that the data techniques were tested in a cumulative way starting with the addition of Gaussian noise, to which we added the change of contrast and brightness, then the scale change and finally the rotation and translation augmentation added to all previous ones. We finally obtained the results presented in the figure 8 when applying pre-trained PSE+LTAE on local data.



	With out augmentation	Gaussian noise	Contrast and Brightness	Scale	Rotation and translation
IoU	0,39	0,69	0,82	0,73	0,67
OA	0,56	0,81	0,91	0,84	0,79

multiplied by a factor of 3leading to 4548 samples.

5.3 Study of the sensitivity of the parameters

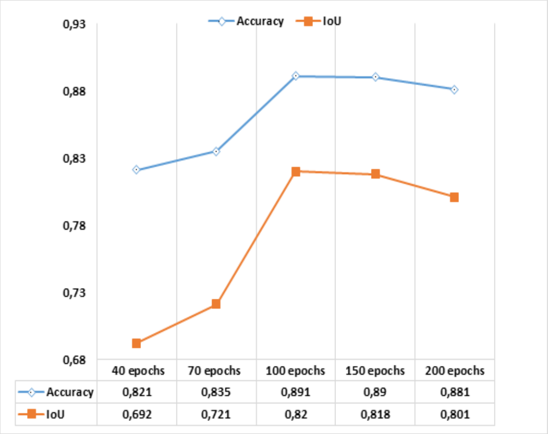
This phase consists in changing configuration parameters of the model on order to make a sensitivity study of their impact on performances and to optimize them. We studied the impact of number of epochs, the batch size, the number of pixels to be sampled in PSE and the number of workers.

5.3.1 Sensitivity of the number of epochs on the performance of the model

The number of epochs is defined by the number of epochs (i.e. passages on the whole set of samples of the learning base) during the gradient descent. Figure 9 illustrates the impact of varying the number of epochs on model performance while keeping other parameters at constant values.

5.3.2 Sensitivity of the batch size on the model performance

The Batch size is the number of samples used to estimate the gradient of the cost function. A batch size of 128 means that 128 samples of the training data set will be used to estimate the error gradient before the model weights are updated. Different batch sizes were tested and their impacts on model performances are shown in Figure 10. The other parameters are kept at constant values.



	40 epochs	70 epochs	100 epochs	150 epochs	200 epochs
Accuracy	0,821	0,835	0,891	0,89	0,881
IoU	0,692	0,721	0,82	0,818	0,801

Figure 9. Sensitivity of the number of epochs on the performance of the model

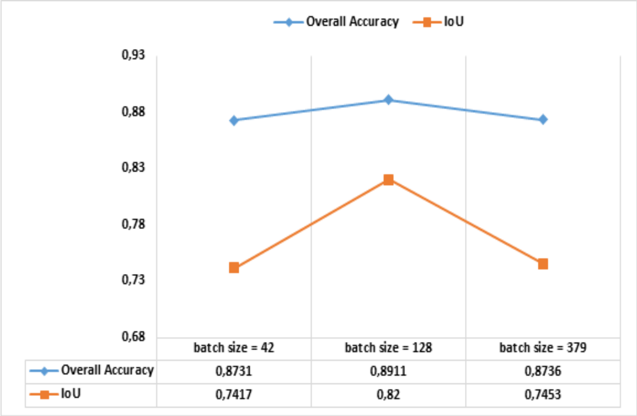


Figure 10. Sensitivity of the batch size on the performance of the model

5.3.3 Sensitivity of the number of pixels to be sampled in PSE on the model performance

The number of pixels to sample S is randomly drawn from the N pixels in the plot. When the total number of pixels in the image is less than S, an arbitrary pixel is repeated to match this fixed size (number of pixels to sample). The same set S is used for sampling all T acquisitions of a given plot. In this section, many simulations were performed by varying the number of pixels to be sampled. The other parameters are kept at constant values. The impact on model performance is illustrated in Figure 11.

5.3.4 Sensitivity of the number of workers on the model performance

The number of workers corresponds to the number of sub-processes to use to load the data. Different simulations have been performed by varying the number of workers while other parameters are kept at constant values. Results are shown in Figure 12.

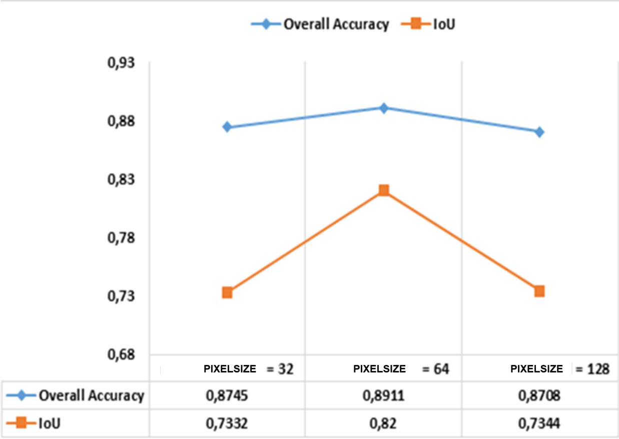


Figure 11. Sensitivity of the number of pixels to be sampled on the model performance

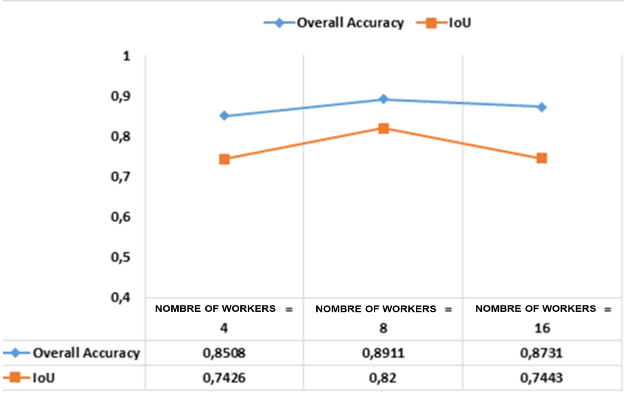


Figure 22. Sensitivity of the number of workers on the model performance

5.4 Classification results

The sensitivity study of parameters allows us to measure the impact of various parameters on model performance and thus to optimize them.

The optimized parameters are resumed in Table 4. Results are presented using best data augmentation configuration.

Table 4. Optimized model parameters

precision metric	Batch size	Number of epochs	Number of co-workers	Number of pixels to sample
Value	100	128	64	8

5.4.1 Application of Transformer models on local few shots

At this step, augmented data is used to train the models on purely local data, acquired on the Tunisian study site.

Table 5. Results of (PSE+LTAE) and (PSE+TAE) classification on local data

Model	**Micro IoU**	**F_ score**	**OA**	**Recall**	**Precision**
PSE+LTAE	**0.77**	0.84	**0.87**	0.82	**0.87**
PSE+TAE	0.75	**0.85**	0.86	**0.85**	0.86

5.4.2 Generalization of pre-trained Transformer models on local few shots

This step consists in applying the models (PSE+LTAE) and (PSE+TAE) pre-trained on the RPG to predict the common classes between the French dataset and the Tunisian study area. Consequently, the weights of the models pre-trained on 200 000 plots of the RPG data are used to initialize our model. Results are presented in Table 6.

Table 6. Results of pre-trained models (PSE+LTAE) and (PSE+TAE) on local data

Model	**Micro IoU**	**F_ score**	**OA**	**Recall**	**Precision**
PSE+LTAE	**0.82**	**0.89**	**0.91**	**0.89**	**0.90**
PSE+TAE	0.78	0.87	0.87	0.87	0.86

https://doi.org/10.5194/isprs-archives-XLIII-B3-2022-899-2022 |

5.4.3 Transfer learning with pre-trained transformer models

It should be noted that, we have a reduced dataset (4548 plots with data augmentation) and different from the French dataset: since we have 17 classes in Tunisia with a more detailed annotation than that of RPG. Based on these two criteria (reduced and different dataset), we need to keep (freeze) only the layers of "feature extractions" (Sainte-Fare Garnot et Landrieu, 2020) which are in our case the layers of PSE.We finally obtained the results presented in the tables below.

Table 7. Results of transfer learning using pre-trained models (PSE+LTAE) and (PSE+TAE)

Model	Micro IoU	F_ score	OA	Recall	Precision
PSE+LTAE	**0.90**	**0.93**	**0.93**	**0.93**	**0.94**
PSE+TAE	0.84	0.9	0.92	0.91	0.91

6. DISCUSSION

6.1 Data augmentation

We notice that scale and rotation translation augmentations do not improve the classification accuracy but decrease it. Indeed these transformations do not change the parcel content and lead to redundant information that is not relevant for the classifiers. In the contrary, the addition of Gaussian Noise increased the OA by 25.8 %. When adding the contrast and brightness changes, the OA is increased by 8.4% reaching 0.9.

6.2 Study of the sensitivity of the parameters

6.2.1 Sensitivity of the number of epochs on the performance of the model

We can see from figure 9 that the number of epochs has a great influence on the performance of the model. Indeed, this parameter should not be too low to avoid the lack of learning and should not be very high to avoid the phenomenon of over-learning. The best performances were obtained with a number of epochs of 100 with a performance reaching almost 0.90 of global accuracy and 0.82 of IoU.

6.2.2 Sensitivity of the batch size on the model performance

We can see from figure 10 that the batch size is an important parameter that influences the dynamics of the learning algorithm. Indeed, it this parameter is too low, the weights of our network can jump around and it can be unable to learn or converge very slowly. Moreover, the batch size should not be very high since it reduces the stochasticity of the gradient descent and may decrease the accuracy of the model during training. The best performances were obtained with a Batch size of 128.

6.2.3 Sensitivity of the number of pixels to be sampled in PSE on the model performance

It can be seen from figure 11 that the number of pixels to be sampled in PSE has a great influence on the performance of the model. The best performances were obtained with a number of pixels of 64 with a performance reaching almost 0.90 of overall accuracy and 0.82 of IoU. Indeed, this parameter depends on the number of pixels of the plots N, So the closer the number of pixels for the observed plots is to S value the better the performance of the model.

6.2.4 Sensitivity of the number of workers on the model performance

The best performances were obtained with a number of workers of 8 with a performance reaching almost 0.90 of global precision and 0.82 of IoU. On the other hand, the global accuracy is lower for the number of workers 16 and 4 (0.85 and 0.87 of global accuracy). Indeed, the performance of the model in relation with the number of workers depends on the occupation of the processor cores for other tasks, the speed of the processor and the speed of the hard disk, ...

No improvement took place when num_workers exceeded the number of CPU cores. Many tips indicate that the number of workers should be twice the number of CPU cores, because it is necessary to leave some CPUs free for other tasks (rather than using them 100% for data loading) allowing the user to concentrate on the most important tasks.

6.3 Application of Transformer models on local few shots

The application of both models lead to similar results with an overall accuracy of 0.87 ± 0.01 and a kappa of 0.84± 0.01. The Recall and accuracy are around 0.85 ± 0.03 which is related to the low rate of false positive classes. That is to say that about 84% of the plots labelled by the model as class i are really so and that the capacity of our model to detect correctly all types of classes is high.

6.4 Generalisation of Transformer models on local few shots

The results show that in terms of accuracy, both methods lead to good results with an overall accuracy up to nearly 87% ± 0.03. Both models (PSE + TAE) and (PSE + LTAE), generalise in a good way to local Tunisian data, which is probably explained by the use of the same Sentinel-2 sensor, and the similarity of some crop types between Kairouan and Southern France. Besides, the lightweight model (PSE+ LTAE) lead to better classification accuracy +3% while minimising model parameters and reducing computing time.

6.5 Transfer learning with pre-trained transformer models

The above results show the contribution of Transfer Learning in a few shot context. We conclude that TL improves our results on average for the most of the accuracy metrics. Effectively, we've got an improvement in terms of overall accuracy and IoU with almost 3% and 8%, respectively.

7. CONCLUSION

In this paper, we focused on crop mapping in a few shot context in a developing country, Tunisia. We have used two state-of-the-art alternative approaches in which convolutional layers are advantageously replaced by encoders operating on unordered sets of pixels (PSE) in order to exploit the generally coarse resolution of publicly available satellite images Sentinel 2. We used neural architectures based on self-attention rather than recurrent networks (Temporal Self-Attention TAE and Lightweight Temporal Self-Attention LTAE). We experimentally demonstrate that these methods are efficient for land cover mapping in agricultural areas in a few shot context with an overall accuracy for both models reaching almost 87% for a detailed classification level with 17 classes. We showed that pre-trained models on French LPIS generalise well on a different geographical site, namely Tunisia improving accuracies by 4%. Finally best results are obtained using the transfer learning (TL) approach by exploiting the potential of existing LPIS data in France and pre-trained models adapted to local few shots training. Thanks to the transferred knowledge, the proposed models become able to successfully classify the studied crop types with a higher overall accuracy for both considered models (PSE+LTAE and PSE+TAE), reaching up to almost 93% (+6% improvement). These results are very promising and open perspectives for producing land cover maps at the national scale and for other African countries that suffer from the lack of geographical data and especially LPIS.

REFERENCES

Bouaziz, M., Eisold, S., and Guermazi, E., 2017: *Semi-automatic approach for land cover classification: a remote sensing study for arid climate in south eastern Tunisia.* Euro-Mediterranean Journal for Environmental Integration. 2(1): p. 1-7.

Bousbih, S., Zribi, M., Pelletier, C., Gorrab, A., Lili-Chabaane, Z., Baghdadi, N., Ben Aissa, N., Mougenot, B, 2019: *Soil Texture Estimation Using Radar and Optical Data from Sentinel-1 and Sentinel-2. Remote Sensing*, *11*, 1520. https://doi.org/10.3390/rs11131520

Bouzaine, S. and Lafforgue, A., 1986: *Monographie des oueds Zéroud et Merguellil*. Tunis, 1037.

Gómez, C., White, J.C., and Wulder M.A., 2016:*Optical remotely sensed time series data for land cover classification: A review*. ISPRS Journal of Photogrammetry and Remote Sensing. 116: p. 55-72.

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S. and Li, C., 2013. Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing*, *34*(7), pp.2607-2654.

Gressin, A., 2014: *Mise à jour d'une base de données d'occupation du sol à grande échelle en milieux naturels à partir d'une image satellite THR*. Paris 5.

Hélène,A.,Denis, B., 2019: *Intelligence artificielle - État de l'art et perspectives pour la France*: Martine Automne, Nicole Merle-Lamoot.

imgaug Development Team
https://imgaug.readthedocs.io/en/latest, access on 17th January 2022

Jia, K., Liang, S., Wei, X., Yao, Y., Su, Y., Jiang, B. ,Wang, X., 2014. Land cover classification of Landsat data with phenological features extracted from time series MODIS NDVI data. *Remote sensing*, *6*(11), pp.11518-11532.

Kingma, D. and Adam, J. Ba, 2014: *ADAM: A Method for Stochastic Optimization*. 3rd International Conference for Learning Representations, San Diego, 2015.

Kondmann, L., Toker, A., Rußwurm, M., Camero, A., Peressuti, D., Milcinski, G., Mathieu, P.P., Longépé, N., Davis, T., Marchisio, G. and Leal-Taixé, L., 2021, August. DENETHOR: The DynamicEarthNET dataset for Harmonized, inter-Operable, analysis-Ready, daily crop monitoring from space. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.

Pelletier, C., 2014: *Cartographie de l'occupation des sols à partir de séries temporelles d'images satellitaires à hautes résolutions: identification et traitement des données mal étiquetées.* Université de Toulouse, Université Toulouse III-Paul Sabatier.

Pelletier, C., Valero, S., Inglada, J., Champion, N., Dedieu, G. , 2016. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment*, *187*, 156-168.

Sainte-Fare Garnot, V. and Landrieu, L., 2020: *Lightweight Temporal Self-Attention for Classifying Satellite Images Time Series*. In International Workshop on Advanced Analytics and Learning on Temporal Data. Springer.

Sainte-Fare Garnot, V., Landrieu, L., Giordano, S. and Chehata, N., 2020: *Satellite image time series classification with pixel-set encoders and temporal self-attention* in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Schneider, M. and Körner, M., 2020: *Satellite Image Time Series Classification with Pixel-Set Encoders and Temporal Self-Attention*. ML Reproducibility Challenge 2020.