

Impact of sampling techniques on crop type mapping using multi-temporal composites from Harmonized Landsat-Sentinel images[☆]

Wilson R.V. Aires^{a,*}, Vitor S. Martins^{a,*}, Lucas B. Ferreira^a, Yanbo Huang^b, Lucas Heintzman^c, Ying Ouyang^d

^a Department of Agricultural and Biological Engineering, Mississippi State University, Mississippi State, MS 39762, United State

^b USDA-ARS Genetics and Sustainable Agriculture Research Unit, Mississippi State, MS 39762, United State

^c USDA-ARS Water Quality and Ecology Research Unit, National Sedimentation Laboratory, Mississippi, MS 38655, United State

^d USDA Forest Service, Southern Research Station, Center for Bottomland Hardwoods Research, Mississippi State, MS 39762, United State

ARTICLE INFO

Keywords:

Crop type classification
Machine learning
Impact of sample size
Spectral indices
Temporal composition

ABSTRACT

Crop type mapping provides key spatial information for agricultural monitoring, management, production, and food security. Satellite multi-spectral imagery and machine learning algorithms have been used for large-area crop-type mapping, which necessitates training sample selection for high-quality mapping results. However, questions persist regarding the optimal sampling techniques and sample size to employ during application development. The novelty of this study focuses on investigating how sampling techniques and sample sizes influence the accuracy and reliability of satellite-based crop type mapping. This topic is often overlooked by practitioners and can have a drastic influence on the resulted maps. This study investigated how different sampling techniques and sample sizes influence crop type mapping in the Yazoo – Mississippi Delta (YMD), a complex agroecological landscape of northwestern Mississippi, United States. We derived a set of temporal composites from 2022 Harmonized Landsat Sentinel-2 (HLS) images and utilized the Cropland Data Layer (from the United States Department of Agriculture National Agricultural Statistics Service) as the ground reference to train and validate the classification model. The influence of four sampling techniques (grid, random, stratified, and cluster techniques) and sample sizes (0.05%, 0.1%, 0.50%, and 1.00% of the total pixels of the HLS tile and a minimum size of 100 samples per class) was evaluated with Artificial Neural Network classification of six primary crops in the YMD, including corn, cotton, rice, sorghum, soybeans, and wheat. Our results demonstrated that the stratified sampling technique achieved the highest performance across all crop-type classes, with an average F1-score of 0.81. In contrast, the grid sampling technique yielded the lowest performance, with an F1-score of 0.63. The cluster sampling technique also presented good results, with an average F1-score of 0.79. Classes representing small areas in the YMD region, such as sorghum and wheat, were not adequately captured by the grid sampling technique (F1-scores ranging from 0 to 0.2) or the random sampling technique (F1-scores ranging from 0.10 to 0.65). Allocating the samples considering a proportion between the area represented by the classes, and increasing the number of samples improved the model's performance. The highest accuracy was obtained using 0.05% and 1.00% of the total number of pixels in the image tile as training samples, with an average F-score ranging from 0.81 to 0.84, respectively. When an equal number of samples was distributed per class, the results revealed significant mapping confusion among crop-type classes, particularly for other land use classes, which achieved an F1-score of 0. Our findings suggest that the optimal sampling technique involves allocating samples per image layer based on the proportional area of each class to determine the number of samples. Also, HLS-derived temporal composites are effective for crop type mapping, which may assist producers and stakeholders in more effective management of agricultural activities.

[☆] This article is part of a special issue entitled: 'Agricultural Cybernetics' published in Computers and Electronics in Agriculture.

* Corresponding authors.

E-mail address: wilson.aires@msstate.edu (U.R.V. Aires).

1. Introduction

Crop type mapping is essential for numerous agricultural production activities, including monitoring crop growth, optimizing resource allocation, improving yield predictions, and supporting sustainable land management practices (Gao et al., 2023; Li et al., 2023; Zhang et al., 2020). The use of cropland maps is crucial for global food security as it enables the quantification of annual food production, yield forecasting, and production trends, including agricultural area expansions or reductions (Becker-Reshef et al., 2023). Accurate crop type maps support further public and private initiatives that enhance agricultural productivity and sustainability (Basso and Antle, 2020; Kavvada et al., 2020; Wolfert and Isakhanyan, 2022), making crop maps indispensable information for addressing food availability and supply uncertainties. Traditional crop type maps typically integrate field surveys, statistical sampling reports, and satellite/aerial data interpretation. Due to spatial and temporal coverage, satellite images have become one of the most cost-effective methods to annually generate crop maps at multiple scales (Gao et al., 2023; Li et al., 2023).

Several crop maps derived from freely available satellite images have demonstrated their effectiveness in accurately classifying the world's primary crops, including corn, soybeans, wheat, and rice. Often these crop maps are derived from the 250/500-m Moderate-Resolution Imaging Spectroradiometer (MODIS) (Clausen et al., 2016; Massey et al., 2017; Pittman et al., 2010; Wardlow et al., 2007; Wardlow & Egbert, 2008), 30-m Landsat satellite series (Cai et al., 2018; Potapov et al., 2021; Song et al., 2022; Zhong et al., 2014; Zhong et al., 2019), 10/20-m Sentinel-2 (Belgiu & Csillik, 2018; Cai et al., 2018; Lambert et al., 2018; Li et al., 2023; Xu et al., 2024), and 3-m PlanetScope images (Acharki et al., 2023; Bueno et al., 2023; Kpienbaareh et al., 2021; Seong et al., 2024). Moreover, the Harmonized Landsat Sentinel-2 (HLS) product has brought a new perspective to crop mapping. It provides a virtual satellite constellation by combining the Operational Land Imager (OLI/OLI-2) and Multi-Spectral Instrument (MSI) sensors from Landsat 8/9 and Sentinel-2 A/B satellites, respectively. This constellation offers a 2- to 3-day temporal revisit and includes a reliable quality assessment band for masking clouds, cloud shadows, snow/ice, and water pixels (Claverie et al., 2018). Recent studies have demonstrated the use of HLS products for crop phenology measurement, crop type, and crop area mapping (Bolton et al., 2020; Gao & Zhang, 2021; Hao et al., 2020; Kyere et al., 2020). The combination of machine learning methods with enhanced satellite data products has significantly improved the accuracy of crop maps (Joshi et al., 2023; Venter et al., 2022). In particular, tree-based algorithms like random forest and extreme gradient boost machines have been used (Bolfé et al., 2023; Hao et al., 2020; Kyere et al., 2020; Li et al., 2023), as well as support vector machines (Duro et al., 2012; Gilbertson et al., 2017; Zheng et al., 2015). Artificial neural networks (ANN), including deep learning models (Feng et al., 2023; Wang et al., 2021, 2023; Zhang et al., 2019; Zhong et al., 2019) have also been employed. However, algorithm effectiveness is significantly affected by the input data quality and training data sampling (Stehman, 2009; Yuan et al., 2020).

As reported in previous studies, the use of different sampling techniques and sample sizes directly impacts the quality of the classification maps, as they are responsible for creating representative samples for target classification (Stehman & Wagner, 2024; Gao et al., 2023; Olofsson et al., 2014). Most commonly used sampling techniques for crop type mapping are random sampling (Goodbody et al., 2023), cluster sampling (Li et al., 2023), stratified random sampling (Yu et al., 2023), and systematic/grid-based sampling (Yan et al., 2021). These techniques influence the allocation of samples for reference labeling and can eventually result in selecting fewer representative samples (Goodbody et al., 2023; Stehman & Wagner, 2024). Stehman and Wagner (2024) revised sampling techniques for land use and land cover mapping and indicated that stratified sampling is more effective in representing classes, as it allocates samples based on the spatial location

of the pixel's class. Yu et al. (2023) developed crop mapping using stratified random sampling and the unsupervised cluster technique to select optimal sampling locations for crop-type mapping. They found that less representative classes such as soybeans (0.84 % of the study area) were not mapped properly using the cluster technique. If simple random or systematic sampling techniques are implemented, there is a high chance of underrepresentation of less representative classes while overrepresenting dominant classes due to their area-proportional biases (Stehman & Wagner, 2024). Additionally, sample size considerations have been a long-standing topic in remote sensing analysis, significantly influencing the accuracy of satellite image products. Inadequate sample sizes can lead to the under- or over-representation of specific crop classes, affecting classification reliability (Farhadpour et al., 2024; Hamunyela et al., 2022; Stehman, 2012; Wagner & Stehman, 2015).

The novelty of this study focuses on investigating how sampling techniques and sample sizes influence the accuracy and reliability of satellite-based crop type mapping. This topic is often overlooked by practitioners and can have a drastic influence on the resulted maps. The sampling data selection, especially regarding spatial representation and number of samples, remains a challenge and subjective. Previous studies on cropland mapping have determined the number of samples using assumptions about coverage area representation or a minimum number of sampling points that is feasible for labeling (Huang et al., 2024; Li et al., 2023; Potapov et al., 2021; Zheng et al., 2023). While a single solution does not exist, further discussion is needed to demonstrate the impacts of sampling strategy and sample size. This study brings new perspectives to these gaps by evaluating the performance of four sampling techniques (grid, random, stratified, and cluster) and multiple sample sizes (0.05 %, 0.1 %, 0.50 % and 1.00 % of the total pixels of the HLS tile and a minimum size of 100 samples per class) on 2022 crop type classification using HLS-based temporal composites. The study was developed in a portion of the Yazoo-Mississippi Delta (YMD). The diverse and dynamic agricultural landscape of the YMD makes it a critical area for studying crop mapping and sampling techniques. The region's complex variability in cropping systems, driven by multiple crop types, varying field sizes, and seasonal transitions, presents significant challenges for machine learning-based classification. Ensuring accurate crop classification in this area requires well-structured sampling techniques and an appropriate sample size to capture its heterogeneity effectively. A large set of spectral input variables was generated during the crop growing season from HLS images, and the 2022 U.S. Cropland Data Layer (CDL) was integrated as ground reference data to create-train-validate an Artificial Neural Network (ANN). The ANN model was trained to map six crop types (corn, cotton, rice, sorghum, soybeans, and wheat). Our main contributions include (1) evaluating sampling techniques for generating crop-type samples and analyzing their impact on classification performance, (2) demonstrating the effect of varying sample sizes on crop-type classification using a neural network model and assessing the sensitivity of classification to different sampling strategies, and (3) leveraging HLS temporal composites for crop-type mapping in a heterogeneous agricultural region.

2. Materials

2.1. Overview

The workflow presented in Fig. 1 illustrates our overall methodology for this study, and is organized as follows: (1) reclassification of the CDL crop type reference data to obtain the six main crop classes (maize, cotton, rice, sorghum, soybeans, and wheat), and "other" class including urban areas, forest, and water resources for a sub-portion of the YMD; (2) creation of sample points using four sampling techniques for model training and validation using the reclassified CDL as ground reference; (3) acquisition and processing of HLS time series images to mask the clouds and cloud shadow in the temporal stack; (4) generation of seasonal composites with the percentiles of 10 %, 25 %, 50 %, 75 %, and 90 % per

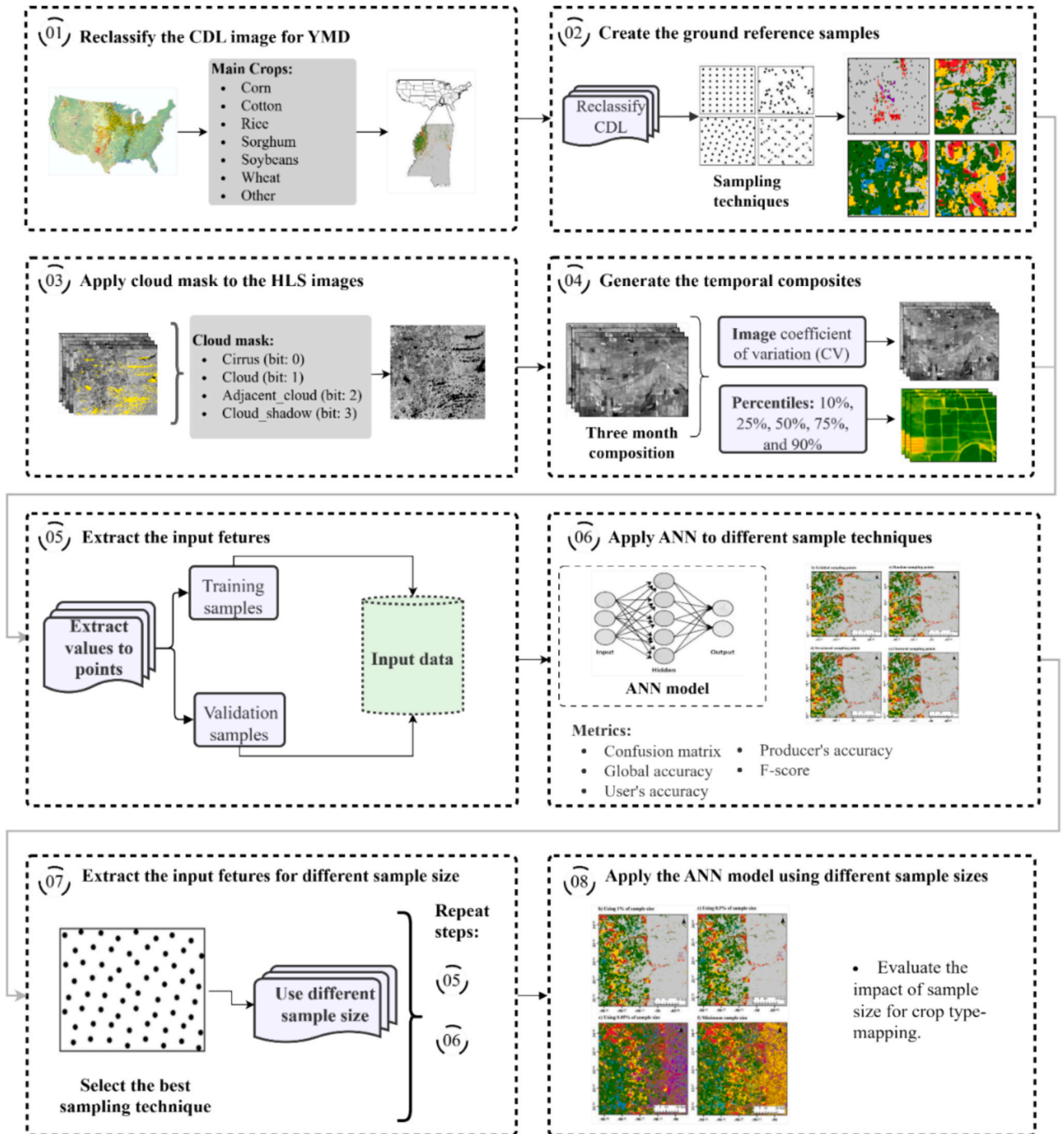


Fig. 1. Workflow to evaluate the impacts of different sampling techniques and sample sizes on crop type mapping using HLS-derived temporal composites.

spectral band, calculate spectral indices, and coefficient of variation; (5) extraction of per-pixel input variables and CDL-based ground reference samples derived from the sample points; (6) application and evaluation of the crop type classification model using different sampling techniques; (7) extracting the training samples considering different sample sizes; and (8) application and evaluation the crop type classification model under different sample sizes.

2.2. Study area

The study area is a sub-portion of the YMD (Heintzman et al. 2024) (Fig. 2) and comprises an entire HLS tile (12,056.04 km²) covering 13

Mississippi counties (Lafayette, Sunflower, Tallahatchie, Panola, Yalobusha, Leflore, Quitman, Bolivar, Washington, Coahoma, and Grenada). The region is representative of Mississippi's agricultural activities and is a critical agroecoeconomic activity (Yasarer et al., 2020, Dash et al., 2023). The climate is characterized by temperate winters, long and hot summers, and the average annual precipitation is 1,412 mm with a wet period from November to April and a dry season from May to October (Dash et al., 2023, Ouyang et al., 2023). The YMD region has a low population density, with a significant portion of its land dedicated to agricultural activities (Coupe et al., 2012). The primary crops cultivated in the basin include soybean, cotton, rice, and corn. The floodplains are a critical ecological feature of the region, providing nutrients for

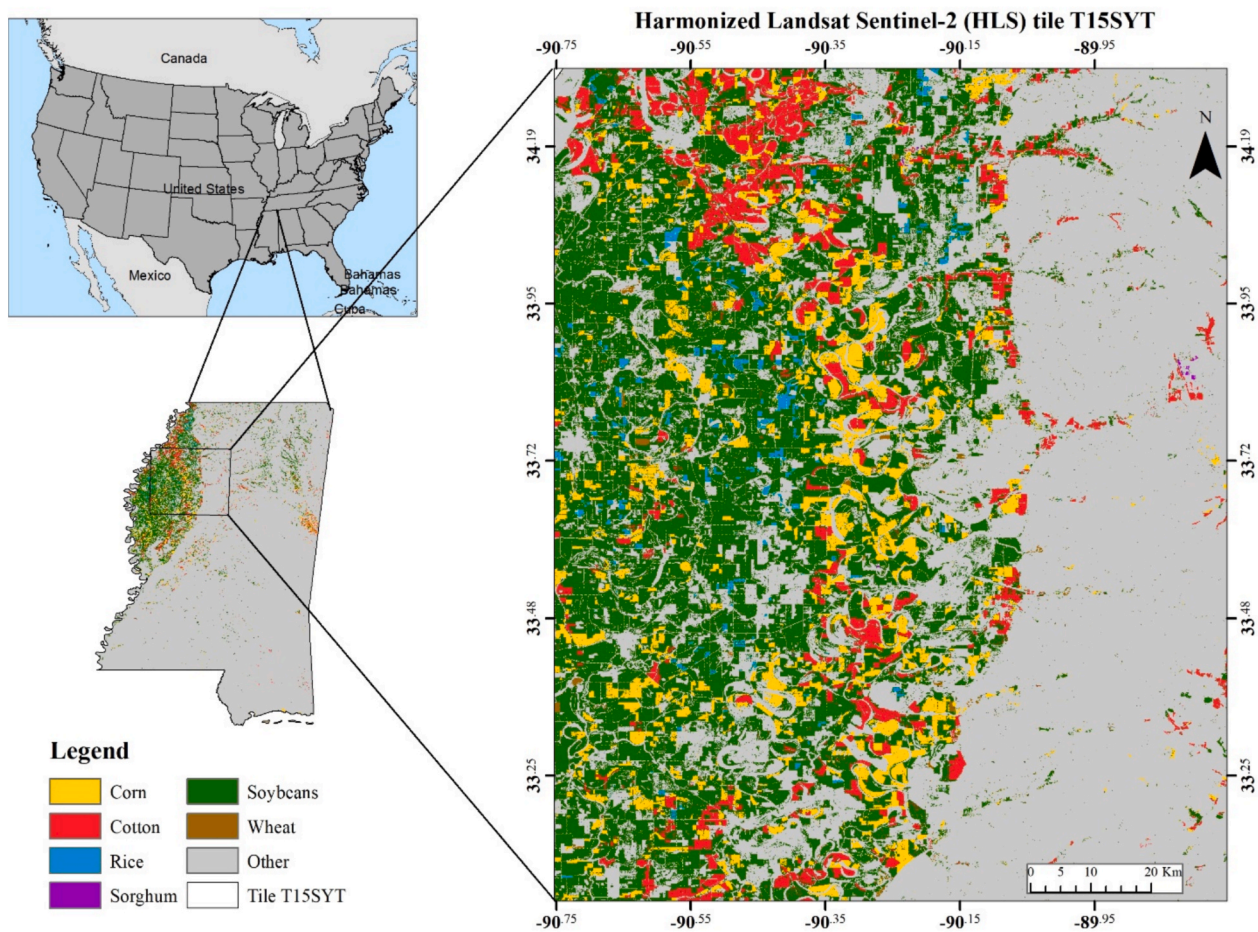


Fig. 2. Location of the study area, which comprises a Harmonized Landsat Sentinel-2 (HLS) tile, including the reclassified cropland data layer (CDL) to represent the main crops of the Mississippi Delta.

Table 1

Area and the percentage of area per crop and their planting and harvest dates within the HLS tile T15SYT.

Crop type	Area (ha)	Area (%)	Usual Planting Dates ⁱ	Usual Harvesting Dates ⁱ
Corn	73,310.0	6.18 %	March 24 – April 27	August 23 – September 23
Cotton	70,345.0	5.85 %	April 27 – May 19	September 27 – October 29
Rice	12,727.0	1.07 %	April 18– May 16	September 5 – October 6
Sorghum	204.0	0.02 %	April 14 – May 21	August 29 – September 27
Soybeans	404,838.0	32.87 %	April 14 – May 21	September 13 – October 31
Wheat	3,399.0	0.26 %	October 10 – November 18	June 2 – June 21
Other	640,007.0	53.75 %	–	–

ⁱ Source: Dash et al. (2023), NASS (2010).

agricultural activities and serving as habitats for natural areas that sustain diverse wildlife. (Heintzman et al., 2024). The multiple crop types in the YMD make it a critical area for studying crop mapping and sampling techniques. Fig. 2 represents the HLS T15SYT tile where CDL crop type data were reclassified to obtain the main crops of the area: corn, cotton, rice, sorghum, soybeans, and wheat. All the remaining land cover classes, including urban areas, forest, and water resources, were placed into the “other” class. Table 1 lists the area and the percentage of area per crop, along with their regionally typical planting and harvest dates.

2.3. Data sources

2.3.1. Cropland data layer (CDL)

The CDL data are freely available annually, providing information on approximately 110 land cover and crop type classes at a 30-meter satellite imagery resolution with extensive ground truthing for the entire

contiguous USA (CONUS) since 2008 (Boryan et al., 2011; Han et al., 2014). This dataset is published by the US Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) and can be obtained from the CDL website (<https://nassgeodata.gmu.edu/CropScape/>). Validation of CDL products has demonstrated a global accuracy of 84.3 % and accuracy ranging from 85 % to 95 % for major crops, including corn, soybeans, and wheat (Boryan et al., 2012; Johnson, 2013). In this study, a CDL image corresponding to the HLS tile for 2022 was utilized and reclassified to identify the primary crops in the YMD, which include corn, cotton, rice, sorghum, soybeans, and wheat. The reclassified CDL image served as the ground reference for creating samples to train and test the classification model, which was evaluated among sampling techniques and sample sizes in this study. A majority filter was applied to the CDL image to eliminate the salt-and-pepper effects, minimizing the problem of isolated pixels in certain locations (Wang et al., 2022).

2.3.2. Harmonized Landsat Sentinel-2 image data

The HLS product is a virtual constellation derived from data acquired by the Operational Land Imager (OLI/OLI-2) and multi-spectral instrument (MSI) sensors of Landsat 8/9 and Sentinel-2 A/B, respectively. The OLI sensors onboard Landsat 8/9 satellites provide 11 multispectral bands with 30 m resolution for the visible spectral region (B02 to B04), near infra-infrared (NIR), and short wave infra-red bands (SWR1 and SWR2), with a return interval of eight days (Niclòs et al., 2023). The MSI on board Sentinel-2 A/B presents 13 multispectral bands with 10-m resolution for the visible spectral region (B02, B03, B04) and NIR, 20-m for red edge (B05 to B08A), and 60-m for deep blue (B01) and short-wave infra-red bands (SWR1 and SWR2). The return interval has been every 5 days when Sentinel-2A and -2B were paired, since March 2017 (ESA, 2015). The HLS project incorporates a range of algorithms to generate a unified image time series from OLI and MSI data, ensuring reliable surface reflectance measurements. The process includes atmospheric correction using the LaSRC algorithm, which employs the 6S radiative transfer code and was initially designed for Landsat-8 (Vermote et al., 2016) before being adapted for Sentinel-2. Additionally, cloud and cloud-shadow masking is performed using the FMask algorithm (Zhu et al., 2015), which accounts for aerosols, water vapor, and other atmospheric components that could distort the signal. The process also involves identifying and removing pixels contaminated by clouds or their shadows, as well as resampling the data onto a consistent spatial grid to ensure proper alignment of information from both sensors. Further adjustments are made for illumination and viewing angles to reduce discrepancies caused by different sensor perspectives. Additionally, the process normalizes the Bidirectional Reflectance Distribution Function (BRDF) effects using the c-factor approach (Roy et al., 2016) to correct for changes in reflectance caused by varying sensor angles and surface properties. It also aligns the spectral bands from OLI and MSI to ensure accurate correspondence in terms of wavelength ranges and bandwidth. The HLS dataset contains both HLS-L30 (Landsat 8/9) and HLS-S30 (Sentinel-2 A/B) using the same tiling system of original Sentinel-2 images. Both products offer a 30-meter spatial resolution and a temporal resolution of 2–3 days globally. A detailed description of the HLS product can be found in Claverie et al. (2018) and GSFC (2018). Table 2 presents the spectral band nomenclature of the HLS bands corresponding to the OLI and MSI bands, along with the wavelength of each band and the spatial resolution.

In this study, HLS images were acquired from March to November 2022 to characterize the crop patterns in the YMD (Table 1), resulting in a total of 72 images for tile 15SYT. For each image, the HLS quality assurance (QA) flags were extracted from the images to identify pixels affected by atmospheric conditions, including cirrus (bit 0), clouds (bit 1), adjacent clouds (bit 2), and cloud shadows (bit 3). Pixels flagged by

any of these criteria were then masked as “no data” to ensure that only reliable, clear-sky observations were used in subsequent analyses. Fig. 3 presents the cloud percentage observed in the used HLS scene for Landsat (L30) and Sentinel-2 (S30).

3. Methods

3.1. Input variables for machine learning-based crop type mapping

In this study, we used HLS-based temporal composites from cloud-free pixels for three periods: March to May, June to August, and September to November 2022. These composites served two key purposes: (i) reducing temporal gaps caused by cloud cover and shadows, and (ii) capturing crucial temporal information related to crop growth stages. The latter is particularly valuable for distinguishing different crop types, as each exhibits unique phenological patterns (Yan & Roy, 2016). Fig. 4 shows the true color composite for the spectral average in three selected periods, illustrating the quality of the seasonal composition for crop variability in the study area. In each period, we calculated the coefficient of variation (CV) and the percentiles of 10 %, 25 %, 50 %, 75 %, and 90 % over time for the HLS-based blue (B02), green (B03), red (B04), near-infrared (NIR), short-wave infrared 1 (SWIR1), and short-wave Infrared 2 (SWIR2) bands. The coefficient of variation (CV) is defined as the ratio of the standard deviation to the mean per pixel over time. We calculated the CV for the band features to normalize their variability since different bands have varying ranges and levels of noise. Using the CV standardizes these differences and ensures that the model can effectively utilize the full dynamic range of each band (Donoghue et al., 2007). Based on the temporal composites, we calculated the spectral indices and extracted multiple percentiles from them (Table 3). We computed these percentiles exclusively for the spectral indices because they are normalized versions of the spectral bands, which reduces noise. Moreover, using percentiles allows us to capture various aspects of the distribution that can be linked to specific phenological crop stages, for example, the 10th percentile may correspond to the beginning of the growing stage and helps to distinguish crops from other land uses (Carrasco et al., 2022). A total of 93 features were used to train the classification model, with 75 features derived from the 5 spectral indices $\times$ 5 percentiles $\times$ 3 periods, and 18 coefficients of variation (6 bands $\times$ 3 periods).

3.2. Sampling techniques used to create the crop type reference dataset

The application of sampling techniques influences the training sample quality and balance for supervised machine learning (Li et al., 2023; Yu et al., 2023). Choosing the optimal sampling techniques

Table 2

Spectral band nomenclature of the HLS bands corresponding to the OLI and MSI bands, wavelength, and spatial resolution. Selected bands were used in this study in bold.

Band name	OLI band number	HLS-L30	MSI band number	HLS-S30	Wavelength (μm)	Spatial Resolution (m)
Coastal Aerosol	1	B01	1	B01	0.43–0.45	30
Blue	2	B02	2	B02	0.45–0.51	30
Green	3	B03	3	B03	0.53–0.59	30
Red	4	B04	4	B04	0.64–0.67	30
Red-Edge 1	—	—	5	B05	0.69–0.71	30
Red-Edge 2	—	—	6	B06	0.73–0.75	30
Red-Edge 3	—	—	7	B07	0.77–0.79	30
NIR Broad	—	—	8	B08	0.78–0.88	30
NIR Narrow	5	B05	8A	B8A	0.85–0.88	30
SWIR 1	6	B06	11	B11	1.57–1.65	30
SWIR 2	7	B07	12	B12	2.11–2.29	30
Water vapor	—	—	9	B09	0.93–0.95	30
Cirrus	9	B09	10	B10	1.36–1.38	30
Thermal Infrared 1	10	B10	—	—	10.60–11.19	30
Thermal Infrared 2	11	B11	—	—	11.50–12.51	30

Source: GSFC (2018).

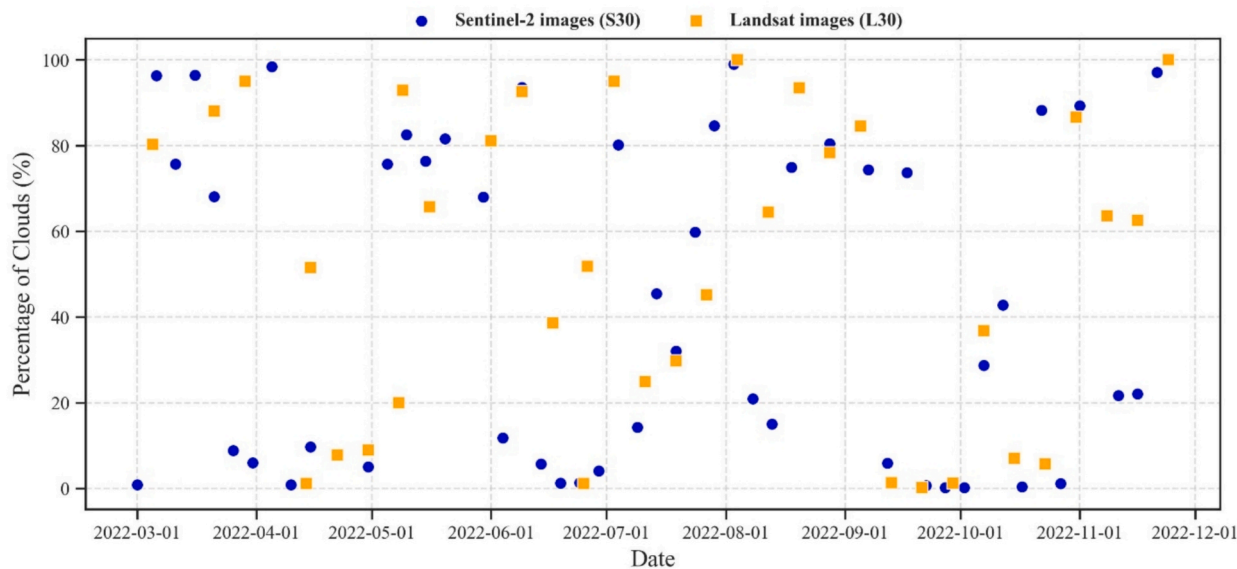


Fig. 3. Cloud percentage for Landsat (L30) and Sentinel-2 (S30) images in used HLS tile.

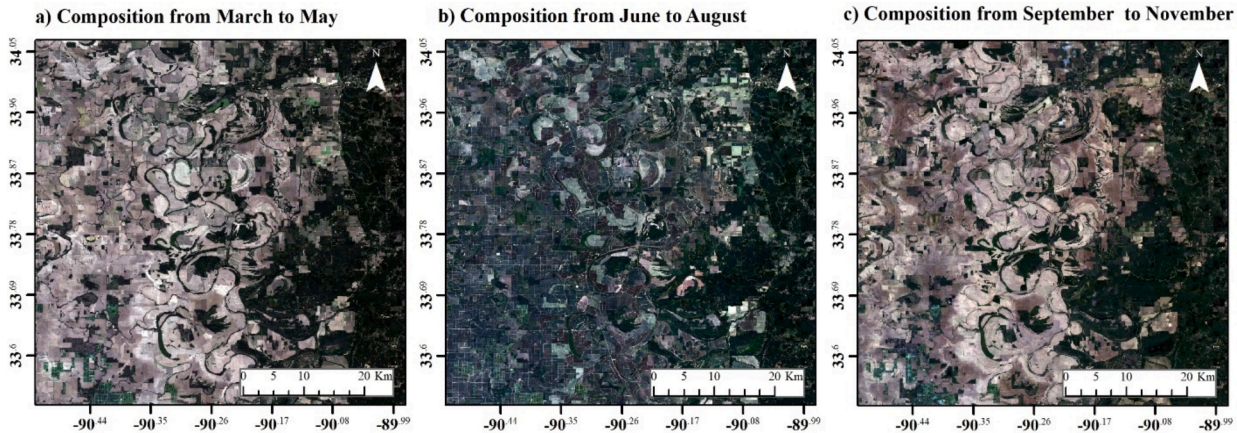


Fig. 4. Temporal composite for natural color (RGB) images in the study: Temporal composites used (a) March to May, (b) June to August, and (c) September to November in 2022 images.

Table 3
Spectral indices are used for crop classification, formula, and reference.

Spectral Indices	Acronym	Formula	Reference
Enhanced Vegetation Index	EVI	$2.5 \cdot (\text{NIR} - \text{Red}) / (\text{NIR} + 6 \cdot \text{Red} - 7.5 \cdot \text{Blue} + 1)$	Huete et al. (2002)
Normalized Burn Ratio	NBR	$(\text{NIR} - \text{SWIR2}) / (\text{NIR} + \text{SWIR2})$	Key & Benson (2006)
Normalized Difference Built-up Index	NDBI	$(\text{SWIR1} - \text{NIR}) / (\text{SWIR1} + \text{NIR})$	Zha et al. (2003)
Normalized Difference Tillage Index	NDTI	$(\text{SWIR1} - \text{SWIR2}) / (\text{SWIR1} + \text{SWIR2})$	Deventer et al. (1997)
Normalized Difference Water Index	NDWI	$(\text{NIR} - \text{SWIR1}) / (\text{NIR} + \text{SWIR1})$	Gao (1996)

depends on the specified accuracy objectives and the desirable spatial representation (Olofsson et al., 2014). The sampling techniques are essential for generating high-quality crop maps because they ensure that the collected data accurately represent the diversity of crops across different regions. These techniques improve classification accuracy, minimize biases, and enhance the ability of machine learning models to distinguish between different crop types. Effective sampling also optimizes resource use, making data collection more efficient while maintaining reliability. In this study, we used four probabilistic sampling techniques to extract training samples for a crop type reference dataset used in the classification:

- **The grid (or systematic) sampling technique** selects samples at a regular distance interval between sampling points, forming a grid-like pattern. Sampling points are spaced at a fixed distance (grid spacing) in both the x (horizontal) and y (vertical) directions. The grid ensures that the entire interest region is covered systematically, avoiding clustering of points. To reduce bias, the first sampling point is often randomly selected, and the grid is then applied from that starting location and is useful to minimize the spatial correlation among the samples (Gallego, 2005). It can generate meaningful results when the area has homogeneous classes because it ensures an effective spatial distribution of the samples with relatively simple

Table 4
Number of samples per crop class used to evaluate the impact of different sampling techniques and sample sizes for crop type mapping.

Crop type	Minimum size	Equally distributed	0.05 %*	0.10 %	0.50 %	1.00 %
Corn	100	957	269	538	2,690	5,380
Cotton	100	957	401	801	4,002	8,003
Rice	100	957	100	145	724	1,448
Sorghum	100	957	100	100	100	100
Soybeans	100	957	2,300	4,600	22,998	45,997
Wheat	100	957	100	100	194	387
Other	100	957	3,430	7,115	36,288	72,678
Total	700	6,700	6,700	13,399	66,996	133,993

* Number of samples used to evaluate the impact of different sampling techniques and to create an independent set of data to analyze the accuracy of the classification model.

implementation. Also, grid sampling is often used in combination with other techniques to enhance their capabilities, especially random sampling. Although it works well for large, homogeneous areas where variability is low, the systematic sample selection does not assure that small crop areas will be adequately represented in the samples (Stehman, 2009). Application for crop type mapping using

this method can be found in Khan et al. (2010), and Yan et al. (2021). In this study, the distance between samples was defined based on the equal distribution of points, and it changed according to the total number of desired samples within the HLS tile dimensions (Table 4). For instance, the distance between points was 1347 m when working with 0.05 % of the sample size. The rationale for the spatial

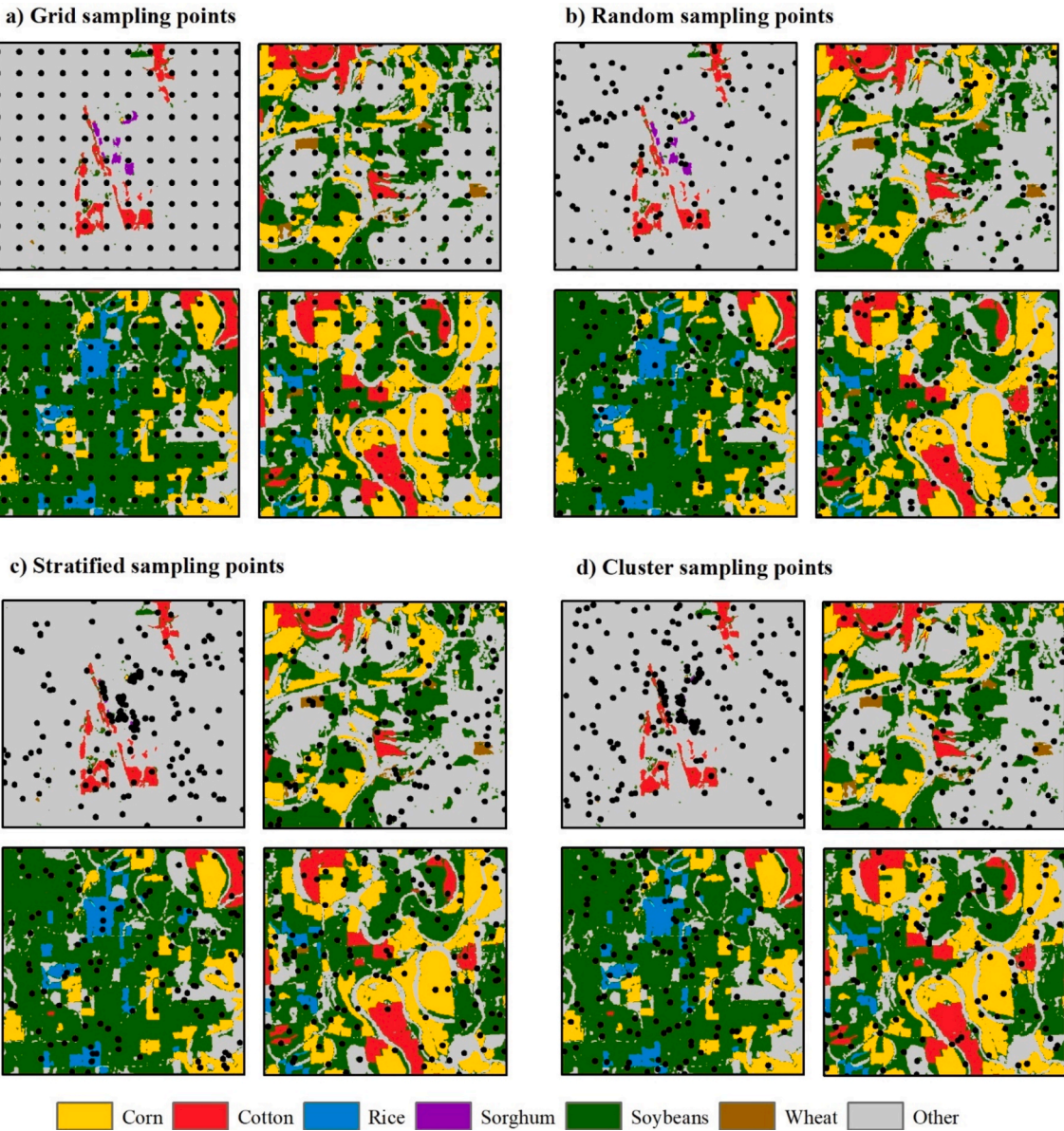


Fig. 5. Spatial distribution of training sample locations under different sampling techniques for different regions of the HLS-tile.

distribution of the sample points (generated from each sampling technique) for regions with crop-type classes is presented below and in Fig. 5.

- **The random sampling technique** generates randomly distributed points within an interest region, and it is designed to eliminate subjective bias in the sample collection process (Gallego, 2005). The process begins by defining the geographic boundaries of the region of interest and determining the sample size based on statistical or practical considerations. This technique is one of the easiest applied sampling options, but it may encounter challenges in achieving spatially distributed samples and adequately representing small land use and land cover classes (Stehman, 2009). Applications using the random sampling design for crop classification are presented in Xiao et al. (2023), and Goodbody et al. (2023). In this work, the samples were randomly distributed across the study area, considering the dimensions of the HLS tile and the total number of samples presented in Table 4.
- **The stratified random sampling** works by dividing the study area into distinct strata based on specific characteristics, such as the different land use and land cover classes in reference maps, to ensure samples' representation of all existing classes (Stehman, 2009). After stratification, the samples are randomly created within the strata until the number of samples per class pre-defined is archived. This is a practical design that satisfies the basic accuracy assessment and sample representation criteria by increasing the sample size in classes that occupy small portions of the interest region (Olofsson et al., 2014). However, this technique requires a reference map or a strategy to ensure that the strata are divided properly. This sampling technique has been applied in several crop type mapping studies and can also be combined with other sampling techniques (Castro et al., 2022; Khan et al., 2021; King et al., 2017; Yu et al., 2023), including cluster sampling in a two-stage sampling design (Li et al., 2023). In this study, the 2022 CDL layer was used to delineate the individual crop classes and generate the sample points. Since stratified random sampling is recommended as an effective practice for map accuracy (Olofsson et al., 2014). It was also adopted to generate an independent testing dataset for model validation, as we need to ensure that all classes are represented.
- **Cluster sampling technique** works by identifying areas with similar characteristics, representing the same class (stratum), and meeting a criterion of block size for a pre-defined number of pixels (e.g., a 100×100 -pixel block). If the criteria are met, a sample point is created to represent this block (Olofsson et al., 2014). Clusters are selected randomly, ensuring that the samples are chosen without bias. The primary purpose of clusters is to reduce the number of samples, as they represent a portion of the landscape with similar characteristics (Zimmerman et al., 2013). However, special attention is necessary when choosing the cluster block; depending on the configuration, it might affect crop type representation. Additionally, the quality of the reference map is crucial for accurately identifying pixels belonging to the same class. A combination of cluster and stratified sampling techniques is often used as a two-stage sample design to ensure spatial representation, although this approach increases the process complexity (Olofsson et al., 2014). This technique has been used for crop-type mapping and effective results were obtained by Li et al. (2023) and Song et al. (2017). In this work, we used a two-stage cluster sampling design. The CDL was used to identify regions with pixels belonging to the same class and informed adequate spatial distribution of samples, which serves as an indicator of representativeness for the crop type class. To create the samples, the cluster configuration used a block with at least 100 pixels of the same classes and a minimum of 5 clusters together with a density of 0.50 (calculated as the ratio of the cluster size to the total number of pixels in the stratum). This configuration was used to guarantee that rare crop classes would be properly sampled.

3.3. Sample size experiment

In this study, we evaluated six sample sizes per crop type class to train the proposed ANN model, as presented in Table 4. The percentage values represent a portion of the total number of available pixels in the HLS image area (13,399,260 pixels), and the minimum size was defined based on the recommendation of Congalton (1991), which suggests a minimum value of 100 samples. For cases where this suggestion was not met, we considered a proportion based on the representation of the crop type class (Table 1), ensuring that the minimum value should be at least 100 samples. To evaluate the impacts of different sampling techniques on the classification model, we used only 0.05 % of the sample size with varying values among the crop classes. It is important to note that only the stratified random sampling and the cluster random samples guarantee that each class will have the specified number of samples (Table 4). The grid and random sampling will only distribute the total number of samples within the HLS tile. To assay sample size effects, we examined sample sizes of 0.05 %, 0.10 %, 0.50 %, and 1.00 %, with different proportions between the classes. We also generated an equally distributed number of samples using 0.05 % and a minimum size of 100 samples per class. The percentage of sample size choice reflects the feasibility of analysts collecting samples when the ground reference values are not available.

3.4. Machine learning crop type classification

A multi-layer perceptron algorithm, also known as a feed-forward neural network, was proposed in this study. This ANN model integrates an input layer, an output layer, and one or more hidden layers and learns non-linear relationships in the dataset along the network (Schmidhuber, 2015). Each layer has a set number of neurons and interconnected nodes with adjustable weights and biases. The hidden layers use nonlinear activation functions, which allow the network to approximate different functions (Svozil et al., 1997). During training, the network uses a loss function to measure the error between its predictions and the actual values. The back-propagation algorithm then calculates the gradient, which shows how each weight and bias should change to reduce the error. Using gradient descent, the network makes small adjustments to these parameters in each training step, gradually lowering the error until the predictions closely match the actual values (Kumaraswamy et al., 2024). These neural network models are widely used in crop type mapping (Kumar et al. (2015), Sonobe et al. (2017), Sonobe (2019), and Sildir et al. (2020)). In our study, two hidden layers were incorporated with 64 and 32 neurons. Rectified Linear Unit (ReLU) activation was used in the hidden layers to introduce non-linearity, while the output layer employed the softmax activation function for class probability estimation. The model was optimized using the Adam optimizer and categorical cross-entropy loss. Categorical Crossentropy was used as the loss function, which is appropriate for multiclass classification when labels are one-hot encoded. The neural network model was trained using 50 epochs and batch size of 32 and a learning rate of 0.001, achieving a balance between computational efficiency and convergence. To avoid overfitting, 0.20 % of the training samples were used in validation during the training process to inspect the model's performance. Note that several machine learning algorithms are feasible for studies similar to ours. We tested random forest, support vector machine, and light gradient boosting machine models, and the results were comparable for each sampling technique used (Appendix A), with the ANN model slightly outperforming the others. In this work, we presented only the result for the ANN because our main goals were to evaluate the (i) influence of sampling technique and (ii) sample size on classification performance, rather than the evaluation of different classifiers and machine learning algorithm structuring.

3.5. Accuracy assessment

A confusion matrix and four accuracy metrics (overall accuracy (OA), producer's accuracy (PA), user's accuracy (UA), and F score) were calculated to evaluate the classification performance (Congalton, 1991, Stehman, 2009). An independent set of samples was generated to test the ANN model using 0.05 % of the total pixels from the HLS tile (Table 4) using stratified random sampling as recommended by Olofsson et al. (2014). The OA offers a general metric of the model's performance by comparing the total number of correctly classified samples to the total number of samples. The UA measures how well the predicted pixels for each crop class align with independent reference data. It reflects the reliability of classification indicating the proportion of correctly classified pixels out of all pixels predicted for that class. A high UA suggests that most of the predicted pixels for a crop type are truly that crop, reducing commission errors. The PA evaluates how well the reference data matches the model's predictions, representing the proportion of actual pixels of a given crop that were correctly classified. A high PA indicates that most reference pixels for a crop were correctly identified, minimizing omission errors. The F1-score is the harmonic mean of UA and PA, balancing precision and recall to provide a single performance measure for each class. It is particularly useful when there is an imbalance in class distribution. The average F1-score, calculated across all crop classes, was used to assess the overall performance of the classification model, ensuring a comprehensive evaluation of its accuracy (Zhong et al., 2019). The OA, PA, UA, and F1-score were calculated below, respectively.

$$OA = N_C/N_A \quad (1)$$

$$UA = N/C \quad (2)$$

$$PA = N/R \quad (3)$$

$$F1 - score = 2 * [(UA * PA) / (UA + PA)] \quad (4)$$

where N denotes the number of correctly identified validation samples of a class, R denotes the number of validation samples of a class, C denotes the number of validation samples classified as a class, and N_C and N_A denote the number of correctly identified and total number of validation samples. The values obtained in each metric range from 0 to 1, where values close to 1 are indicative of high accuracy for crop maps.

4. Results

4.1. Overview of spectral variability among the crop types

Fig. 6 illustrates the principal component analysis (PCA) and linear discriminant analysis (LDA) for different crop types, using the extracted values of the 93 features employed to train the ANN model as input data. This highlights the importance of using different features to distinguish among crop types. For instance, the PCA shows that most classes are clustered, with the "other" and soybeans classes exhibiting the highest variability, making them prone to misclassification with the remaining classes. The LDA demonstrates that cotton and wheat may have different spectral responses, potentially reducing misclassification between these two categories. Additionally, Fig. 7 presents the boxplot of the important features for crop mapping used in this study. It demonstrates that the distributions of EVI and NIR are distinct among the crop types, suggesting that they exhibit different responses to varying canopy structures and vegetation vigor. These features can potentially provide valuable information for the classification model. Similarly, the NDTI can help distinguish regions with crop residues and areas with low reflectance, such as bare soil. The NDWI is useful for water masking, where positive values indicate water surfaces. NDTI and NDWI exhibited higher average values for the "other" class than the crop-type classes, making them important features to assist in classifying this group, as it represents 54 % of the study area.

4.2. Evaluation of crop type classification using different sampling techniques

Fig. 8 presents the confusion matrix obtained for each sampling technique, including grid sampling, random sampling, stratified sampling, and cluster sampling, used for crop type mapping. Note that we adopted 0.05 % (6,700 samples) of the total number of pixels in the HLS tile to define the number of samples that were used for ANN training. An independent dataset using the stratified sampling technique with 0.05 % samples was also used to test the model's performance. The main diagonal represents correctly classified pixels for each class, while the off-diagonal values indicate pixels classified among incorrect classes. In general, we observed that soybeans and "other" classes were well represented by all sampling techniques, given their higher representation in the study area. In contrast, the sorghum and wheat classes were poorly represented when samples were extracted using grid and random sampling techniques. We observed that when grid sampling was used, no pixel of the sorghum class was correctly classified and only 19 out of 100 pixels of the wheat class were correctly classified. Using random sampling, only four pixels were classified as sorghum. Among the crop

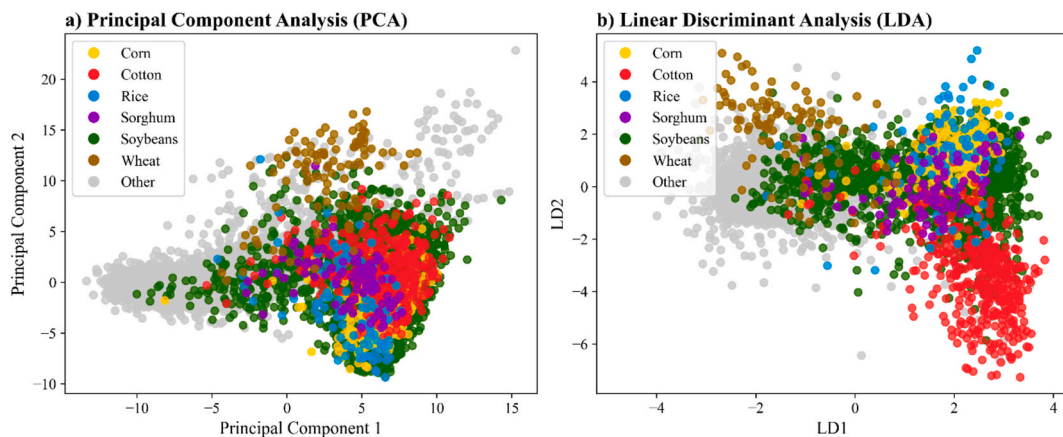


Fig. 6. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) of the crop classes evaluated. The analyses were conducted using the extracted values of the 93 features employed to train the ANN model as input data.

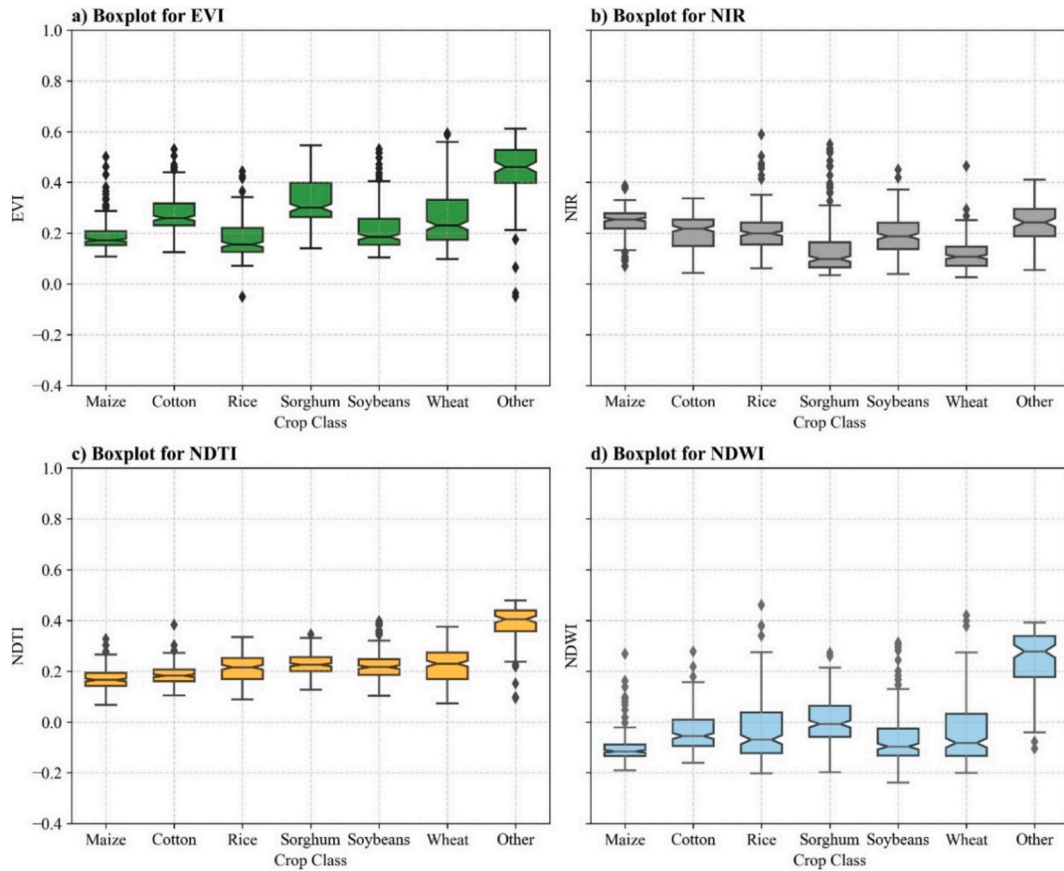


Fig. 7. Boxplots of the spectral responses of important features used to differentiate crop types and other land covers derived from an average composition from the 93 input features: (a) average Enhanced Vegetation Index (EVI); (b) average Near Infrared (NIR); (c) average Normalized Difference Tillage Index (NDTI); (d) average Normalized Difference Water Index (NDWI).

classes, soybeans had the highest number of misclassified pixels, (often with the “other” and corn classes) in all sampling techniques. Wheat and sorghum were often confused with “other” and soybeans classes. Rice class had most pixels classified correctly, indicating that the classification model effectively distinguished the patterns of this crop across all sampling techniques. Corn and cotton had a strong representation; however, the misclassification pixels for these classes were mainly with soybeans.

Table 5 shows the overall accuracy, producer’s accuracy, user’s accuracy, and F1-score derived from the classified results using different sampling techniques. The stratified and cluster sampling techniques exhibited superior performance in most of the four metrics, with the stratified sampling technique slightly outperforming the others, achieving an average F1-score equal to 0.81. The random sampling technique achieved an average producer’s accuracy of 0.85, but there is a bias effect caused by the sorghum classes that reached 100 % producer’s accuracy. It is important to use metrics other than overall accuracy to evaluate the model because the user’s accuracy, average producer’s accuracy, and the average F1-score penalize models that poorly classified certain classes.

Fig. 9 presents the evaluation metrics for the crop type classification results using different sampling techniques. For classes with a large area proportion in the HLS tile, such as soybeans (34 %) and “other” (54 %), all four sampling techniques were able to provide sufficient information for the ANN training, resulting in accuracy metrics above 0.80. As these land cover classes have a large area, the methodology employed for the sampling technique does not have a significant impact on creating the sample points, as the chance of a class representation is high. The classes with poor performance were sorghum and wheat, which present a small proportion of the total area, 0.02 % and 0.26 %, respectively. For these

classes, the grid and random sampling techniques performed poorly because the allocation of samples from these methods does not guarantee that all classes would have sufficient samples to train the model effectively. The stratified and cluster methods were able to capture samples for these classes and performed well in all classes in general. However, for the rice class, the cluster sampling technique presented the worst results compared to the other methods, with a user’s accuracy of 0.63 and an F1-score of 0.71. In contrast, the stratified sampling technique reached 0.81 for both metrics. Since this sampling technique performed well in all cases and is relatively simple to implement, the stratified sampling technique was selected to evaluate the impacts of different sample sizes for crop classification.

Fig. 10 compares the classification maps obtained for each sampling technique with the reclassified CDL base map used as ground truth. Most of the crop types were correctly classified by the ANN model, especially when using the stratified (Fig. 10d) and cluster (Fig. 10e) sampling techniques. However, the pixel-based classification typically presents a salt-and-pepper noise effect, and it is possible to notice that this effect occurred between soybeans and the “other” class, and this problem can be observed in the classified images where the noise is more concentrated between these classes.

4.3. Impact of sample sizes on crop type classification

The sample sizes of minimum size, equal number of samples, 0.10 %, 0.50 %, and 1.00 % of the total number of available pixels in the HLS tile, along with a minimum size of 100 samples per class, were used to evaluate the impact of different sample sizes on the classification performance. The sampling proportion of 0.05 % of the total number of pixels of the HLS was used to evaluate the sampling techniques in the

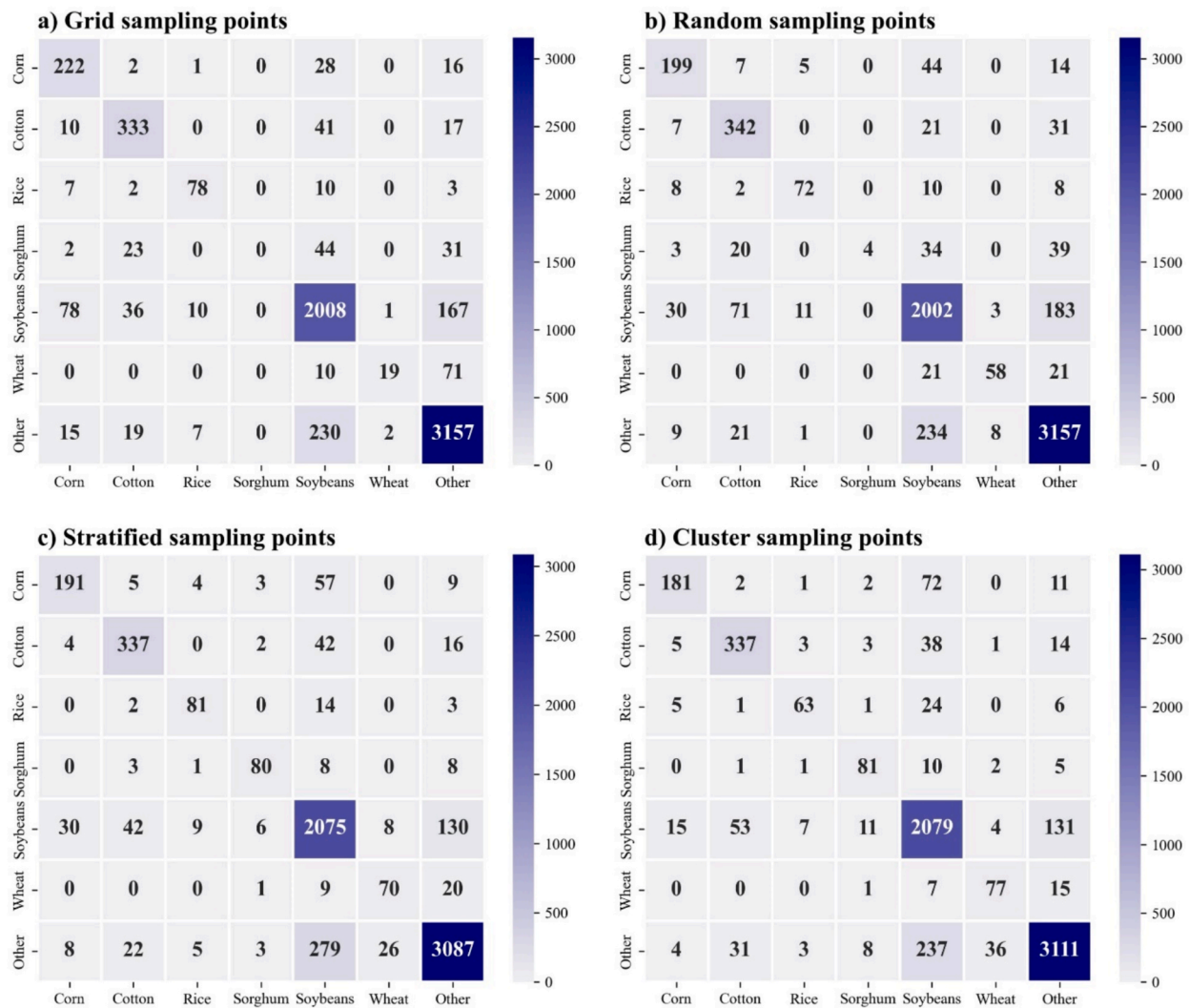


Fig. 8. Confusion matrices for crop classification results based on each sampling technique.

Table 5

Overall accuracy, producer's accuracy, user's accuracy, and F1-score for the classified crop type results obtained from different sampling techniques.

Sampling technique	Overall accuracy	Producer's accuracy	User's accuracy	F1-Score
Grid	0.87	0.70	0.63	0.63
Random	0.87	0.85	0.67	0.69
Stratified	0.88	0.82	0.81	0.81
Cluster	0.88	0.81	0.79	0.79

previous section of this paper; for this reason, we tested this proportion of samples distributed equally among the classes (957 samples per class).

Fig. 11 presents the confusion matrix obtained using different sample sizes. The strongest impact on the classification model occurred when using a minimum sample size of 100 samples and an equal distribution of 0.05 % of samples per class. In both cases, the class "other" did not have any correctly classified pixels; instead, the model confused it with all classes, especially sorghum (Fig. 11a) and wheat (Fig. 11b), which represent small areas. Increasing the number of samples to 0.10 %, 0.50 %, and 1.00 % improved the model's capability to distinguish the crop types. The most misclassified pixels using these sample sizes were soybeans and "other", regardless of the percentage used, but the 0.50 % sample size resulted in fewer misclassified pixels (Fig. 11d). Considering

these three sample sizes, rice had the highest number of correctly classified pixels, indicating that the classification model was able to capture the spatial distribution of this class with the number of samples provided. Corn and cotton also had relatively effective classification.

Table 6 presents the accuracy metrics for the crop type mapping results using different sample sizes. Despite the increase in the number of samples from 0.10 % (13,399) to 1.00 % (133,399), there is little accuracy gain, especially between 0.50 % and 1.00 % of the sample size. The combination of using a minimum size of 100 samples per class or 0.05 % of the total number of pixels in the HLS tile equally distributed between the classes was not effective in capturing the spatial distribution and spectral response of the classes. Thus, the ANN model performed poorly.

Using 0.10 %, 0.50 %, and 1.00 % of the total number of pixels in the HLS tile, most classes achieved >0.80 for all accuracy metrics (Fig. 12). However, the user's accuracy ranged from 0.73 to 0.78 for corn and from 0.71 to 0.77 for sorghum when using 0.50 % and 0.10 % of samples. Due to the large number of samples used among these three percentage values, a trade-off between the desired accuracy and the number of samples should be considered, as increasing the number of observations did not result in a substantial improvement in the model's performance. Considering a minimum sample size of 100 or equally distributing 0.05 % of the total number of HLS pixels only provided good spatial information for the ANN model for the class soybeans, with values above 0.70 for all metrics. Thus, the sampling size should be proportional to the

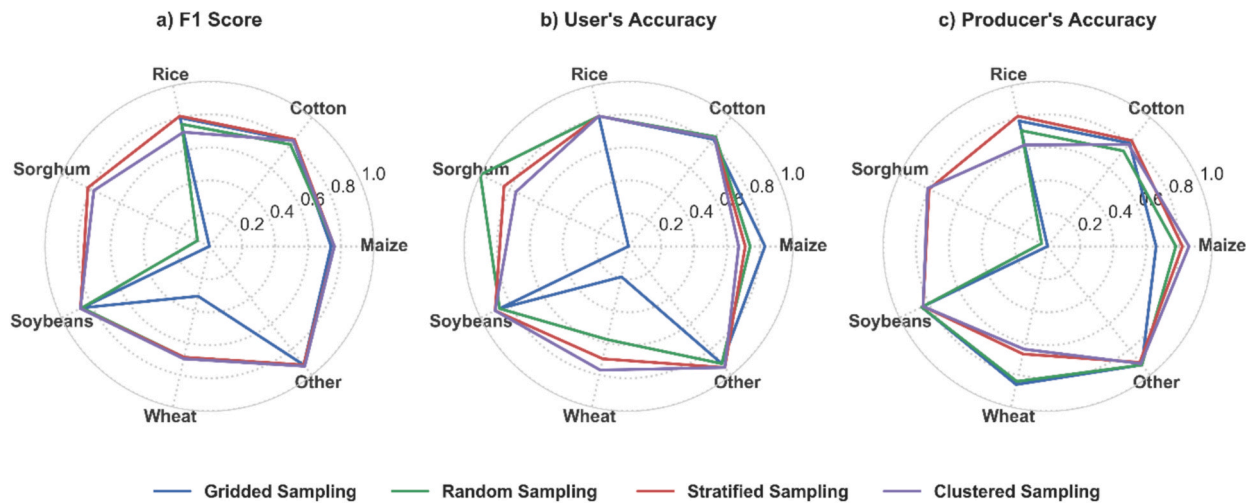


Fig. 9. Accuracy metrics for each class were obtained from the classified results using the ANN model for grid, random, stratified, and cluster sampling techniques.

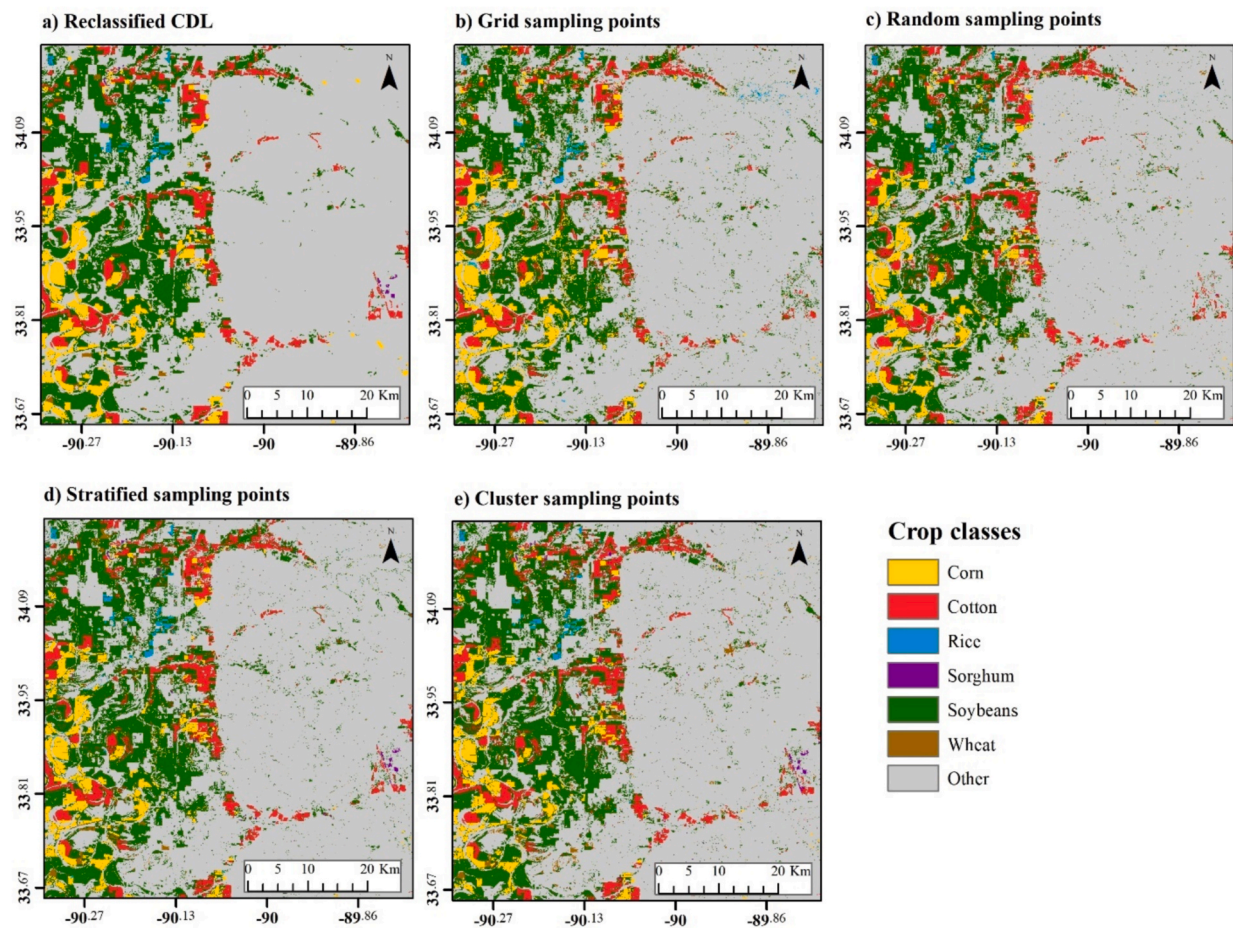


Fig. 10. Crop-type classification results from the ANN model using different sampling techniques.

areal extent.

Fig. 13 shows the crop-type maps using different sample sizes and the stratified sampling technique. Compared to the reclassified CDL product (Fig. 13a), the sample sizes of 0.10 %, 0.50 %, and 1.00 % effectively captured the spatial representation of the classes. Additionally, the salt-and-pepper noise effect was reduced as the number of samples increased, as shown in Fig. 13d, e, and f. Notably, the number of pixels classified as soybeans decreased in the highlighted region. Salt-and-

pepper noise occurs in pixel-based classification because each pixel is analyzed individually and assigned a class based on its reflectance. Variations in pixel reflectance can lead to misclassification, resulting in scattered, isolated pixels that differ from their surroundings. Using a sample size of 0.05 %, equally distributed between classes, and a minimum size of 100 samples per class resulted in significant confusion between classes. The “other” class was primarily classified as sorghum and wheat (Fig. 13c) or soybeans (Fig. 13b), which are classes that

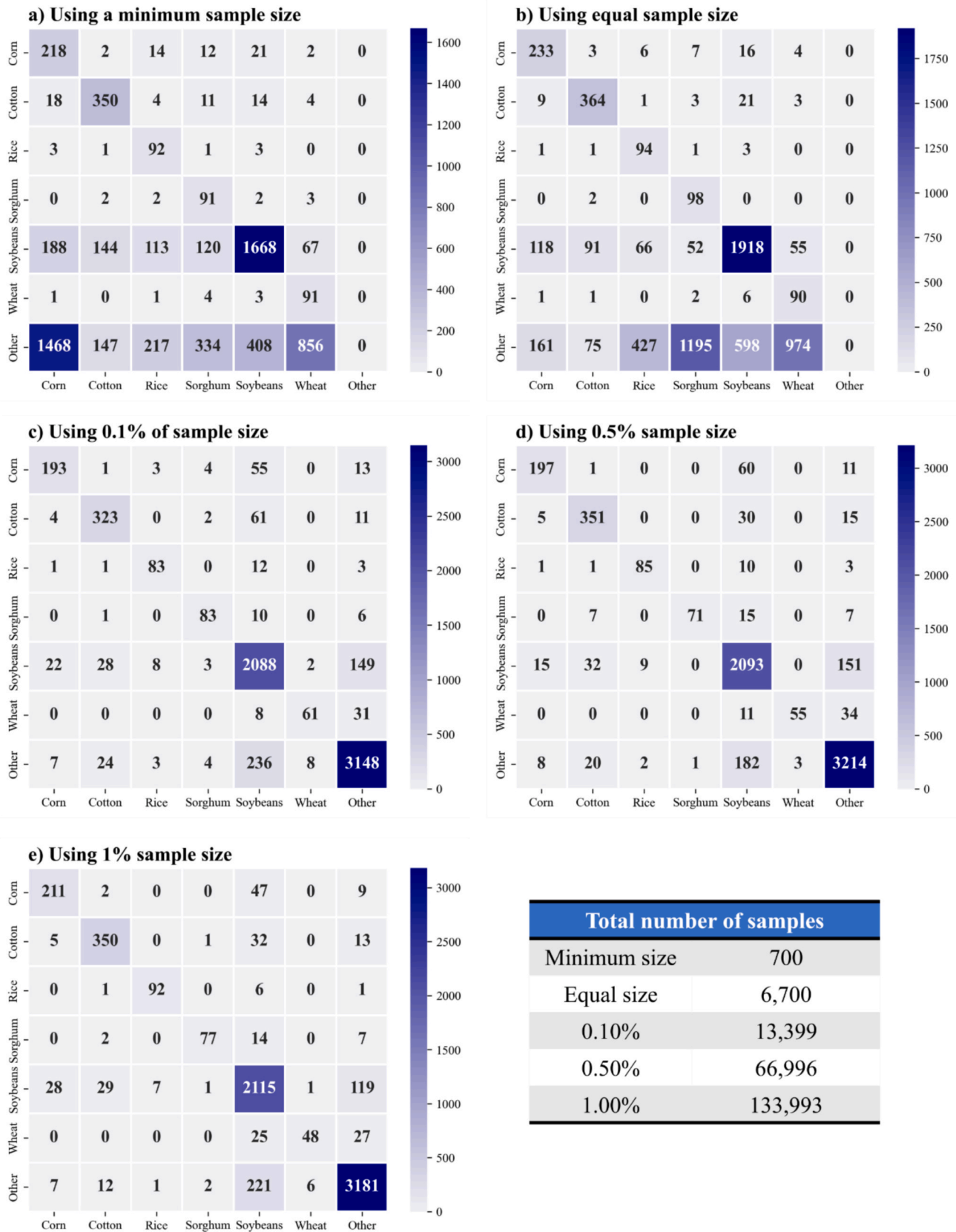


Fig. 11. Confusion matrix for the classified images from the ANN model using different sample sizes and stratified sampling technique.

Table 6

Overall accuracy, average user's accuracy, average producer's accuracy, and the average F1-score for crop type mapping obtained using different sample sizes.

Sample size	Overall accuracy	Average producer's accuracy	Average User's accuracy	Average F1-score
Minimum	0.37	0.27	0.74	0.34
Equal size	0.42	0.31	0.78	0.39
0.10 %	0.89	0.87	0.80	0.83
0.50 %	0.91	0.91	0.79	0.84
1.00 %	0.91	0.90	0.81	0.84

occupy a small proportion of the area in the HLS tile.

The difference in sample size directly impacts the quality of classification due to its influence on the model's ability to capture the spatial and spectral variability of the crop types. Larger sample sizes typically provide a more comprehensive representation of the diverse conditions within the study area, reducing classification errors and enhancing accuracy metrics. However, as observed in this study, beyond a certain threshold, increasing the sample size does not necessarily lead to significant improvements in classification performance. This suggests that while sufficient samples are necessary to represent variability, excessive sampling may not add new information to improve the model's training process. On the other hand, smaller sample sizes, though computationally efficient, may not capture the full variability of crop classes, particularly for minority classes or crops with smaller areal extents, such as wheat and sorghum, leading to reduced accuracy metrics. For instance, the lower user's accuracy observed for wheat and sorghum at smaller sample sizes indicates that the model struggled to differentiate these crops due to inadequate representation. Additionally, sample size proportionality to the area extent of each crop type is critical to ensure balanced class representation, as imbalances can bias the model toward more dominant classes, reducing classification reliability for less-represented crops.

5. Discussion

Our study implemented a series of experiments with different sampling techniques to classify crop types using multi-temporal HLS composites. Multi-temporal HLS composites were calculated, and spectral metrics (percentiles 10 %, 25 %, 50 %, 75 %, 90 %) and coefficient of variation were derived. The input dataset consisted of 93 spectral features, and the output was the main crop labels (corn, cotton, rice, sorghum, soybeans, and wheat) derived from the reclassified CDL image. Different sampling techniques and sample sizes were used to train the

ANN model. This study provides strategies for agriculture monitoring and policymakers to create high-quality samples to develop crop maps, which are fundamental for several agricultural activities, particularly for proper agricultural planning and food security. The crop type maps' overall accuracy ranged from 0.37 to 0.91 depending on the sampling technique and sample size used. Several studies have presented similar performance of crop type mapping using multi-temporal spectral metrics (Dash et al., 2023; Li et al., 2023; Zhong et al., 2014). Li et al. (2023) achieved 92 % overall accuracy when mapping soybeans and corn using 10-meter resolution Sentinel-2 data. Acharki et al. (2023) achieved 95 % overall accuracy when mapping land use and land cover classes, including crops such as wheat, corn, and rice, using 3-meter resolution PlanetScope data. Our results illustrated the typical issues with pixel-based classification, such as salt-and-pepper effects and the misclassification of rare classes. This corroborates with previous studies that are implemented in large areas using standard machine learning models such as random forest (Tran et al., 2022) and artificial neural network (Kumar et al., 2015).

Our findings revealed that stratified random and cluster sampling techniques presented the best performances in the crop type maps, with an overall accuracy of 88 %, which is consistent with the results of Dash et al. (2023), who mapped crop types in the YMD region and obtained an overall accuracy of 84 % using Landsat images. This is particularly important when the study area has an unbalanced representation of certain classes (Stehman & Wagner, 2024), which was the case with sorghum and wheat crops in this study, and the grid and random sampling techniques did not capture a representative number of samples for these classes. These techniques create samples based on grid distance and a randomly defined starting point, respectively, which does not guarantee that all classes will have representative samples. Relevant findings using stratified and cluster sampling techniques for crop type mapping can be found in Castro et al. (2022), King et al. (2017), Li et al. (2023), Song et al. (2017), and Yu et al. (2023). By exploring the model performance for each crop type, we observed that classes with a significant proportion of the study area, such as soybeans and "other," were well mapped, regardless of the sampling techniques. The mapping results using the cluster sampling technique were inferior for the rice class compared to other methods, which may be attributed to the lack of representative samples across the area since each block is represented by only one sample (Zimmerman et al., 2013). Since the stratified random sampling performed well for all crop classes, we selected it to evaluate the impact of different sample sizes.

While there is no consensus or standard practice for a minimum number of samples, other studies suggest 75 to 100 samples as a minimum sample size if the classification involves a large number of classes

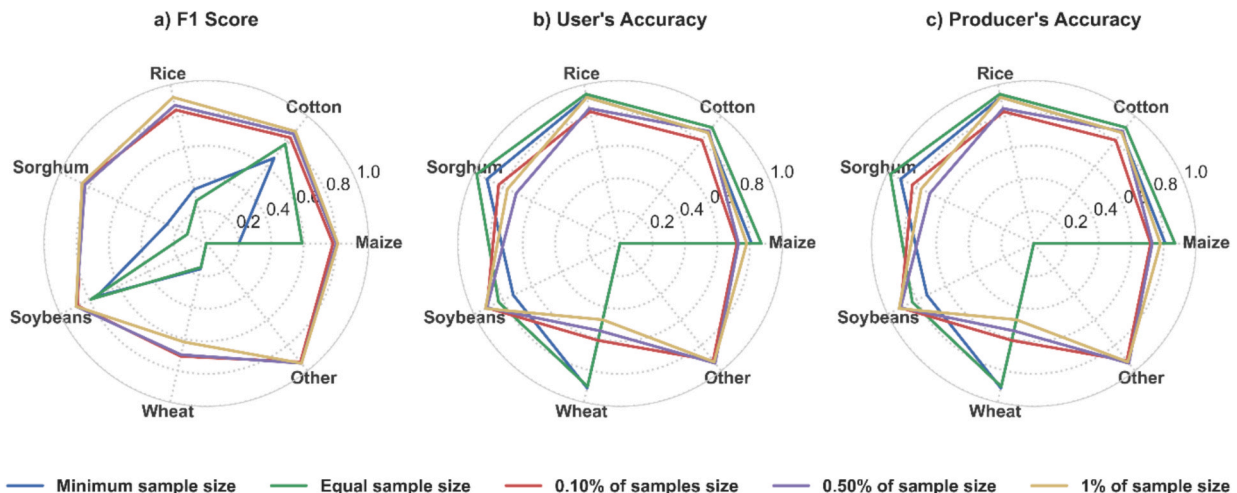


Fig. 12. Evaluation metrics for each class were obtained from the classified image using the ANN model for different sample sizes and stratified sampling technique.

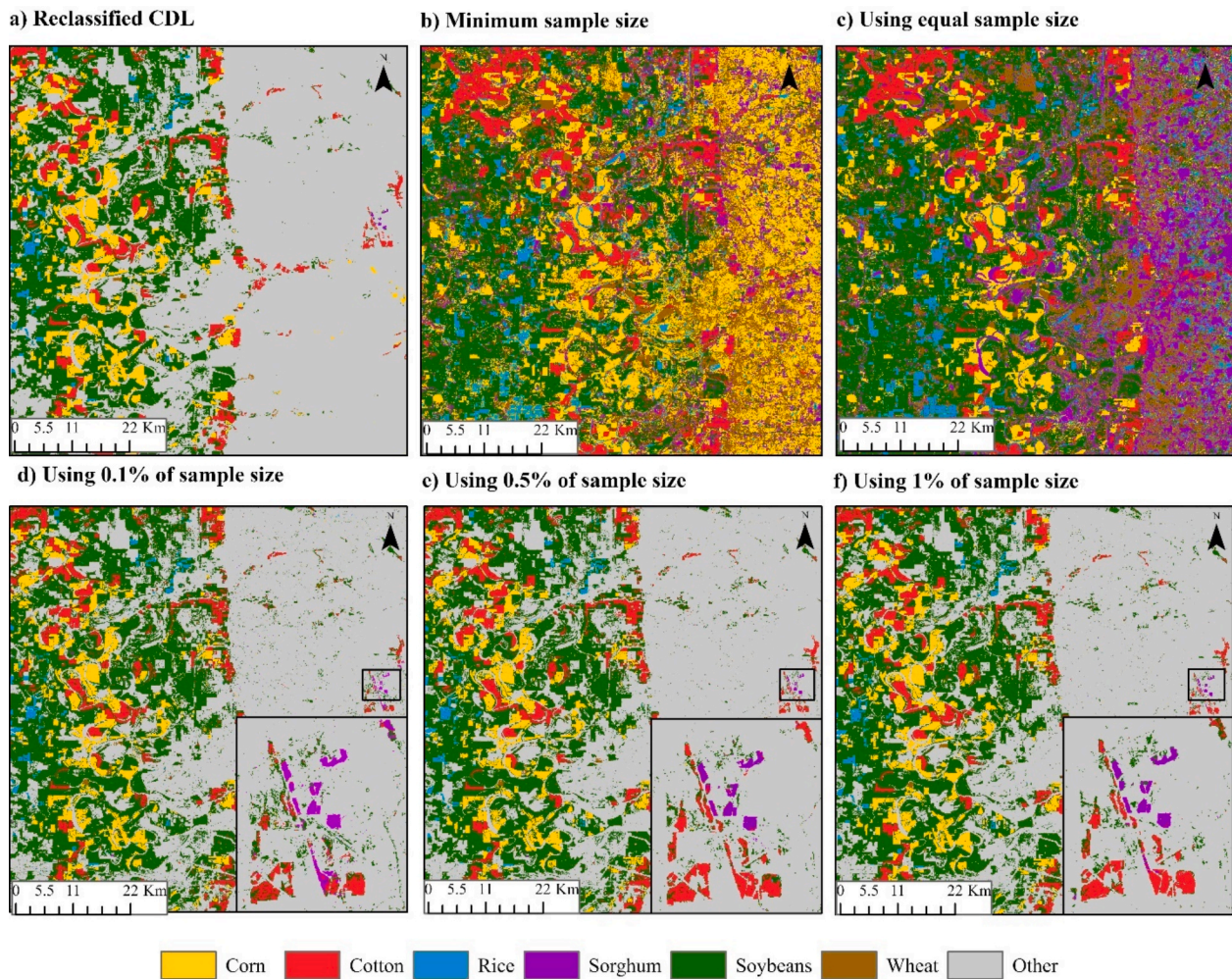


Fig. 13. Crop-type classification maps using ANN model trained with different sample sizes. (a) Reclassified CDL as reference map, (b) minimum sample size (700 samples), (c) 0.05% (6,700 Samples) equally distributed between crop classes, (d) 0.1% (13,399 Samples), (e) 0.5% (66,996 samples), and f) 1% (133,993 samples). It highlights the salt-and-pepper effect in sample sizes of 0.1%, 0.5%, and 1% of the total number of pixels in HLS images.

(Congalton, 1991; Wagner & Stehman, 2015). The number of samples also impacts the number of features used to train the model in its capability of generalization in new data. In regression models, for example, a proportion of 20 observations of one feature is recommended (Hair et al., 2010). By increasing the proportion of training samples from 0.05 % (6,700 samples) to 1.00 % (133,993 samples) of the total number of pixels in the study area, we observed improvements in the classification accuracy and reduction of the salt-and-pepper effects. However, a trade-off between mapping accuracy and sample size should be considered because reference samples are time-consuming to obtain and a large dataset demands high computer performance (Zhang et al., 2020). Additionally, (Moraes et al., 2021) demonstrated that the model performance can be negatively affected by increasing the number of samples depending on the sample representation. They showed that the mapping accuracy was reduced by ~ 2.00 % when the training samples increased from 2000 to 6000. The sample size plays a major role in land use and land cover classification as well and has been a concern reported in many studies (Gao et al., 2023; Hao et al., 2020; Li et al., 2023; Stehman & Wagner, 2024; Wagner & Stehman, 2015). The number of samples allocated for each class can lead to underestimation or overestimation of certain classes (Olofsson et al., 2014). We also observed this issue when an equal distribution of the number of samples per class was used. This approach resulted in the overestimation of class representing a small proportion of area, such as sorghum and wheat, and the underestimation of the “other” class. For this reason, it is recommended

to use a proportional distribution of samples across classes, ensuring their representativeness within the study area.

The CDL was used as a ground reference to allocate the samples for each crop class, and this product is one of the most reliable crop type information for the United States (Boryan et al., 2011) and has been used in other crop type mapping studies (Tran et al., 2022; Wang et al., 2023b; Xu et al., 2024). Although this product is fully assessed and has high accuracy for major crops (Boryan et al., 2012; Johnson, 2013), it is an automated product and still has issues related to misclassified and isolated pixels (Zhang et al., 2019), and preprocessing of these data is required before using them as a reference. For global crop-mapping applications, other datasets need to be explored to generate the reference data. Developing a refined ground reference database with high spatial resolution is a challenge for remote sensing since most applications demand ground observations to validate the models.

The operational HLS product substantially increased the number of available images and has the potential for crop mapping, as increased the number of images with clear pixels providing high-quality data for the model. In this study, seasonal image compositions were used to minimize the temporal gaps created by cloud and cloud shadow and capture the development stages of different crop types, our results were successful on the crop type mapping at 30 m resolution. Additionally, HLS images may not have the appropriate spatial resolution to capture crop type in small fields, and integrating HLS with higher spatial resolution sensors, such as PlanetScope, is an optional strategy to improve

the representation of the crop maps. Rao et al. (2021) demonstrated a gain of accuracy of 10 % when using PlanetScope combined with Sentinel 1 and Sentinel 2, benefiting the mapping of crops in small fields (<600 m²). Harmonized products with Sentinel-2 and PlanetScope should be considered, as they allow for obtaining high spatial resolution images with a reasonable revisit time, providing many images to minimize cloud effects.

6. Conclusions

This study evaluated the impacts of sample size and allocation strategies on crop type mapping accuracy, using multi-temporal HLS composites as the input data. We selected a portion of the Yazoo-Mississippi Delta as the study area because it represents a landscape with a diversity of crops, and the results could be extrapolated to other regions with complex crop systems. The experiments identified stratified random sampling as the most effective method for crop classification, achieving 88 % overall accuracy. This technique ensured the proportional representation of all crop classes within the study area, mitigating biases caused by class imbalance. Cluster sampling also performed well, however, it struggled with less representative classes, especially wheat, highlighting the importance of spatial sample allocation in sampling design for crop mapping.

The results demonstrated that using an equal sample allocation artificially inflated less representative class representation (e.g., sorghum) and underestimated dominant classes, whereas proportional allocation aligned results with ground truth distributions. Increasing the sample size improved classification accuracy and reduced noise. For instance, increasing the sample size from 0.05 % (6,700 pixels) to 1.00 % (133,993 pixels) of the total HLS image pixels improved the overall accuracy from 88 % to 90 %. However, users should evaluate whether this 2 % accuracy gain justifies the increased sample size, as larger datasets may not always provide proportionally valuable new information for the analysis.

This work provides actionable guidelines for crop mapping sample allocation techniques and sample size. While the HLS composites served as the data source, the findings are transferable to other multi-temporal remote sensing datasets. Future studies could extend this framework to integrate higher-resolution sensors or refine ground reference databases. Ultimately, this research advances methodological best practices for sample-driven crop classification, emphasizing that thoughtful sampling technique and sample size are pivotal to achieving accurate, scalable

agricultural monitoring. The lessons learned could be extended to a variety of studies and regions involving machine learning/deep learning and remote sensing applications that require strategies to create high-quality datasets to train the models.

CRediT authorship contribution statement

Uilson R.V. Aires: Writing – review & editing, Validation, Methodology, Formal analysis, Conceptualization. **Vitor S. Martins:** Writing – original draft, Visualization, Software, Methodology, Formal analysis. **Lucas B. Ferreira:** Writing – review & editing, Validation, Investigation, Data curation. **Yanbo Huang:** Writing – review & editing, Visualization, Investigation, Data curation, Conceptualization. **Lucas Heintzman:** Writing – review & editing, Visualization, Conceptualization. **Ying Ouyang:** Writing – review & editing, Visualization, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Uilson Ricardo Venancio Aires reports financial support was provided by US Department of Agriculture. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Funding was provided by the USDA Agricultural Research Service, project numbers 58-0200-0-002, and 58-6060-3-005, and the Data Science for Food and Agricultural Systems program, project award no. 2024-67021-42530 from the U.S. Department of Agriculture's National Institute of Food and Agriculture, and supported by USDA-ARS Project # 6060-13660-009-00D. Additionally, Dr. Martins was supported by USDA Hatch project under MIS-032130, and the Soil Health program. We want to thank the NASA HLS team for providing the available product and the USDA Cropland Data Layer. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. USDA is an equal opportunity provider, employer, and lender.

Appendix A. Evaluation of different machine learning model performances for various sampling techniques using a sample size of 0.05 % of the total number of pixels from the Harmonized Landsat and Sentinel-2 (HLS) image.

ANN				
Sampling techniques	Overall accuracy	Average producer's accuracy	Average user's accuracy	Average F1-Score
Gridded	0.87	0.70	0.63	0.63
Random	0.87	0.85	0.67	0.69
Stratified	0.88	0.82	0.81	0.81
Clustered	0.88	0.81	0.79	0.79
RF				
Sampling techniques	Overall accuracy	Average producer's accuracy	Average user's accuracy	Average F1-Score
Gridded	0.87	0.74	0.56	0.62
Random	0.87	0.76	0.57	0.63
Stratified	0.88	0.74	0.69	0.76
Clustered	0.88	0.73	0.69	0.75

(continued on next page)

(continued)

ANN				
Sampling techniques	Overall accuracy	Average producer's accuracy	Average user's accuracy	Average F1-Score
SVM				
Sampling techniques	Overall accuracy	Average producer's accuracy	Average user's accuracy	Average F1-Score
Gridded	0.88	0.75	0.60	0.65
Random	0.88	0.76	0.60	0.64
Stratified	0.89	0.76	0.74	0.80
Clustered	0.89	0.75	0.74	0.80
LightGBM				
Sampling techniques	Overall accuracy	Average producer's accuracy	Average user's accuracy	Average F1-Score
Gridded	0.77	0.44	0.49	0.46
Random	0.87	0.74	0.58	0.63
Stratified	0.89	0.75	0.71	0.77
Clustered	0.89	0.75	0.73	0.79

Data availability

The datasets generated and/or analyzed during the current study are available in the GitHub repository under an open license (https://github.com/uvaires/crop_classification). The repository contains the complete code used for data preprocessing, analysis, and visualization, sampling techniques used to create the crop type reference dataset, and documentation explaining how to run the scripts and reproduce the results.

References

- Acharki, S., Frison, P.-L., Veettil, B.K., Pham, Q.B., Singh, S.K., Amharref, M., Bernoussi, A.S., 2023. Land cover and crop types mapping using different spatial resolution imagery in a mediterranean irrigated area. *Environ. Monit. Assess.* 195 (11), 1309. <https://doi.org/10.1007/s10661-023-11877-4>.
- Becker-Reshef, I., Barker, B., Whitcraft, A., Oliva, P., Mobley, K., Justice, C., Sahajpal, R., 2023. Crop type maps for operational global agricultural monitoring. *Sci. Data* 10 (1), 172. <https://doi.org/10.1038/s41597-023-02047-9>.
- Belgiu, M., Csillik, O., 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sens. Environ.* 204, 509–523. <https://doi.org/10.1016/j.rse.2017.10.005>.
- Bolfe, É.L., Parreiras, T.C., Silva, L.A.P., Sano, E.E., Bettiol, G.M., de Victoria, D., Sanches, D., Vicente, L.E., 2023. Mapping agricultural intensification in the Brazilian savanna: a machine learning approach using harmonized data from landsat sentinel-2. *ISPRS Int. J. Geo Inf.* 12 (7), 263. <https://doi.org/10.3390/ijgi12070263>.
- Bolton, D.K., Gray, J.M., Melaas, E.K., Moon, M., Eklundh, L., Friedl, M.A., 2020. Continental-scale land surface phenology from harmonized landsat 8 and sentinel-2 imagery. *Remote Sens. Environ.* 240, 111685. <https://doi.org/10.1016/j.rse.2020.111685>.
- Boryan, C., Yang, Z., & Di, L. (2012). Deriving 2011 Cultivated Land Cover Data Sets Using USDA National Agricultural Statistics Service Historic Cropland Data Layers. *IEEE International Geoscience and Remote Sensing Symposium*, 6297–6300. doi: 10.1109/IGARSS.2012.6352699.
- Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US Agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto Int.* 26 (5), 341–358. <https://doi.org/10.1080/10106049.2011.562309>.
- Bueno, I.T., Antunes, J.F.G., Dos Reis, A.A., Werner, J.P.S., Toro, A.P.S.G.D.D., Figueiredo, G.K.D.A., Esquerdo, J.C.D.M., Lamparelli, R.A.C., Coutinho, A.C., Magalhães, P.S.G., 2023. Mapping integrated crop-livestock systems in Brazil with PlanetScope time series and deep learning. *Remote Sens. Environ.* 299, 113886. <https://doi.org/10.1016/j.rse.2023.113886>.
- Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., Li, Z., 2018. A high-performance and in-season classification system of field-level crop types using time-series landsat data and a machine learning approach. *Remote Sens. Environ.* 210, 35–47. <https://doi.org/10.1016/j.rse.2018.02.045>.
- Carrasco, L., Fujita, G., Kito, K., Miyashita, T., 2022. Historical mapping of rice fields in Japan using phenology and temporally aggregated landsat images in Google Earth Engine. *ISPRS J. Photogramm. Remote Sens.* 191, 277–289. <https://doi.org/10.1016/j.isprsjprs.2022.07.018>.
- Castro, P.L.B., Yin, H., Teixeira Junior, P.D., Lacerda, E., Pedroso, R., Lautenbach, S., Vicens, R.S., 2022. Sugarcane abandonment mapping in Rio de Janeiro state Brazil. *Remote Sens. Environ.* 280, 113194. <https://doi.org/10.1016/j.rse.2022.113194>.
- Clauss, K., Yan, H., Kuenzer, C., 2016. Mapping paddy rice in China in 2002, 2005, 2010 and 2014 with MODIS time series. *Remote Sens. (Basel)* 8 (5), 434. <https://doi.org/10.3390/rs8050434>.
- Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.-C., Skakun, S.V., Justice, C., 2018. The harmonized landsat and sentinel-2 surface reflectance data set. *Remote Sens. Environ.* 219, 145–161. <https://doi.org/10.1016/j.rse.2018.09.002>.
- Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* 37 (1), 35–46. [https://doi.org/10.1016/0034-4257\(91\)90048-B](https://doi.org/10.1016/0034-4257(91)90048-B).
- Coupe, R.H., Barlow, J.R.B., Capel, P.D., 2012. Complexity of human and ecosystem interactions in an agricultural landscape. *Environ. Develop.* 4, 88–104. <https://doi.org/10.1016/j.envdev.2012.09.009>.
- Dash, P., Sanders, S.L., Parajuli, P., Ouyang, Y., 2023. Improving the accuracy of land use and land cover classification of Landsat data in an agricultural watershed. *Remote Sens. (Basel)* 15 (16), 4020. <https://doi.org/10.3390/rs15164020>.
- Deventer, A.P.V., Ward, A.D., Gowda, P.H., Lyon, J.G., 1997. Using thematic mapper data to identify contrasting soil plains and tillage practices. *Photogramm. Eng. Remote Sens.* 63, 87–93.
- Duro, D.C., Franklin, S.E., Dubé, M.G., 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sens. Environ.* 118, 259–272. <https://doi.org/10.1016/j.rse.2011.11.020>.
- European Space Agency (ESA), 2015, July 24. *Sentinel-2 User Handbook*. https://sentinels.copernicus.eu/web/sentinel/user-guides/document-library/-/asset_publisher/xslst4309D5h/content/sentinel-2-user-handbook.
- Farhadpour, S., Warner, T.A., Maxwell, A.E., 2024. Selecting and interpreting multiclass loss and accuracy assessment metrics for classifications with class imbalance: guidance and best practices. *Remote Sens. (Basel)* 16 (3), 533. <https://doi.org/10.3390/rs16030533>.
- Feng, F., Gao, M., Liu, R., Yao, S., Yang, G., 2023. A deep learning framework for crop mapping with reconstructed sentinel-2 time series images. *Comput. Electron. Agric.* 213, 108227. <https://doi.org/10.1016/j.compag.2023.108227>.
- Gallego, F.J., 2005. Stratified sampling of satellite images with a systematic grid of points. *ISPRS J. Photogramm. Remote Sens.* 59 (6), 369–376. <https://doi.org/10.1016/j.isprsjprs.2005.10.001>.
- Gao, B.-C., 1996. NDWI: a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* 72, 257–266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3).
- Gao, F., Zhang, X., 2021. Mapping crop phenology in near real-time using satellite remote sensing: challenges and opportunities. *J. Remote Sens.* 2021. <https://doi.org/10.34133/2021/8379391>.
- Gao, Z., Guo, D., Ryu, D., Western, A.W., 2023. Training sample selection for robust multi-year within-season crop classification using machine learning. *Comput. Electron. Agric.* 210, 107927. <https://doi.org/10.1016/j.compag.2023.107927>.
- Gilbertson, J.K., Kemp, J., van Niekerk, A., 2017. Effect of pan-sharpening multi-temporal landsat 8 imagery for crop type differentiation using different classification techniques. *Comput. Electron. Agric.* 134, 151–159. <https://doi.org/10.1016/j.compag.2016.12.006>.
- Goodbody, T.R.H., Coops, N.C., Senf, C., Seidl, R., 2023. Airborne laser scanning to optimize the sampling efficiency of a forest management inventory in south-eastern Germany. *Ecol. Ind.* 157, 11281. <https://doi.org/10.1016/j.ecolind.2023.11281>.
- Goddard Space Flight Center (GSFC), 2018, December. *Harmonized Landsat Sentinel-2 (HLS) Product User's Guide*. https://lpdaac.usgs.gov/documents/1007/HLS_User_Guide_V15_provisional.pdf.
- Hair, J., Black, B., Babin, B., Anderson, R.E., 2010. *Multivariate Data Analysis*, vol. 3, 7th ed. Pearson Education.
- Hamunyela, E., Hipondoka, M., Persendt, F., Sevelia Nghiyalwa, H., Thomas, C., Matengu, K., 2022. Spatio-temporal characterization of surface water dynamics with Landsat in endorheic Cuvelai-Etoshia Basin (1990–2021). *ISPRS J. Photogramm. Remote Sens.* 191, 68–84. <https://doi.org/10.1016/j.isprsjprs.2022.07.007>.

- Han, W., Yang, Z., Di, L., Zhang, B., Peng, C., 2014. Enhancing agricultural geospatial data dissemination and applications using geospatial web services. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 7 (11), 4539–4547. <https://doi.org/10.1109/JSTARS.2014.2315593>.
- Hao, P., Di, L., Zhang, C., Guo, L., 2020. Transfer learning for crop classification with cropland data layer data (CDL) as training samples. *Sci. Total Environ.* 733, 138869. <https://doi.org/10.1016/j.scitotenv.2020.138869>.
- Heintzman, L.J., McIntyre, N.E., Langendoen, E.J., Read, Q.D., 2024. Cultivation and dynamic cropping processes impart land-cover heterogeneity within agroecosystems: a metrics-based case study in the Yazoo-Mississippi Delta (USA). *Landsc. Ecol.* 39 (2), 29. <https://doi.org/10.1007/s10980-024-01797-0>.
- Huang, Y., Qiu, B., Yang, P., Wu, W., Chen, X., Zhu, X., Xu, S., Wang, L., Dong, Z., Zhang, J., Berry, J., Tang, Z., Tan, J., Duan, D., Peng, Y., Lin, D., Cheng, F., Liang, J., Huang, H., Chen, C., 2024. National-scale 10 m annual maize maps for china and the contiguous united states using a robust index from sentinel-2 time series. *Comput. Electron. Agric.* 221, 109018. <https://doi.org/10.1016/j.compag.2024.109018>.
- Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS Vegetation Indices. *Remote Sens. Environ.* 195–213. www.elsevier.com/locate/rse.
- Johnson, D.M., 2013. A 2010 map estimate of annually tilled cropland within the conterminous United States. *Agr. Syst.* 114, 95–105. <https://doi.org/10.1016/j.agsy.2012.08.004>.
- Joshi, A., Pradhan, B., Gite, S., Chakraborty, S., 2023. Remote-sensing data and deep-learning techniques in crop mapping and yield prediction: a systematic review. *Remote Sens. (Basel)* 15 (8), 2014. <https://doi.org/10.3390/rs15082014>.
- Key, C. H., Benson, N. (2006). *Landscape Assessment (LA): Sampling and Analysis Methods*. https://www.fs.usda.gov/rm/pubs/series/rmrs/gtr/rmrs_gtr164/rmrs_gtr164_13_land_assess.pdf.
- Khan, A., Hansen, M.C., Potapov, P., Adusei, B., Stehman, S.V., Steininger, M.K., 2021. An operational automated mapping algorithm for in-season estimation of wheat area for Punjab, Pakistan. *Int. J. Remote Sens.* 42 (10), 3833–3849. <https://doi.org/10.1080/01431161.2021.1883200>.
- Khan, M.R., de Bie, C.A.J.M., van Keulen, H., Smaling, E.M.A., Real, R., 2010. Disaggregating and mapping crop statistics using hypertemporal remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* 12 (1), 36–46. <https://doi.org/10.1016/j.jag.2009.09.010>.
- King, L., Adusei, B., Stehman, S.V., Potapov, P.V., Song, X.-P., Krylov, A., Di Bella, C., Loveland, T.R., Johnson, D.M., Hansen, M.C., 2017. A multi-resolution approach to national-scale cultivated area estimation of soybean. *Remote Sens. Environ.* 195, 13–29. <https://doi.org/10.1016/j.rse.2017.03.047>.
- Kpienbaareh, D., Sun, X., Wang, J., Luginaah, I., Bezner Kerr, R., Lupafya, E., Dakishoni, L., 2021. Crop type and land cover mapping in northern malawi using the integration of sentinel-1, sentinel-2, and planetscope satellite Data. *Remote Sens. (Basel)* 13 (4), 700. <https://doi.org/10.3390/rs13040700>.
- Kumar, P., Gupta, D.K., Mishra, V.N., Prasad, R., 2015. Comparison of support vector machine, artificial neural network, and spectral angle mapper algorithms for crop classification using LISS IV Data. *Int. J. Remote Sens.* 36 (6), 1604–1617. <https://doi.org/10.1080/2150704X.2015.1019015>.
- Kumaraswamy, S., Subathra, K., Dattathreya, G., Ramkumar, G., Metwally, A.S.M., Ansari, M.Z., 2024. An ensemble neural network model for predicting the energy utility in individual houses. *Comput. Electr. Eng.* 114, 109059. <https://doi.org/10.1016/j.compeleceng.2023.109059>.
- Kyere, I., Astor, T., Graß, R., Wachendorf, M., 2020. Agricultural crop discrimination in a heterogeneous low-mountain range region based on multi-temporal and multi-sensor satellite data. *Comput. Electron. Agric.* 179, 105864. <https://doi.org/10.1016/j.compag.2020.105864>.
- Lambert, M.-J., Traoré, P.C.S., Blaes, X., Baret, P., Defourny, P., 2018. Estimating smallholder crops production at village level from sentinel-2 time series in Mali's Cotton Belt. *Remote Sens. Environ.* 216, 647–657. <https://doi.org/10.1016/j.rse.2018.06.036>.
- Li, H., Song, X.P., Hansen, M.C., Becker-Reshef, I., Adusei, B., Pickering, J., Wang, L., Wang, L., Lin, Z., Zalles, V., Potapov, P., Stehman, S.V., Justice, C., 2023. Development of a 10-m resolution maize and soybean map over China: matching satellite-based crop classification with sample-based area estimation. *Remote Sens. Environ.* 294. <https://doi.org/10.1016/j.rse.2023.113623>.
- Massey, R., Sankey, T.T., Congalton, R.G., Yadav, K., Thenkabail, P.S., Ozdogan, M., Sánchez Meador, A.J., 2017. MODIS phenology-derived, multi-year distribution of conterminous U.S. Crop Types. *Remote Sens. Environ.* 198, 490–503. <https://doi.org/10.1016/j.rse.2017.06.033>.
- Moraes, D., Benevides, P., Costa, H., Moreira, F. D., & Caetano, M. (2021). Influence of Sample Size in Land Cover Classification Accuracy Using Random Forest and Sentinel-2 Data in Portugal. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 4232–4235. doi: 10.1109/IGARSS47720.2021.9553924.
- Niclos, R., Perelló, M., Puchades, J., Coll, C., Valor, E., 2023. Evaluating landsat-9 TIRS-2 calibrations and land surface temperature retrievals against ground measurements using multi-instrument spatial and temporal sampling along transects. *Int. J. Appl. Earth Obs. Geoinf.* 125, 103576. <https://doi.org/10.1016/j.jag.2023.103576>.
- Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good Practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* 148, 42–57. <https://doi.org/10.1016/j.rse.2014.02.015>.
- Pittman, K., Hansen, M.C., Becker-Reshef, I., Potapov, P.V., Justice, C.O., 2010. Estimating global cropland extent with multi-year MODIS Data. *Remote Sens. (Basel)* 2 (7), 1844–1863. <https://doi.org/10.3390/rs2071844>.
- Potapov, P., Turubanova, S., Hansen, M.C., Tyukavina, A., Zalles, V., Khan, A., Song, X.-P., Pickens, A., Shen, Q., Cortez, J., 2021. Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century. *Nat. Food* 3 (1), 19–28. <https://doi.org/10.1038/s43016-021-00429-z>.
- Rao, P., Zhou, W., Bhattacharai, 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 Sens. (Basel)* 13 (10), 1870. <https://doi.org/10.3390/rs13101870>.
- Roy, D.P., Kovalsky, V., Zhang, H.K., Vermote, E.F., Yan, L., Kumar, S.S., Egorov, A., 2016. Characterization of landsat-7 to landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote Sens. Environ.* 185, 57–70. <https://doi.org/10.1016/j.rse.2015.12.024>.
- Schmidhuber, J., 2015. Deep learning in neural networks: an overview. *Neural Netw.* 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>.
- Seong, S., Chang, A., Mo, J., Na, S., Ahn, H., Oh, J., Choi, J., 2024. Crop classification in South Korea for multitemporal planetscope imagery using SFC-Densenet-AM. *Int. J. Appl. Earth Obs. Geoinf.* 126, 103619. <https://doi.org/10.1016/j.jag.2023.103619>.
- Sildir, H., Aydin, E., Kavzoglu, T., 2020. Design of feedforward neural networks in the classification of hyperspectral imagery using superstructural optimization. *Remote Sens. (Basel)* 12 (6), 956. <https://doi.org/10.3390/rs12060956>.
- Song, X.-P., Li, H., Potapov, P., Hansen, M.C., 2022. Annual 30 m soybean yield mapping in Brazil using long-term satellite observations, climate data and machine learning. *Agric. For. Meteorol.* 326, 109186. <https://doi.org/10.1016/j.agrformet.2022.109186>.
- Song, X.-P., Potapov, P.V., Krylov, A., King, L., Di Bella, C.M., Hudson, A., Khan, A., Adusei, B., Stehman, S.V., Hansen, M.C., 2017. National-scale soybean mapping and area estimation in the united states using medium resolution satellite imagery and field survey. *Remote Sens. Environ.* 190, 383–395. <https://doi.org/10.1016/j.rse.2017.01.008>.
- Sonobe, R., 2019. Parcel-based crop classification using multi-temporal TerraSAR-X dual polarimetric data. *Remote Sens. (Basel)* 11 (10), 1148. <https://doi.org/10.3390/rs11101148>.
- Sonobe, R., Yamaya, Y., Tani, H., Wang, X., Kobayashi, N., Mochizuki, K., 2017. Assessing the suitability of data from sentinel-1A and 2A for crop classification. *GIScience & Remote Sensing* 54 (6), 918–938. <https://doi.org/10.1080/15481603.2017.1351149>.
- Stehman, S.V., 2009. Sampling designs for accuracy assessment of land cover. *Int. J. Remote Sens.* 30 (20), 5243–5272. <https://doi.org/10.1080/0143116090331000>.
- Stehman, S.V., 2012. Impact of sample size allocation when using stratified random sampling to estimate accuracy and area of land-cover change. *Remote Sens. Lett.* 3 (2), 111–120. <https://doi.org/10.1080/01431161.2010.541950>.
- Stehman, S.V., Wagner, J.E., 2024. Choosing a sample size allocation to strata based on trade-offs in precision when estimating accuracy and area of a rare class from a stratified sample. *Remote Sens. Environ.* 300, 113881. <https://doi.org/10.1016/j.rse.2023.113881>.
- Svozil, D., Kvasnicka, V., Pospichal, J., 1997. Introduction to multi-layer feed-forward neural networks. *Chemom. Intel. Lab. Syst.* 39 (1), 43–62. [https://doi.org/10.1016/S0169-7439\(97\)00061-0](https://doi.org/10.1016/S0169-7439(97)00061-0).
- Tran, K.H., Zhang, H.K., McMaine, J.T., Zhang, X., Luo, D., 2022. 10 m crop type mapping using sentinel-2 reflectance and 30 m cropland data layer product. *Int. J. Appl. Earth Obs. Geoinf.* 107, 102692. <https://doi.org/10.1016/j.jag.2022.102692>.
- Venter, Z.S., Barton, D.N., Chakraborty, T., Simensen, T., Singh, G., 2022. Global 10 m land use and land cover datasets: a comparison of dynamic world, world cover and esri land cover. *Remote Sens. (Basel)* 14 (16), 4101. <https://doi.org/10.3390/rs14164101>.
- Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the landsat 8/OLI land surface reflectance product. *Remote Sens. Environ.* 185, 46–56. <https://doi.org/10.1016/j.rse.2016.04.008>.
- Wagner, J.E., Stehman, S.V., 2015. Optimizing sample size allocation to strata for estimating area and map accuracy. *Remote Sens. Environ.* 168, 126–133. <https://doi.org/10.1016/j.rse.2015.06.027>.
- Wang, L., Jin, G., Xiong, X., 2022. Improvement in land cover classification using multitemporal sentinel-1 and sentinel-2 satellite imagery. In: *4th International Conference on Information Technology and Computer Communications (ITCC)*, pp. 16–24. <https://doi.org/10.1145/3548636.3548639>.
- Wang, Y., Huang, H., State, R., 2023b. Early crop mapping using dynamic ecoregion clustering: a USA-Wide Study. *Remote Sens. (Basel)* 15 (20), 4962. <https://doi.org/10.3390/rs15204962>.
- Wang, Y., Zhang, Z., Feng, L., Ma, Y., Du, Q., 2021. A new attention-based CNN approach for crop mapping using time series sentinel-2 images. *Comput. Electron. Agric.* 184, 106090. <https://doi.org/10.1016/j.compag.2021.106090>.
- Wardlaw, B.D., Egbert, S.L., 2008. Large-area crop mapping using time-series MODIS 250 m NDVI data: an assessment for the U.S. Central Great Plains. *Remote Sens. Environ.* 112 (3), 1096–1116. <https://doi.org/10.1016/j.rse.2007.07.019>.
- Wardlaw, B.D., Egbert, S.L., Kastens, J.H., 2007. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains. *Remote Sens. Environ.* 108 (3), 290–310. <https://doi.org/10.1016/j.rse.2006.11.021>.
- Xiao, X., Jiang, L., Liu, Y., Ren, G., 2023. Limited-samples-based crop classification using a time-weighted dynamic time warping method, sentinel-1 imagery, and google earth engine. *Remote Sens. (Basel)* 15 (4), 1112. <https://doi.org/10.3390/rs15041112>.
- Xu, Y., Ma, Y., Zhang, Z., 2024. Self-supervised pre-training for large-scale crop mapping using sentinel-2 time series. *ISPRS J. Photogramm. Remote Sens.* 207, 312–325. <https://doi.org/10.1016/j.isprsjprs.2023.12.005>.
- Yan, L., Roy, D.P., 2016. Conterminous United States crop field size quantification from multi-temporal landsat data. *Remote Sens. Environ.* 172, 67–86. <https://doi.org/10.1016/j.rse.2015.10.034>.

- Yan, S., Yao, X., Zhu, D., Liu, D., Zhang, L., Yu, G., Gao, B., Yang, J., Yun, W., 2021. Large-scale crop mapping from multi-source optical satellite imageries using machine learning with discrete grids. *Int. J. Appl. Earth Obs. Geoinf.* 103, 102485. <https://doi.org/10.1016/j.jag.2021.102485>.
- Yu, Q., Duan, Y., Wu, Q., Liu, Y., Wen, C., Qian, J., Song, Q., Li, W., Sun, J., Wu, W., 2023. An interactive and iterative method for crop mapping through crowdsourcing optimized field samples. *Int. J. Appl. Earth Obs. Geoinf.* 122, 103409. <https://doi.org/10.1016/j.jag.2023.103409>.
- Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., Zhang, L., 2020. Deep learning in environmental remote sensing: achievements and challenges. *Remote Sens. Environ.* 241, 111716. <https://doi.org/10.1016/j.rse.2020.111716>.
- Zha, Y., Gao, J., Ni, S., 2003. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int. J. Remote Sens.* 24 (3), 583–594. <https://doi.org/10.1080/01431160304987>.
- Zhang, C., Di, L., Lin, L., Guo, L., 2019. Machine-learned prediction of annual crop planting in the U.S. Corn Belt based on Historical Crop Planting Maps. *Comput. Electron. Agric.* 166, 104989. <https://doi.org/10.1016/j.compag.2019.104989>.
- Zhang, H., Kang, J., Xu, X., Zhang, L., 2020. Accessing the temporal and spectral features in crop type mapping using multi-temporal sentinel-2 imagery: a case study of Yi'an County, Heilongjiang Province, China. *Comp. Electr. Agric.* 176, 105618. <https://doi.org/10.1016/j.compag.2020.105618>.
- Zheng, B., Myint, S.W., Thenkabail, P.S., Aggarwal, R.M., 2015. A support vector machine to identify irrigated crop types using time-series landsat NDVI Data. *Int. J. Appl. Earth Obs. Geoinf.* 34, 103–112. <https://doi.org/10.1016/j.jag.2014.07.002>.
- Zheng, Q., Ha, T., Prishchepov, A.V., Zeng, Y., Yin, H., Koh, L.P., 2023. The neglected role of abandoned cropland in supporting both food security and climate change mitigation. *Nat. Commun.* 14 (1), 6083. <https://doi.org/10.1038/s41467-023-41837-y>.
- Zhong, L., Gong, P., Biging, G.S., 2014. Efficient corn and soybean mapping with temporal extendability: a multi-year experiment using landsat imagery. *Remote Sens. Environ.* 140, 1–13. <https://doi.org/10.1016/j.rse.2013.08.023>.
- Zhong, L., Hu, L., Zhou, H., 2019. Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* 221, 430–443. <https://doi.org/10.1016/j.rse.2018.11.032>.
- Zhu, Z., Wang, S., Woodcock, C.E., 2015. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for landsats 4–7, 8, and sentinel 2 images. *Remote Sens. Environ.* 159, 269–277. <https://doi.org/10.1016/j.rse.2014.12.014>.
- Zimmerman, P.L., Housman, I.W., Perry, C.H., Chastain, R.A., Webb, J.B., Finco, M.V., 2013. An accuracy assessment of forest disturbance mapping in the western great lakes. *Remote Sens. Environ.* 128, 176–185. <https://doi.org/10.1016/j.rse.2012.09.017>.