Multimodal Datasets for Crop Classification in Precision Agriculture

Salma Oumoussa

7 March 2025

Abstract

Remote sensing and data-driven approaches play a pivotal role in precision agriculture, particularly in crop classification. This document provides a structured overview of publicly available multimodal datasets, including satellite, UAV-based, hyperspectral, and soil-related sources. By categorizing datasets based on modality, resolution, and accessibility, we assist researchers and practitioners in selecting suitable data for robust crop classification models. Key challenges in integration, preprocessing, and AI-driven classification improvements are also discussed to guide future research..

Introduction

The advancement of remote sensing technology has significantly improved agricultural monitoring and decision-making processes. Crop classification, a key application of precision agriculture, relies on diverse datasets spanning multiple modalities, including optical, radar, LiDAR, hyperspectral, and UAV-based imagery. However, integrating these datasets presents challenges, such as varying spatial and temporal resolutions, preprocessing complexities, and interoperability constraints.

This document presents a **non-exhaustive** repository of datasets essential for intelligent crop classification, detailing their sources, characteristics, and applications. The focus is on multimodal fusion techniques, which enhance classification accuracy by combining spectral, structural, and temporal information. By systematically reviewing these resources, we aim to bridge the gap between data availability and practical implementation in agricultural research. Additionally, we outline key considerations for dataset selection, preprocessing requirements, and future research directions in AI-driven crop classification.

Satellite-Based Datasets for Direct Remote Sensing Imagery

Satellite-based datasets provide large-scale, high-resolution imagery crucial for crop classification. These datasets include publicly available remote sensing data from satellites specializing in multispectral, hyperspectral, radar (SAR), thermal, LiDAR, and microwave imaging. Researchers and practitioners can use these direct sources to acquire imagery for precision agriculture applications.

Category	Type	Source	Resolution	Direct Access
		Optical Sensors		
Multispectral	Satellite	Landsat Series	30m-60m	Access Here
		Sentinel-2 (MSI)	10m-60m	Access Here
Hyperspectral	Satellite	Hyperion	30m	Access Here
		EnMAP	30m	Access Here
		Radar Sensors (SAR)		
SAR	Satellite	Sentinel-1	10m-40m	Access Here
		ALOS PALSAR	10m-100m	Access Here

(Table Continued)

Category	Type	Source	Resolution	Direct Access
		LiDAR Sensors		
Spaceborne	Satellite	ICESat-2	Global	Access Here
		GEDI	25m footprint	Access Here
Airborne	Aircraft	Airborne LiDAR	¡1m	Access Here
		Thermal Infrared Sensors		
TIR	Satellite	Landsat-8 TIRS	100m	Access Here
		MODIS TIR	1km	Access Here
		Passive Microwave Sensors		
Passive Microwave	Satellite	SMAP	10km-40km	Access Here
		GRACE	Global	Access Here
		High-Resolution Commercial Sensors		
Commercial	Satellite	WorldView-3	0.3m	Access Here
		Pléiades	0.5m	Access Here

Multimodat Datasets

Satellite-Based Datasets

Satellite-based datasets provide large-scale, high-resolution imagery crucial for crop classification. These datasets typically include multispectral, hyperspectral, and radar data, enabling robust analysis of vegetation patterns, growth cycles, and land cover classification.

Disclaimer: This section presents a curated selection of publicly available datasets relevant to crop classification in precision agriculture. While comprehensive, this is not an exhaustive list.

Dataset Name	Description	Use Case	Access
CropScape - Cropland Data Layer	National-level dataset for land cover classification.	Crop classification, land use monitoring.	Explore
National-scale Crop Type Maps for Germany	Sentinel-1, Sentinel-2, and Landsat-based dataset (2017-2019).	Crop classification, segmentation.	Read more
SICKLE: A Multi-Sensor Satellite Imagery Dataset	Annotated dataset for key cropping parameters.	Crop classification, multi-spectral analysis.	Read more
Sen4AgriNet (Sentinel-2 Benchmark)	Sentinel-2 dataset with farmer-declared crop labels.	Agricultural monitoring, crop classification.	Github access Explore
Sentinel-2 Benchmark Dataset	Multi-year Sentinel-2 dataset for classification.	Land cover analysis, crop segmentation.	Access
USDA Cropland Data Layer	Landsat and Sentinel-based high-resolution cropland dataset.	Land cover classification, crop monitoring.	View dataset
Planted: Dataset for Planted Forest Identification	Multi-satellite time-series dataset for forest and crop identification.	Crop classification, land monitoring.	Read more
BigEarthNet	A Large-Scale Benchmark Archive For Remote Sensing Image Understanding	Image Classification, Semantic Segmentation, Multi-Label Image Classification	Documentation

UAV-Based Datasets

UAV-based datasets provide high-resolution, flexible, and cost-effective imaging solutions for precision agriculture. These datasets typically include multispectral, thermal, and hyperspectral imagery for small to medium-scale crop classification applications.

Dataset Name	Description	Use Case	Access
BREIZHCROPS: Time Series Dataset	UAV-based time series data for crop type mapping.	Precision agriculture, crop health assessment.	Read more
ClarkCGA Multi-Temporal Crop Classification Data	Multi-temporal training data for crop classification.	Remote sensing applications in agriculture.	Github Access
CropAndWeed Dataset	UAV-based dataset for crop and weed classification.	Machine learning, crop identification.	Explore paper Github access
Early Crop Classification Dataset	Fusion of Sentinel-1, Sentinel-2, and UAV time series data for early-season crop identification.	Enables early crop predictions for timely agricultural decisions.	Published in MDPI Remote Sensing

Hyperspectral & Multimodal Remote Sensing Datasets

These datasets combine multiple sensor modalities, including hyperspectral, multispectral, LiDAR, and microwave, providing deep insights into crop classification and stress monitoring.

Dataset Name	Description	Use Case	Access
MDAS: Multimodal Benchmark Dataset	Multimodal dataset integrating hyperspectral, LiDAR, and microwave data.	Remote sensing applications, precision agriculture.	Read more
Multi-Modal Temporal Attention Models	A model trained on satellite time-series data for improved crop mapping.	Temporal crop analysis, classification improvement.	Explore Github access
CropSpectral Dataset	Multi-temporal hyperspectral dataset for high-resolution crop classification.	Crop health monitoring, stress detection, and species identification.	Download
Hierarchical Crop Classification Fusion	A fusion dataset that integrates satellite imagery, crop rotation history, and contextual data for improved classification.	Enhancing crop classification accuracy and generalization.	Available on arXiv

Soil & Fertility Datasets

Datasets related to soil nutrients, soil moisture, and fertility assessments, aiding precision agriculture strategies.

Dataset Name	Description	Use Case	Access
GEOBON Soil Nutrient Database	High-resolution soil nutrient database derived from Earth observation and ground sampling.	Supports regional and global soil fertility monitoring.	Visit GEOBON
Global Crop Fertilization Dataset (1961-2019)	Contains nitrogen, phosphorus, and potassium application rates for 13 major crop groups worldwide.	Useful for studying fertilization trends and optimizing nutrient application.	Browse on arXiv
Global Plant Nitrogen Traits Dataset	Dataset compiling nitrogen-related plant traits across different crop species and regions.	Helps researchers study plant nutrition and improve fertilizer recommendations.	Explore on Nature Scientific Data
LandPKS Soil	Open-source soil and land potential dataset for agricultural analysis.	Helps assess land suitability for various crops.	Access LandPKS database
OpenLandMap Soil Dataset	Global soil property dataset based on remote sensing, geostatistics, and field observations.	Essential for soil classification, precision agriculture, and climate studies.	View dataset on OpenLandMap
Soil Macronutrient Assessment	Reviews remote sensing applications for monitoring soil macronutrients (NPK) in agricultural lands.	Supports precision agriculture strategies for soil fertility management.	Read full study on MDPI

SoilGrids250m	Global soil property and class predictions based on machine learning and remote sensing.	Helps map soil characteristics for targeted agricultural applications.	Explore SoilGrids
Variable Rate Nitrogen Management Dataset	Satellite-based nitrogen zone delineation, integrating crop models and soil variability assessments.	Optimizes fertilizer application by mapping spatial nitrogen needs.	Download research paper
World Soil Information (WoSIS)	Comprehensive soil profile database integrating field and remote sensing data.	Supports soil fertility analysis and precision agriculture studies.	Explore ISRIC WoSIS

Crop Classification & Yield Prediction Datasets

Datasets specifically focused on crop classification, segmentation, phenology tracking, and yield prediction.

Dataset Name	Description	Use Case	Access
Climate Change-aware Crop Yield Predictions	An open, large-scale dataset for crop yield predictions considering climate change factors.	Supports climate-resilient agricultural strategies. (USE the CropNet package)	Read more
Crop Performance & Yield Trials	Crop performance, aerial, and satellite data from multistate maize yield trials.	Supports crop yield estimation and performance benchmarking.	Access dataset
Crop Recommendation Dataset	Uses soil parameters and climate data to recommend crops best suited for an area.	Precision agriculture, automated crop recommendation systems.	Explore dataset
Crop Yield Mapping	Analyzes remote sensing capabilities for estimating global crop yields.	Helps integrate satellite imagery with yield prediction models.	Kaggle Access
CalCROP21	Global dataset of 173 crop types at a 5.6km resolution		
EuroCropsML	A large-scale dataset for crop classification using machine learning techniques.	Crop type mapping, supervised learning applications.	View dataset
Fine-Scale Crop Classification Dataset	High-resolution dataset for detailed crop mapping.	Supports fine-scale agricultural monitoring and crop identification.	Read full study on Frontiers

Open-Access & General Agricultural Data Repositories

Datasets that compile agricultural data from multiple sources, serving as repositories for research and precision agriculture applications.

Dataset Name	Description	Use Case	Access
Callisto-Dataset-Collection	Collection of datasets for agricultural research, integrating multiple data sources.	Supports machine learning applications in precision agriculture.	Explore
MDAS: Multimodal Benchmark Dataset	Comprehensive dataset combining remote sensing, climate, and field data.	Used for evaluating multimodal models in agricultural classification.	Read more
Machine Learning for Precision Agriculture	Uses NPK levels, soil pH, and climate data to recommend crops and fertilizers.	Supports AI-driven decision-making in precision farming.	Read full paper on PubMed Related dataset
NASA Harvest Agricultural Dataset	Collection of datasets for global agricultural monitoring integrating satellite and in-situ data.	Used for food security analysis, yield estimation, and crop classification.	Explore NASA Harvest

Conclusion

This document compiles a diverse range of datasets essential for precision agriculture, particularly for crop classification. By structuring datasets based on modality and application, we aim to assist researchers in selecting and integrating multimodal data sources effectively. Future work should focus on real-time data integration, improving preprocessing methodologies, and leveraging AI-driven approaches to enhance classification accuracy and decision-making in agricultural monitoring.