Towards Improved Crop Type Classification: a Compact Embedding Approach Suitable for Small Fields

Madeline C. Lisaius[1,*], Andrew Blake[2], Clement Atzberger[3], Srinivasan Keshav[1]

[1] Department of Computer Science and Technology, University of Cambridge, Cambridge, United Kingdom - mcl66@cam.ac.uk
[2] Clare Hall, University of Cambridge, Cambridge, United Kingdom
[3] Cyclops.ai, New York, United States

Keywords: Foundation Model, Self-Supervised Learning, Remote Sensing, Crop Classification, Agriculture

Abstract

Satellite -based crop classification and maps are important tools for food security and climate change mitigation, but existing approaches are not effective for small field systems. To address this, crop type classification using embeddings generated by a global foundation model, TESSERA, are compared to standard classification approaches in the literature. We find that our embedding -based approach offers a triple win: 1) consistent and statistically significant performance improvement over current methods, 2) greater simplicity due to the elimination of feature engineering, and 3) the reduction of computational cost. Our embedding -based approach achieves significantly higher F1 scores in the classification of 5 of 7 crop types for small fields in Austria (over 10% improvement in one case). Additionally, the TESSERA embedding -based method uses 8% of compute compared to the raw data method. These results indicate that embeddings are an effective approach for crop type classification tasks in small field systems.

1. Introduction

Satellite remote sensing using publicly available sensors such as Sentinel-1 and Sentinel-2 allows low-cost, global land surface monitoring of crop types. This monitoring and mapping can inform important agricultural decisions that are closely tied to global food security such as crop yield estimates, management characterizations, crop area measurements, and climate impact estimation (Weiss et al., 2020).

However, remote sensing for crop classification has been underdeveloped for small field contexts, which disproportionately impacts smallholder and subsistence agricultural systems, due to two challenges. First, small fields include a high proportion of border pixels, which are often mixed pixels and reflection different management practices compared to the field center (Maurya et al., 2021). Many machine learning methods focusing on agricultural remote sensing are designed using pixels from field interiors, making them a poor fit for small fields (Yuan et al., 2022). Second, there is a lack of publicly available small field crop classification datasets with well-defined field boundaries, making development of appropriate methods challenging (Wu et al., 2023).

Current approaches for classifying small field crops largely come from literature on smallholders in the Global South and are (i) pixel-wise classical methods such as Random Forest (Ayman et al., 2024) (Lebourgeois et al., 2017) and Support Vector Machine (SVM) (Rao et al., 2021) (Ayman et al., 2024), and (ii) spatially informed methods (Xie et al., 2019) (Pribadi et al., 2023) (Zheng et al., 2024) although patch -based methods have been found in some cases to be less effective than pixel-wise methods (Maolan et al., 2024). Recent advances in deep learning focus on spatial-spectral-temporal methods (Rustowicz et al., 2019) (Turkoglu et al., 2021) (Mohammadi et al., 2023) and rarely test in small field environments. Furthermore, where small field environments are tested, proposed methods do not consistently beat baselines (Rußwurm et al., 2023). This work will focus on making methodological comparisons to other approaches working on small field environments at the pixel level. In lieu of access to a bench mark small field dataset, we explore this using small fields in Austria.

Self-supervised learning (SSL) embedding -based approaches have shown potential for crop type classification due to their ability to outperform current baseline methods (Lisaius et al., 2024) (Yuan and Lin, 2020). SSL approaches extract meaningful and compact embeddings of unlabeled input data, in this case of satellite remote sensing time series, which can be utilized for downstream tasks. While SSL embedding -based approaches are increasingly being used in the remote sensing context, to the best of our knowledge, none have been applied to the task of crop classification in small fields. Foundation Models (FMs), which are built on SSL, have similarly shown promise for crop type classification, but none appear to have been leveraged for small field agriculture (Lu et al., 2025).

To address the potential of embeddings for small field crop classification, we show an approach to test small fields, characterize the challenges of edge pixels, and demonstrate the efficacy of our embedding approach to crop type classification in these contexts.

2. Related Work

There is considerable recent literature related to pixel-wise small field crop classification using public remote sensing spectral time series data. Of the existing work, we present the three closest examples:

Ibrahim et al. (Ibrahim et al., 2021) use Sentinel-1, Sentinel-2 and SkySat high resolution imagery to create a wall-to-wall classification map for the small field interest region in Nigeria consisting of five crop classes and average field size ranging between 0.3 ha and 0.5 ha. They use seasonally informed spectral temporal metrics (STMs) from Sentinel-1 and Sentinel-2

* Corresponding author

sensors to first classify agricultural land from other land covers with RF, and SkySat to verify. They then classify crop types from identified fields using only Sentinel-2 -based STMs, also with RF. However, this work requires extensive crop-specific feature engineering to prepare data for use by RF, requiring customization to adapt to new regions.

Soler et al. (Soler et al., 2024) curate street view images of small fields and label them to create a set of ground references. From a Sentinel-2 time series they take Red-edge 4, both SWIR bands, NIR, and additionally add a spectral vegetation index (VI), and use harmonic regression to extract frequency domain features. These features are used with the cultivated label set to classify all of Thailand. This work benefits from the extensive collection of labels and is not possible in regions with limited street view imagery or inaccessible roads across the small plot areas. Not all fields are small fields or are explicitly part of smallholder regions of Thailand.

Ren et al. (Ren et al., 2022) classify 5 crop classes - one small field region with 2 classes and one small field region with 3 classes - in Mongolia using Sentinel-1 and Sentinel-2 data. They report that more than 50% of the fields in the study area are smaller than 2.56 ha. For both optical and SAR data, a single annual composite is generated of the selected bands and five vegetation indices (VIs). Classification is performed with Random Forest after a croplands mask has been applied. This study is limited by the extensive annual compositing approach and custom feature engineering.

Outside of small field work, Lisaius et al. (Lisaius et al., 2024) introduce Spectral–Temporal Barlow Twins (ST-BT), a novel SSL method for pixel-wise embedding generation using Sentinel-2 time series. Embeddings are a data representation where raw inputs are projected into a multidimensional space by a pretrained architecture, and can be leveraged with few labels for various downstream tasks. Designed to be invariant to cloud cover and limited labeled data, ST-BT embeddings achieves high classification accuracy on two downstream crop type classification tasks and outperforms comparable best-in-class approaches. This method shows potential under real-world data constraints but is not evaluated for small field or field-edge conditions and requires bespoke training for the region of interest.

Our review of prior work suggests that embedding -based approaches to crop type classification are yet to be leveraged in this small field setting.

3. Methods and Data

3.1 TESSERA

In this work, we propose the use of FM embeddings from TESSERA (Feng et al., 2025). This approach is based on ST-BT (Lisaius et al., 2024) and generates embeddings from Sentinel-2 multi-spectral and Sentinel-1 synthetic aperture radar (SAR) time series. Across multiple tasks including crop type classification, TESSERA embedding -based methods outperform other FMs and baselines, showing promise for small field crop classification. In particular, the augmentation strategy involves encouraging the model to learn the same embedding from two distinct sparse samples of a time series, which builds invariance to temporal variations. Implementation and downstream evaluation performance details can be found in the full paper (Feng et al., 2025). We hypothesize that this may contribute to a more stable crop type signature at boundary pixels compared to raw data inputs because the embedding is encouraged to capture intrinsic and continuous characteristics such as crop biophysical evolution rather than stochastic spectral fluctuations caused by mixed-pixel effects or phenological asynchrony at field edges.

Annual TESSERA embeddings, encoding data from from January to December for each year, are of 128 dimensions and are downloaded using the GeoTessera Python Package (**?**). A sample visualization is shown in Figure 1. As these embeddings are pre-generated, freely available, and appropriate for many use cases, we do not count the pretraining of TESSERA or generation of embeddings in the compute and complexity comparisons although we acknowledge that these steps incur meaningful computational costs.

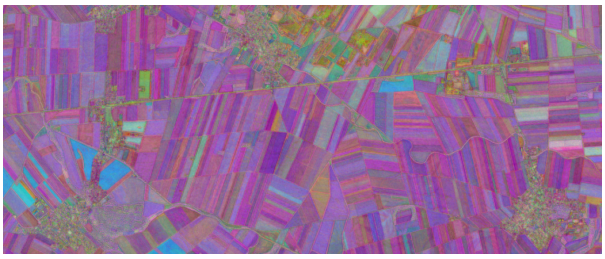

Figure 1. An example area in the Austria region is visualized using the first three dimensions of the TESSERA embeddings as RGB. Fields and other features are distinguishable in the unlabeled embeddings.

We additionally compare the TESSERA embeddings to embeddings from AlphaEarth (Brown et al., 2025), which offers the same pre-generated embedding paradigm. AlphaEarth is a foundation model that takes in optical, radar, LiDAR, environmental, and annotated text in training. The model produces 64 -dimension embeddings and reports best-in-class performance across various downstream tasks when compared against existing foundation models. Pre-generated embeddings are freely available via Google Earth Engine.

3.2 Austria dataset

Labeled field polygons come from the INVEKOS Austria 2022 dataset and include 43,738 fields representing 154 classes over 1850 km^2 to the East of Vienna (INVEKOS Schläge Österreich 2022 - Open Government Data Austria, 2022). This dataset is also used in the documented TESSERA downstream tasks (Feng et al., 2025) but is grouped differently by phenology. In this work, for example, the grouped class ‘maize’ encompasses ‘maize’, ‘seed corn’, ‘silo corn’ and ‘sweet corn’. Classes that correspond to non-vegetation such as rock walls, and classes with less than 50 fields observed are discarded. Remaining crops that represent a small subgroup such as grapes and mustard family are grouped together in ‘other crops’. The resulting classes and field counts are shown in Table 1.

3.3 Remote sensing dataset

We prepare Sentinel-1 and -2 time series data which are referred to as “raw” to compare to TESSERA embeddings. We collected 83 Sentinel-1 and 99 Sentinel-2 timesteps from 1 January 2022 to 1 October 2022, broadly corresponding to the growing season. We find that the Sentinel-2 images are well-aligned to our crop boundaries and consistent across the time series, and therefore did not employ georeferencing correction. From Sentinel

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-873-2026 | © Author(s) 2026. CC BY 4.0 License.

Crop	Field Count
Winter grain	10,872
Other crop	9,986
Maize	2,702
Soy	2,053
Legume	1,319
Potato	598
Squash	551
TOTAL	**28,081**

Table 1. Number of fields per crop class, which is dominated by the winter grain class.

2, band numbers 2, 3, 4, 5, 6, 7, 8, 8a, 11, and 12 are used. All bands are normalized and resampled to 10 m spatial resolution. The provided QAI band is used for cloud masking. We compute five common optical and SAR VIs: Normalized Difference Vegetation Index (NDVI), Green Chlorophyll Vegetation Index (GCVI), Enhanced Vegetation Index (EVI), Land Surface Water Index (LSWI) and the Sentinel-1 SAR -based Radar Vegetation Index (RVI).

Additionally, we adopt the spectral-temporal metrics (STMs) approach as described in Ibrahim et al. because of their robust approach to metric creation and documentation for a small field region (Ibrahim et al., 2021). We generate pixel -based Sentinel-2 metrics for four time windows: (i) all available dates, (ii) the beginning of season April - May, (iii) the peak season June - July and (iv) end of season August - September. For each STM we produce: minimum, 25% quantile, 50% quantile, 75% quantile, maximum, average, standard deviation, range, inter-quartile range, skewness, kurtosis, NDVI, EVI, Normalized Difference Water Index (NDWI) and the Tasseled Cap transformations of wetness (TCW), greenness (TCG), and brightness (TCB) using Tasseled Cap values for Sentinel-2 from Jiang et al. (Jiang et al., 2024). We use these Sentinel-2 STMs alone similar to Ibrahim et al. and also with the Sentinel-1 time series and RVI.

As part of our understanding of border pixel and near-border dynamics of small fields we consider the distance of pixels to the field border as a proxy to understand behavior of these difficult pixels. Past work has characterized spectral mixing in pixels (Lobell and Asner, 2004), common at borders, but this was done with simulated data which will not capture real world management heterogeneity at field borders. To create distance labels, we use the Python Rasterio package to rasterize the field polygons, identify edge pixels, and create bands of 10 m width moving into the interior. Pixels that intersect a polygon border including those that are partially outside the polygon are considered ‘0 m’, pixels fully inside the polygon and intersecting the 10 m distance band are considered ‘10 m’, and each distance band is sequentially considered (20 m, 30 m, etc.). Pixels are labeled with the first distance band they intersect as shown in 2. While distance to border is a 1-dimensional proxy to a 2-dimensional field size dynamics, distance to border will relate to pixel characteristics regardless of field shape and therefore is a good experimental base.

3.4 Evaluation Details

Due to the lack of publicly available benchmark datasets focused specifically on small fields and with well delineated field boundaries, we define two proxy approaches for evaluating embeddings.

First, we establish a subset of small fields in the Austrian dataset, defined as having pixels with distance-from-border labels

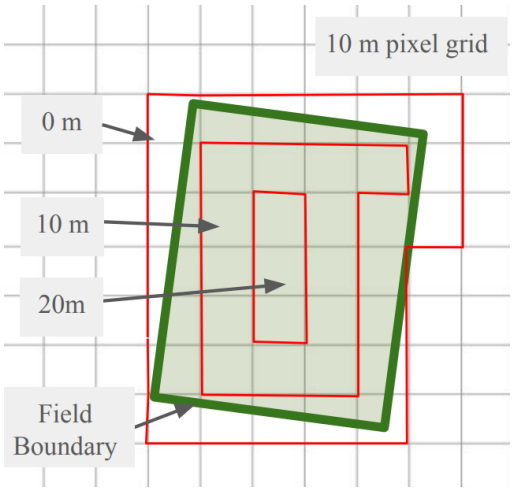

Figure 2. Visualization of how pixel distance labels are defined.

of 20 meters or less. This translates to fields with a maximum width or height of 50 meters (noting that the second dimension can be larger) - usually these are long and thin geometries. When we filter fields to this size, we find that, by class, the average field size varies from 0.5 hectares (ha) to 1 ha as shown in Table 2, and comprise about half of available fields. This size is comparable to the size of small fields in smallholder regions; globally, smallholder farmers typically manage plots under 2 ha (Lowder et al., 2016). This approach provides a reproducible approach for testing models for small fields.

Second, we characterize classification performance as a function of distance to the field boundary. This allows us to capture key challenges of small fields, where a large proportion of pixels lie near field edges, by analyzing classification variation and accuracy as a function of distance to field border.

Table 2. Austrian small field statistics.

Crop	No. Fields	Average Area (m²)	Average Area (ha)
Other Crops	6462	4052	0.40
Winter grain	4414	9821	0.98
Maize	1073	11146	1.11
Legume	753	7691	0.77
Soy	691	12366	1.24
Squash	236	8140	0.81
Potato	234	10878	1.88

3.5 Notation

Table 3 sets out the notation that will be used to evaluate the stability of crop type classification at field boundaries using standard deviation and absolute difference.

Table 3. Description of notation used.

Symbol	Description
F	Set of fields; each field is a set of pixels
f	Field ID
i	Index of pixel in field
B	Number of spectral bands
b	Spectral band index, $0 \leq b \leq B$
$d(f,i)$	Distance label (see sec. 3.3) for ith pixel in field f
j	Distance group index, from o–10 towards center
t	Number of time steps
$x^t_{f,i,b}$	Sensed value of field f, pixel i, time t, band b

The mean time series value $\bar{x}$ for band b of each pixel i and for each field f is defined as:

$$\bar{x}_{f,i,b} \triangleq \frac{1}{T}\sum_{t=1}^{T} x^{t}_{f,i,b} \tag{1}$$

3.5.1 Standard deviation The standard deviation across all values in a field f for distance group j and by band b is:

$$\sigma_{f,b,j} \triangleq s.d.\,\{\bar{x}_{f,i,b} \mid d(f,i) = j\} \tag{2}$$

We consider the average standard deviation, $\bar{\sigma}$, as a metric of the stability of the data within a group j and band b:

$$\bar{\sigma}_{b,j} \triangleq \sum_{f \in F} \sigma_{f,b,j} \tag{3}$$

A group that is homogeneous, for example at the center of a field, should be stable and as consequence have low average standard deviation.

3.5.2 Absolute difference of inputs We define groups g of mean band values $\bar{x}$ for field f, for pixels i and across all bands b as:

$$g_{f,b,j} \triangleq \{\bar{x}_{f,i,b} \mid d(f,i) = j\} \tag{4}$$

And the mean of each group $g_{f,i,b,j}$ as:

$$\bar{g}_{f,b,j} \triangleq \sum_{i} \frac{g_{f,i,b,j}}{|g|} \tag{5}$$

The absolute difference a of each group mean $\bar{g}_{f,b,j}$ relative to the group mean at 70 m $\bar{g}_{f,b,70}$, by field and distance, is then calculated as:

$$a_{f,b,j} \triangleq |\bar{g}_{f,b,j} - \bar{g}_{f,b,70}| \tag{6}$$

And the average of absolute differences by distance j and by band b across all fields f is calculated as:

$$\bar{a}_{b,j} \triangleq \sum_{f \in F} a_{f,b,j} \tag{7}$$

We view the average absolute difference as a metric of the similarity of groups at varying distances from field edge, where pixels at the edges of fields are expected to have greater average absolute difference relative to center pixel groups due to spatial separation and greater heterogeneity at field borders.

3.6 Classifier selection

We compare the performance of multiple data preparation and classification approaches to select appropriate classification heads for each data input. As data inputs, we consider ‘raw’ (Sentinel-1 and Sentinel-2) time series data that is cloud masked (CM), raw data that is cloud masked with VIs added, STMs, STMs with S1 data and RVI, TESSERA embeddings and AlphaEarth embeddings. With these inputs we test 5 classifiers: logistic regression (LR), multi-layer perceptron (MLP), RF, SVM and XGBoost; each combination of data input and classifier is considered a scenario. We use 10% of fields for training and the remaining 90% are used for evaluating and validation when appropriate (taking 10% of data for validation). To mitigate spatial autocorrelation, all pixels from each field were assigned to a single subset of train, validation, or test. While neighboring fields may share similar spectral signatures, the relatively high number of fields and their distribution across the study area reduce potential spatial leakage. To select the scenarios to use for downstream analysis of border pixels, we evaluate overall classification accuracy, balanced F1 score, and overall F1 score using 1% of data over 20 runs with different random seeds to assess statistical confidence. The metrics for each scenario are shown in Table 4.

For small field and distance to border evaluation, we select the following classification heads based on the evaluation metrics for tested scenarios: 2) raw CM data with VIs classified with XGBoost for the ‘raw’ approach, 4) STMs with S1 and RVI classified with XGBoost for the STM approach, and 5) TESSERA embeddings using XGBoost for the embedding approach. Of all methods, the AlphaEarth embeddings meaningfully under perform and therefore are not considered further. While it is difficult to postulate as to why AlphaEarth embeddings underperform the other methods without open source code and model weights, we believe that the incorporation of spatial contextual information may meaningfully “blur” the agricultural signal at field boundaries which will have a disproportionate impact on regions with small fields or with fields including a high proportion of boundary pixels. AlphaEarth evaluates on a number of land cover based tasks, although not in any high boundary context scenarios and thus it is difficult to state with high confidence (Brown et al., 2025). As noted, the three selected input and classification scenarios use XGBoost (XGBClassifier, v3.0.2), and we report the tuned hyperparameters in Table 5.

4. Results

The proposed TESSERA embedding -based approaches achieve statistically significant higher classification accuracy in the most challenging settings: small fields.

4.1 Small field classification by crop

When classifying crop type of the small field subset, the TESSERA -based method has significantly higher F1 scores for five of seven classes compared to the other approaches, sometimes surpassing a 10% difference as shown in Figure 3. This margin is non-trivial in the context of agricultural monitoring, where even modest gains can enhance decision-making and resource allocation, underscoring the usefulness of the TESSERA approach for these systems.

4.2 Impact to distance to border on classification

For all distances to border, the TESSERA embedding -based classification approach has comparable or higher classification accuracy compared to the methods using raw data as seen in Figure 4. Additionally, we see classification results stabilize for pixels at 30 meters or greater from the field border; this stabilization occurs closer to the border than the raw data. Even in this approach considering all field sizes, the majority of pixels will lie in the less stable classification zone between 0 and 30 m which underscores the importance of classification accuracy at borders.

4.3 Few shot classification performance

Given that many crop classification tasks are limited by label availability, we evaluate F1 scores in a few-shot approach. Few-shot approaches train on few labels - in this case, between 5

	1) Raw + CM			2) Raw + CM + VIs			3) STMs		
	Accuracy	Macro F1	Bal F1	Accuracy	Macro F1	Bal F1	Accuracy	Macro F1	Bal F1
LR	69.27 ± 0.08	54.39 ± 0.07	67.65 ± 0.09	70.12 ± 0.08	56.09 ± 0.07	68.36 ± 0.09	65.43 ± 0.07	48.80 ± 0.07	65.07 ± 0.08
MLP	72.70 ± 0.28	58.98 ± 0.41	70.47 ± 0.25	71.89 ± 0.52	58.02 ± 0.71	69.56 ± 0.40	60.77 ± 0.65	43.19 ± 0.90	60.60 ± 0.57
RF	70.85 ± 0.09	52.60 ± 0.09	65.66 ± 0.11	71.20 ± 0.09	53.73 ± 0.10	66.05 ± 0.12	71.32 ± 0.09	53.45 ± 0.11	66.55 ± 0.12
SVM	70.84 ± 0.07	56.58 ± 0.07	68.35 ± 0.09	70.58 ± 0.07	56.27 ± 0.05	67.81 ± 0.09	68.98 ± 0.08	53.77 ± 0.07	66.61 ± 0.09
XGBoost	**73.93 ± 0.08**	**60.67 ± 0.06**	**69.85 ± 0.11**	**74.21 ± 0.09**	**60.75 ± 0.08**	**70.18 ± 0.11**	**73.42 ± 0.09**	**58.18 ± 0.10**	**69.45 ± 0.11**

	4) STMS + S1			5) TESSERA			6) AlphaEarth		
	Accuracy	Macro F1	Bal F1	Accuracy	Macro F1	Bal F1	Accuracy	Macro F1	Bal F1
LR	68.41 ± 0.08	52.91 ± 0.08	66.92 ± 0.09	72.96 ± 0.09	58.80 ± 0.09	71.09 ± 0.10	58.57 ± 0.09	40.52 ± 0.10	58.85 ± 0.10
MLP	64.80 ± 1.85	48.70 ± 2.82	63.48 ± 1.73	75.36 ± 0.14	63.10 ± 0.19	72.87 ± 0.12	**62.17 ± 0.16**	**42.07 ± 0.21**	**59.98 ± 0.14**
RF	71.41 ± 0.09	51.17 ± 0.10	66.57 ± 0.12	73.33 ± 0.09	56.35 ± 0.10	69.01 ± 0.11	60.75 ± 0.09	31.92 ± 0.08	53.21 ± 0.11
SVM	71.18 ± 0.09	55.77 ± 0.08	68.11 ± 0.10	73.39 ± 0.08	59.55 ± 0.08	71.05 ± 0.09	57.06 ± 0.09	38.02 ± 0.10	57.10 ± 0.10
XGBoost	**73.81 ± 0.08**	**59.74 ± 0.08**	**69.85 ± 0.10**	**75.88 ± 0.08**	**62.36 ± 0.10**	**72.74 ± 0.09**	52.73 ± 0.09	36.40 ± 0.12	57.03 ± 0.11

Table 4. Performance metrics (overall accuracy, macro F1, and balanced F1) across scenarios. Top: Raw + CM, Raw + CM + VIs, and STMs. Bottom: STMs + S1 + RVI, TESSERA embeddings, and AlphaEarth embeddings. CM stands for “cloud masked”, “VIs” stands for including vegetation indices as described in Section 3.3. The best performing scenarios are **in bold**.

Table 5. Comparison of XGBoost hyperparameters across different input approaches.

Parameter	**TESSERA**	**STM(S1,RVI)**	**Raw(CM,VIs)**
learning_rate	0.1	0.1	0.1
n_estimators	400	400	200
max_depth	6	3	6
min_child_weight	1	1	5
subsample	0.8	0.8	0.8
gamma	0	0	0.1
reg_alpha	0	0	0
reg_lambda	1	1	1.5
n_jobs	2	2	2

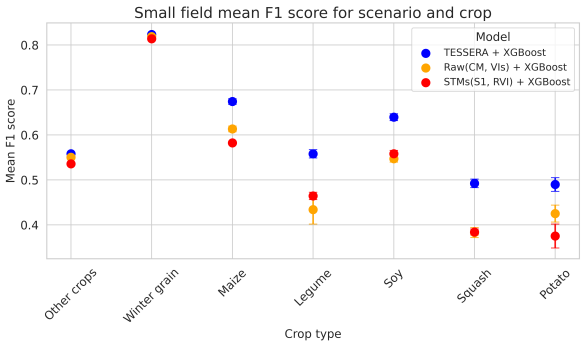

Figure 3. Small field F1 scores by crop class for each scenario with confidence intervals. Classes are ordered by number of fields in the dataset. The TESSERA -based methods have significantly higher F1 scores than the raw data -based methods for five classes, and on-par for two.

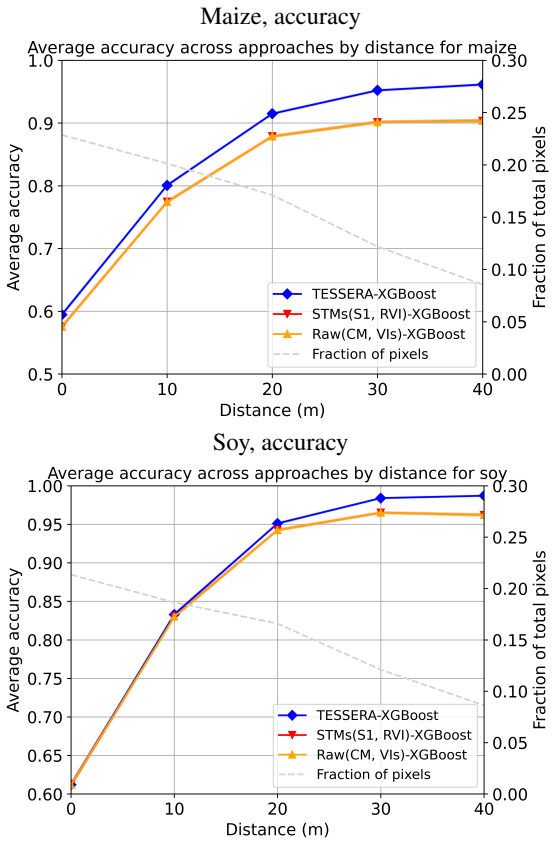

Figure 4. The average accuracy of each classification method (input and classifier combination) for each distance from the border is shown, along with the fraction of total pixels in each distance group. For all distances from the border, TESSERA -based methods have equal or higher accuracy for both corn and soy than raw data approaches.

and 100 labeled fields used for training. The TESSERA -based method consistently maintains a higher or comparable F1 score for two globally important staple crops, maize and soy, as shown in Figure 5.

4.4 Computational comparison

We additionally find TESSERA -based classification methods are computationally leaner than the other approaches. Compared to the best raw data approach, the TESSERA approach uses only 8% of the CPU used by the other methods, as shown in Table 6. Additionally, the TESSERA -based approach requires no feature engineering such as VI or STM generation, reducing further CPU demand, and necessitates no specialized knowledge about cropping seasons, meaningfully simplifying the classification process. Reduced computational needs and no requirement of feature engineering make TESSERA -based approaches more accessible and more scalable compared to the best raw data methods.

Table 6. Relative CPU performance ratios across methods

Method	**Ratio of CPU use**
Raw (CM, VIs) + XGBoost	1
STMs generation	92.82±1.06
STMs (S1, RVI) + XGBoost	0.9188±0.0045
TESSERA + XGBoost	0.0766±0.0013

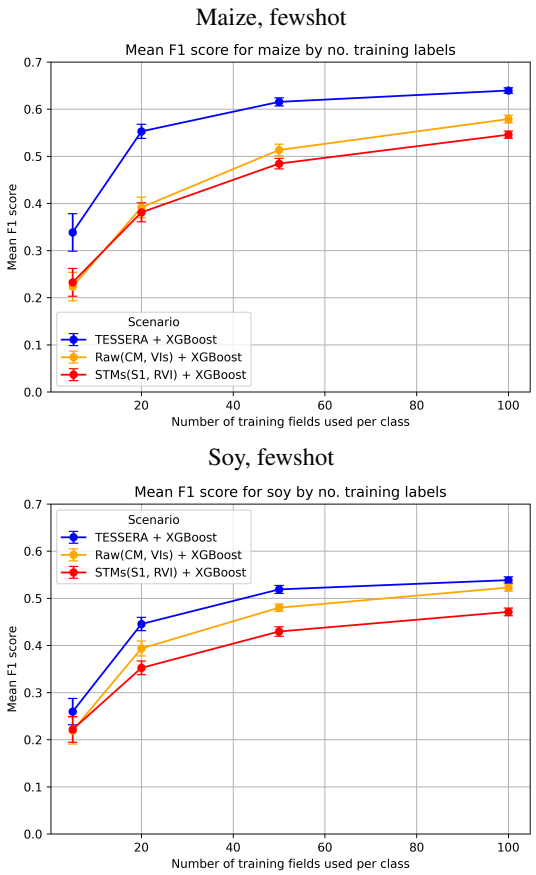

Figure 5. F1 scores for maize and soy, by scenario, as a function of the number of fields used to train the downstream classifier. Confidence intervals are from 20 runs. The TESSERA -based method maintains significantly higher F1 scores through most numbers of fields used for training.

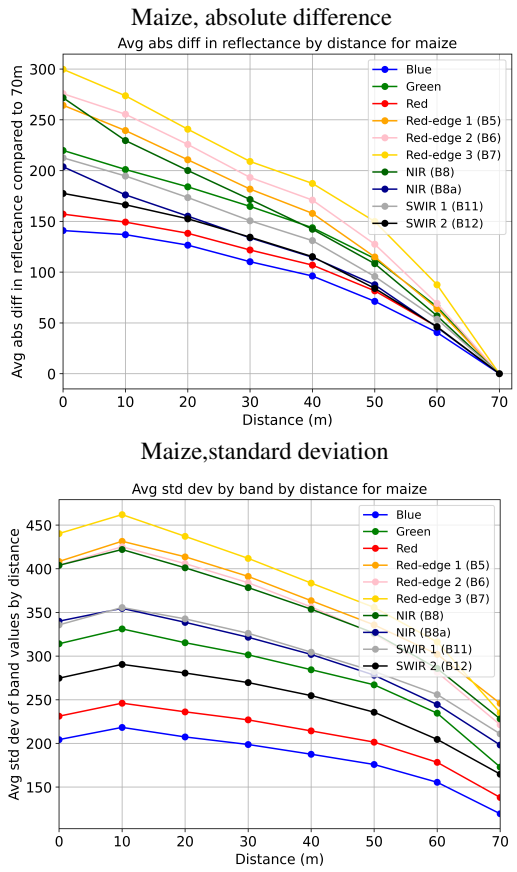

Figure 6. Absolute difference of average values compared to average values at 70 m ($\bar{a}_{b,j}$ of Equation 7) and average standard deviation ($\bar{\sigma}_{b,j}$ of Equation 3) among each distance group, for **maize**. Each of the 10 Sentinel-2 bands used are visualized.

4.5 Characterization of data at field edges

To contextualize the improved F1 score on the small field subset, we characterize the variability and stability of the Sentinel-2 spectral signal as a function of distance to border. We do this with average standard deviation, $\bar{\sigma}_{b,j}$, as defined in Section 3.5 and absolute average difference, $\bar{a}_{b,j}$, as defined in Section 3.5.2. Crop-specific plots of these two characterizations are shown for maize and soy, two globally common staple crops, in Figures 6 and 7 for maize and soy respectively.

For both maize and soy, the average standard deviation between pixels at the border is higher than at central pixels. Additionally, with distance from center pixels towards the border, the values of pixels across all bands differ increasingly in value from central pixels. This is consistent with the first law of geography, which posits that ‘everything is related to everything else, but near things are more related than distant things’ (Tobler, 1970). Reflecting this, spectral signature of a pixel in the field center is more consistent to other pixels in its distance bucket due to its spatial proximity to other pixels with similar environmental and management conditions. In contrast, as the distance from the center increases, the relationship between pixels weakens as they are increasingly spatially separated along increasing sized annuli, resulting in greater spectral variability and reduced stability. The variability and stability trends at the field border are exacerbated by the influence of mixed pixels including information outside the field as well as the changes to field management common at field edges.

The same trends of increasing homogeneity towards field centers appear when looking at the TESSERA embeddings as shown in Figure 8. However, the TESSERA embeddings stabilize for both maize and soy at 40 m into the interior of the field. Their greater spatial stability and robustness to edge effects are particularly relevant for small fields, where a large proportion of pixels lie near the border and are affected by spectral noise, mixed pixels, and field management irregularities. By stabilizing at a smaller distance to border than raw inputs and maintaining higher accuracy across all distances to the field edge, the TESSERA embeddings demonstrate an advantage in these spatially heterogeneous zones and explain the improved classification accuracy and F1 score.

5. Discussion and Conclusion

Our work demonstrates that pre-generated embeddings from TESSERA offer meaningful improvements for small field crop type classification. By creating a small field dataset in the absence of a standardized public benchmark, we provide a reproducible framework for assessing classification methods in difficult border pixel settings. We emphasize the urgent need for standardized public datasets in this area.

Crucially, our results demonstrate that the use of embeddings from TESSERA improves classification accuracy compared to the current baseline methods for small fields, with few labels, and while being more computationally efficient and eliminating the need for region-specific hand-engineered features. We

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-873-2026 | © Author(s) 2026. CC BY 4.0 License.

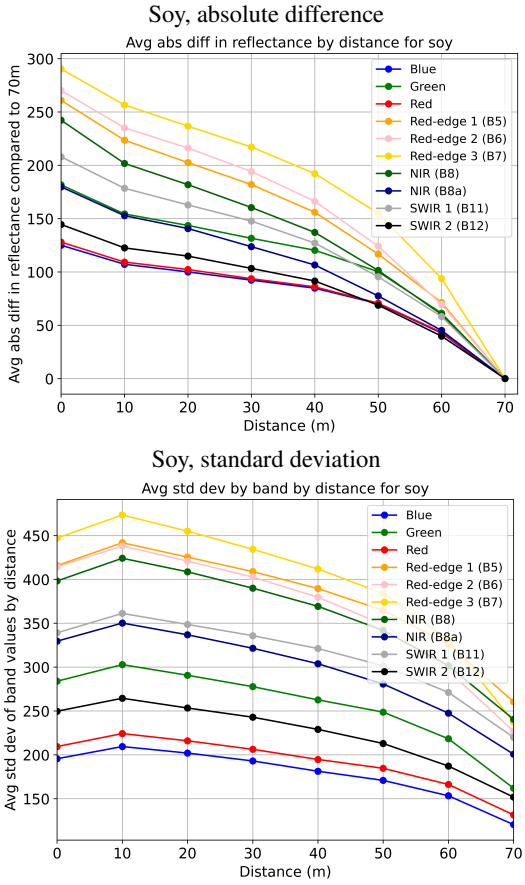

Figure 7. Absolute difference of average values to average values at 70 m ($\bar{a}_{b,j}$ of Equation 7) and average standard deviation ($\bar{\sigma}_{b,j}$ of Equation 3) among each distance group, for **soy**. Each of the 10 Sentinel-2 bands used are visualized.

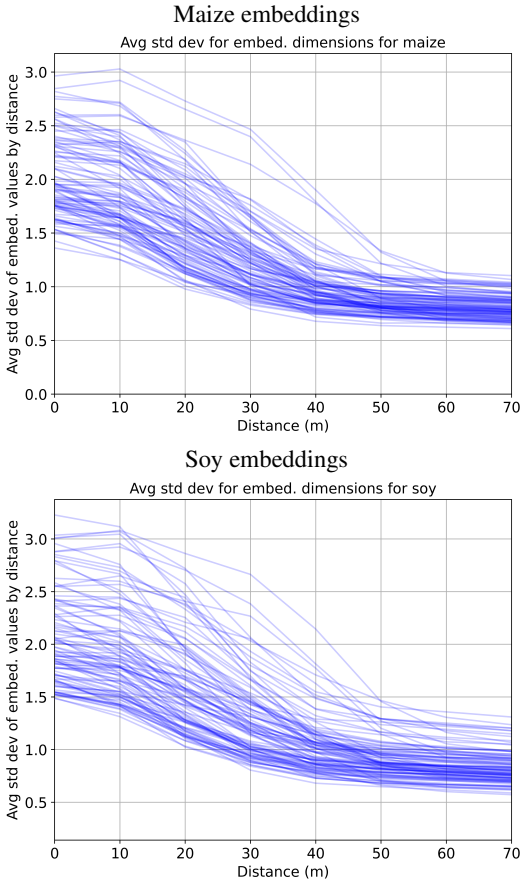

Figure 8. Average standard deviation ($\bar{\sigma}_{b,j}$ of Equation 3) among each distance group for TESSERA embeddings, for maize and soy. All dimensions are visualized, each as a transparent line. The average standard deviation stabilizes at 40 m from the border.

show that the TESSERA embeddings have greater spatial stability and robustness to edge effects compared to the raw data, which supports their improved performance on crop type classification in the small field context. As noted in Section 3.1, we hypothesize that this is due to the choice of augmentation strategy in TESSERA focusing on temporal invariance, which reduces embedding sensitivity to boundary conditions. This reduced sensitivity also supports why TESSERA -based few-shot learning for small fields yields higher F1 scores compared to other methods. Although discarded in the early approach selection phase described in Section 3.6, we also note that AlphaEarth performed meaningfully below all other methods for the crop classification task. We believe this could be due to the incorporation of spatial contextual information in the AlphaEarth embeddings that will likely disadvantage performance at field boundaries due to an information ’blurring’ effect likely to occur. In this way, we therefore emphasize that not all embedding -based approaches will offer advantages for small field crop classification tasks.

There are also a number of limitations to this work. This work focuses on one Austrian agricultural system and future work including case studies in small field systems across geographies and cropping systems would further validate findings. Establishing a standardized benchmark dataset for small fields would enable more consistent evaluation across studies. In this work, TESSERA embeddings are compared against methods developed for small field regions, but not against new deep learning approaches. Future work benchmarking all state of the art deep learning methods would offer a greater landscape of benefits and trade-offs, including around interpretability of methods. Furthermore, this work focuses on pixel-wise approaches and formal evaluation of the impact on spatially contextualized approaches at field boundaries would further test the claim that spatial approaches do not always benefit classification at field boundaries. We also note that future work could consider a full fine tuning of TESSERA for downstream tasks, which would open consideration of the utility of full fine tuning against other similarly positioned approaches. Finally, while this work focuses on small fields it does not address the additional challenges present in smallholder agricultural systems, and this remains to be further investigated.

Overall, our findings support the effectiveness of compact embeddings from TESSERA for crop type classification in small field systems and lay the groundwork for broader applications in data-sparse environments, with implications for both scientific advancement and food systems policy.

6. Acknowledgments

Thank you to UKRI and Mantle Labs for funding supporting this work.

References

Ayman, O., Gamal, M. E., ElSayed Atwa, M., Khaled, Y., El-Din Saad, N. G., 2024. Exploiting Machine Learning and Re-

mote Sensing for Precision Crop Mapping in Africa. *2024 34th International Conference on Computer Theory and Applications (ICCTA)*, 193–199. ISSN: 2770-6575.

Brown, C. F., Kazmierski, M. R., Pasquarella, V. J., Rucklidge, W. J., Samsikova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., Ilyushchenko, S., Gorelick, N., Zhang, L. L., Alj, S., Schechter, E., Askay, S., Guinan, O., Moore, R., Boukouvalas, A., Kohli, P., 2025. AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data. arXiv:2507.22291 [cs].

Feng, Z., Jaffer, S., Knezevic, J., Sormunen, S., Young, R., Lisaius, M., Immitzer, M., Ball, J., Atzberger, C., Coomes, D. A., Madhavapeddy, A., Blake, A., Keshav, S., 2025. TESSERA: Temporal Embeddings of Surface Spectra for Earth Representation and Analysis. arXiv:2506.20380 [cs].

Ibrahim, E. S., Rufin, P., Nill, L., Kamali, B., Nendel, C., Hostert, P., 2021. Mapping Crop Types and Cropping Systems in Nigeria with Sentinel-2 Imagery. *Remote Sensing*, 13(17), 3523. Number: 17 Publisher: Multidisciplinary Digital Publishing Institute.

INVEKOS Schläge Österreich 2022 - Open Government Data Austria, 2022.

Jiang, N., , Changyong, D., , Yunwei, T., , Charles, G., , Lin, Y., , Ding, H., 2024. Derivation of tasseled cap transformation coefficients for SDGSAT-1 Multispectral Imager at-sensor reflectance data. *International Journal of Digital Earth*, 17(1), 2413885. Publisher: Taylor & Francis.

Lebourgeois, V., Dupuy, S., Vintrou, , Ameline, M., Butler, S., Bégué, A., 2017. A Combined Random Forest and OBIA Classification Scheme for Mapping Smallholder Agriculture at Different Nomenclature Levels Using Multisource Data (Simulated Sentinel-2 Time Series, VHRS and DEM). *Remote Sensing*, 9(3), 259. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

Lisaius, M. C., Blake, A., Keshav, S., Atzberger, C., 2024. Using Barlow Twins to Create Representations From Cloud-Corrupted Remote Sensing Time Series. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17. Conference Name: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.

Lobell, D. B., Asner, G. P., 2004. Cropland distributions from temporal unmixing of MODIS data. *Remote Sensing of Environment*, 93(3), 412–422.

Lowder, S. K., Skoet, J., Raney, T., 2016. The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide. *World Development*, 87, 16–29.

Lu, S., Guo, J., Zimmer-Dauphinee, J. R., Nieusma, J. M., Wang, X., vanValkenburgh, P., Wernke, S. A., Huo, Y., 2025. Vision Foundation Models in Remote Sensing: A survey. *IEEE Geoscience and Remote Sensing Magazine*, 2–27.

Maolan, K., , Yusufujiang, R., , Zhang, X., , Kuluwan, Y., 2024. Sentinel-2 image based smallholder crops classification and accuracy assessment by UAV data. *Geocarto International*.

Maurya, K., Mahajan, S., Chaube, N., 2021. Remote sensing techniques: mapping and monitoring of mangrove ecosystem—a review. *Complex & Intelligent Systems*, 7(6).

Mohammadi, S., Belgiu, M., Stein, A., 2023. Improvement in crop mapping from satellite image time series by effectively supervising deep neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 198, 272–283.

Pribadi, D. O., Rustiadi, E., Syamsul Iman, L. O., Nurdin, M., Supijatno, Saad, A., Pravitasari, A. E., Mulya, S. P., Ermyanyla, M., 2023. Mapping smallholder plantation as a key to sustainable oil palm: A deep learning approach to high-resolution satellite imagery. *Applied Geography*, 153, 102921.

Rao, P., Zhou, W., Bhattarai, N., Srivastava, A. K., Singh, B., Poonia, S., Lobell, D. B., Jain, M., 2021. Using Sentinel-1, Sentinel-2, and Planet Imagery to Map Crop Type of Smallholder Farms. *Remote Sensing*, 13(10), 1870. Number: 10 Publisher: Multidisciplinary Digital Publishing Institute.

Ren, T., Xu, H., Cai, X., Yu, S., Qi, J., 2022. Smallholder Crop Type Mapping and Rotation Monitoring in Mountainous Areas with Sentinel-1/2 Imagery. *Remote Sensing*, 14(3), 566. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

Rußwurm, M., Courty, N., Emonet, R., Lefèvre, S., Tuia, D., Tavenard, R., 2023. End-to-end learned early classification of time series for in-season crop type mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196, 445–456.

Rustowicz, R. M., Cheong, R., Wang, L., Ermon, S., Burke, M., Lobell, D., 2019. Semantic Segmentation of Crop Type in Africa: A Novel Dataset and Analysis of Deep Learning Methods. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Soler, J. L., Friedel, T., Wang, S., 2024. Combining Deep Learning and Street View Imagery to Map Smallholder Crop Types. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(20), 22202–22212. Number: 20.

Tobler, W. R., 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46, 234–240.

Turkoglu, M. O., D’Aronco, S., Perich, G., Liebisch, F., Streit, C., Schindler, K., Wegner, J. D., 2021. Crop mapping from image time series: Deep learning with multi-scale label hierarchies. *Remote Sensing of Environment*, 264, 112603.

Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, 236, 111402.

Wu, B., Zhang, M., Zeng, H., Tian, F., Potgieter, A. B., Qin, X., Yan, N., Chang, S., Zhao, Y., Dong, Q., Boken, V., Plotnikov, D., Guo, H., Wu, F., Zhao, H., Deronde, B., Tits, L., Loupian, E., 2023. Challenges and opportunities in remote sensing-based crop monitoring: a review. *National Science Review*, nwac290.

Xie, B., Zhang, H. K., Xue, J., 2019. Deep Convolutional Neural Network for Mapping Smallholder Agriculture Using High Spatial Resolution Satellite Image. *Sensors*, 19(10), 2398. Number: 10 Publisher: Multidisciplinary Digital Publishing Institute.

Yuan, Y., Lin, L., 2020. Self-Supervised Pre-Training of Transformers for Satellite Image Time Series Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, PP.

Yuan, Y., Lin, L., Liu, Q., Hang, R., Zhou, Z.-G., 2022. SITS-Former: A pre-trained spatio-spectral-temporal representation model for Sentinel-2 time series classification. *International Journal of Applied Earth Observation and Geoinformation*.

Zheng, Y., Dong, W., ZhipingYang, Lu, Y., Zhang, X., Dong, Y., Sun, F., 2024. A new attention-based deep metric model for crop type mapping in complex agricultural landscapes using multisource remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, 134, 104204.