Hassan 1st University
National School of Applied Sciences – Berrechid

Final Year Internship Project Report

Major: Computer Engineering

Research and Development of an AI System for Early Crop Classification in Precision Agriculture

A Sentinel-2-Based Proof of Concept for Future Multimodal Integration

Prepared By:

Salma Oumoussa

Supervised By:

- Prof. **Diego Pellufo** Head of SDAS Group
Assistant Professor, College of Computing, Mohammed VI Polytechnic University
- Prof. **Khalid Bouihat**
Full Professor, National School of Applied Sciences of Berrechid
- Prof. **Adil Haddi**
Full Professor, National School of Applied Sciences of Berrechid

Defended On:

XX/06/2026

Defended Before:

- Prof. **Khalid Bouihat**
- Prof. **Adil Haddi**

Academic Year: 2024–2025

Dedication

To my father, **Brahim Oumoussa**, who is no longer with us, but who did everything to ensure I received a high-quality education. May Allah grant him Jannah and eternal peace.

To my mother, **Halima Saadani**, whose countless sacrifices, strength, and unconditional support have shaped the person I am today. May Allah bless and protect her.

To my brother, **Yahya Oumoussa**, who stood by me during moments of weakness and never stopped reminding me how proud he is of me. May Allah preserve him and grant him success.

To **my family and friends**

Thank you for your presence, your encouragement, and your unwavering love.

Salma Oumoussa

Acknowledgments

I would like to express my sincere gratitude to all the individuals and institutions who supported and contributed to the successful completion of this work.

First and foremost, I extend my thanks to the **College of Computing at Mohammed VI Polytechnic University (UM6P)** and the **Smart Data Analysis Systems (SDAS) Group** for providing an enriching and interdisciplinary research environment throughout this project.

My deepest appreciation goes to **Professor Diego Pellufo**, my academic supervisor at UM6P, for his invaluable guidance, continuous support, and insightful feedback at every stage of this research.

I am also thankful to my academic supervisors, **Prof. Khalid Bouihat and Prof. Adil Haddi**, for their mentorship, encouragement, and thoughtful direction during this journey.

Lastly, I would like to thank everyone who contributed to this work in any way—whether through discussions, reviews, or moral support. Your contributions, however big or small, have been instrumental in bringing this project to fruition.

Thank you all.

Abstract

In the era of Agriculture 4.0, the fusion of artificial intelligence (AI) and remote sensing is transforming traditional agricultural practices into data-driven, sustainable systems. This research presents the development of an early-season crop classification system based on Sentinel-2 satellite time series and deep learning techniques. Conducted within the UM6P College of Computing in collaboration with the SDAS Group, the study implements and evaluates a proof-of-concept model, EarlyRNN, designed to address core challenges such as variable sequence lengths and class imbalance.

Using the BavarianCrops dataset, the proposed approach demonstrates significant improvements in early prediction accuracy, positioning itself as a valuable decision-support tool for precision agriculture. The system enables timely crop identification, supporting targeted interventions such as fertilizer application and resource allocation. This work lays the foundation for a broader multimodal integration involving soil data, climate records, and phenological signals—paving the way for intelligent, scalable agricultural systems.

Résumé

À l'ère de l'Agriculture 4.0, l'intégration de l'intelligence artificielle (IA) aux données de télédétection devient essentielle pour moderniser les pratiques agricoles et garantir leur durabilité. Cette recherche propose un système de classification des cultures reposant sur une approche multimodale combinant séries temporelles satellitaires, données pédologiques et phénologie des cultures. Réalisé au sein du Collège de l'informatique de l'UM6P, en collaboration avec le groupe SDAS, le projet repose sur un pipeline intelligent basé sur le modèle **EarlyRNN**, une architecture de deep learning adaptée aux données temporelles, intégrant un mécanisme de décision précoce.

Le jeu de données **BavarianCrops** sert de base expérimentale pour l'entraînement et l'évaluation du modèle, en tenant compte de deux défis majeurs : la variabilité des longueurs de séquence et le déséquilibre entre classes. Des techniques spécifiques de prétraitement, de fusion de données et d'optimisation ont été mises en œuvre pour améliorer la robustesse du modèle.

Ce système permet une classification précoce et fiable des cultures, ouvrant la voie à une meilleure gestion des intrants, à l'optimisation des ressources et à une agriculture plus durable. Ce travail s'inscrit dans une démarche de recherche appliquée en agriculture de précision et contribue à l'enrichissement des solutions intelligentes au service du secteur agricole

List of Abbreviations

Abbreviation	Full Term
AI	Artificial Intelligence
PA	Precision Agriculture
NDVI	Normalized Difference Vegetation Index
SAR	Synthetic Aperture Radar
S2	Sentinel-2
DOY	Day of Year
GRU	Gated Recurrent Unit
LSTM	Long Short-Term Memory
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
EarlyRNN	Early Recurrent Neural Network
ConvLSTM	Convolutional Long Short-Term Memory
UAV	Unmanned Aerial Vehicle
GT	Ground Truth
OA	Overall Accuracy
F1	F1 Score
PRISMA-ScR	Preferred Reporting Items for Systematic Reviews and Meta-Analyses – Scoping Review
SDAS	Smart Data Analysis Systems
UM6P	Mohammed VI Polytechnic University
CropID	Crop Identification Platform
TCN	Temporal Convolutional Network

List of figures

List of Tables

Contents

General Introduction

The agricultural sector stands at a pivotal crossroads, facing the dual imperative of increasing productivity while ensuring sustainability in the face of environmental, economic, and demographic pressures. In this context, **precision agriculture**—powered by artificial intelligence (AI), remote sensing, and data-driven decision-making—emerges as a transformative approach. It enables the optimization of agricultural inputs, enhances monitoring of crop development, and ultimately contributes to more efficient, resilient, and environmentally conscious farming systems.

Within this evolving landscape, crop classification plays a fundamental role. Accurate identification and timely monitoring of crop types are crucial not only for yield estimation and resource planning, but also for implementing adaptive interventions that support sustainable food systems. However, this task remains highly complex due to factors such as environmental variability, class imbalance, inconsistent data quality, and the need for early predictions in dynamic agro-ecological settings.

This research project, conducted at **the College of Computing at UM6P** in collaboration with the **Smart Data Analysis Systems (SDAS) Group**, addresses these challenges by developing a multimodal crop classification system based on satellite time-series data and deep learning techniques. The project focuses on EarlyRNN, a model designed for early and accurate classification, supporting field-level decisions in precision agriculture.

To provide a structured and comprehensive understanding of the research endeavor, this report is organized as follows:

- **Chapter 1** introduces the broader context of precision agriculture and Agriculture 4.0, presents the host institutions, and outlines the rationale, objectives, and scope of the project.
- **Chapter 2** offers an in-depth literature review of existing work on crop classification, remote sensing technologies, and the integration of deep learning models—particularly for time-series and multimodal data analysis.
- **Chapter 3** details the research methodology, including both the systematic literature review approach and the experimental framework implemented using the BavarianCrops dataset.
- **Chapter 4** presents the results obtained from model training and evaluation, highlighting the effectiveness of the proposed system in addressing early classification challenges.
- **Chapter 5** discusses the implications of the findings, methodological constraints, and areas for improvement, in relation to the current state of research.
- **Chapter 6** concludes the report by summarizing key contributions, drawing final reflections, and proposing directions for future work and real-world applications.

Through this structured exploration, the project aims to contribute meaningfully to the advancement of AI-powered agricultural systems and to the broader scientific discourse surrounding sustainable and intelligent farming practices.

Chapter 1: Context and Research Framework

Introduction

This chapter introduces the institutional and scientific context of the project, rooted in the collaboration between UM6P's College of Computing and the SDAS group. It outlines the shift toward data-driven agriculture through Precision Agriculture and Agriculture 4.0, identifies the limitations of current crop classification methods, and presents the project's goals, roadmap, and innovations—laying the groundwork for an AI-powered, multimodal crop classification system.

Presentation of the Host Organization

This research project was carried out within the College of Computing at Mohammed VI Polytechnic University (UM6P), under the joint academic supervision of the Smart Data Analysis Systems (SDAS) Group. Both institutions offer a dynamic research environment centered on artificial intelligence, data science, and the development of intelligent systems. Their collaboration provided the ideal foundation for the development of a project that bridges cutting-edge AI research with real-world agricultural applications.

1.1UM6P College of Computing

The College of Computing at UM6P is a nationally and internationally recognized institution committed to advancing education and research in computing sciences. Its core focus areas include artificial intelligence, data science, cybersecurity, and high-performance computing. With strong ties to industrial and academic partners, the College fosters interdisciplinary innovation and supports projects addressing key societal challenges, including sustainability, agriculture, and digital transformation.

Figure 1: The UM6P College of Computing logo, symbolizing the academic and research backbone of this project.

1.2 SDAS Group (Co-Supervision)

The SDAS Group specializes in AI and data-intensive applications, including remote sensing, which directly aligned with the technical needs of this project. Led by Prof. Diego Pellufo, the group's work spans a variety of domains such as medical diagnostics, remote sensing, embedded AI systems, and software quality engineering, all with a strong emphasis on ethical and explainable AI.

Figure 2 : The SDAS Group logo, representing the technical expertise in AI and data science that guided this research.

The collaboration between the College of Computing and SDAS created the right environment for this internship project, which was designed to explore the integration of AI into precision agriculture. Drawing on both academic rigor and applied innovation, the project was conceived to develop a proof of concept focused on early crop classification using satellite data, laying the groundwork for a more extensive multimodal system in the future.

General Context

1.1 Precision Agriculture and Its Role in Sustainable Farming

Precision agriculture (PA) has emerged as a transformative approach to farming, driven by the need to produce more food while minimizing environmental impact. By leveraging technologies such as **remote sensing**, **geographic information systems (GIS)**, and **machine learning (ML)**, PA enables farmers to make data-informed decisions at the field level, optimizing interventions like irrigation, fertilization, and pest control.

Unlike traditional agricultural practices that treat fields uniformly, PA recognizes spatial and temporal variability, tailoring inputs to specific zones within a field. This improves resource efficiency, reduces input waste, and contributes to long-term soil and ecosystem health, these innovations directly support the core objectives of sustainable agriculture.

Key enabling technologies include:

- **Remote Sensing & GIS** – Provide continuous monitoring of crop health, soil conditions, and environmental variables through satellite or aerial imagery.
- **AI and Machine Learning** – Facilitate crop classification, yield prediction, and anomaly detection through pattern recognition and model-based decision-making.
- **Variable Rate Technology (VRT)** – Enables precise delivery of fertilizers or pesticides only where needed, based on geospatial insights.

These tools have laid the foundation for digital, data-driven agriculture, and form the technological basis for the project at hand.

This figure outlines the process flow of precision agriculture, starting from the collection of key data—such as soil, weather, and crop health—through to data analysis and decision-making. By integrating these insights, precision agriculture optimizes resource application, leading to improved crop yield and more sustainable farming practices.

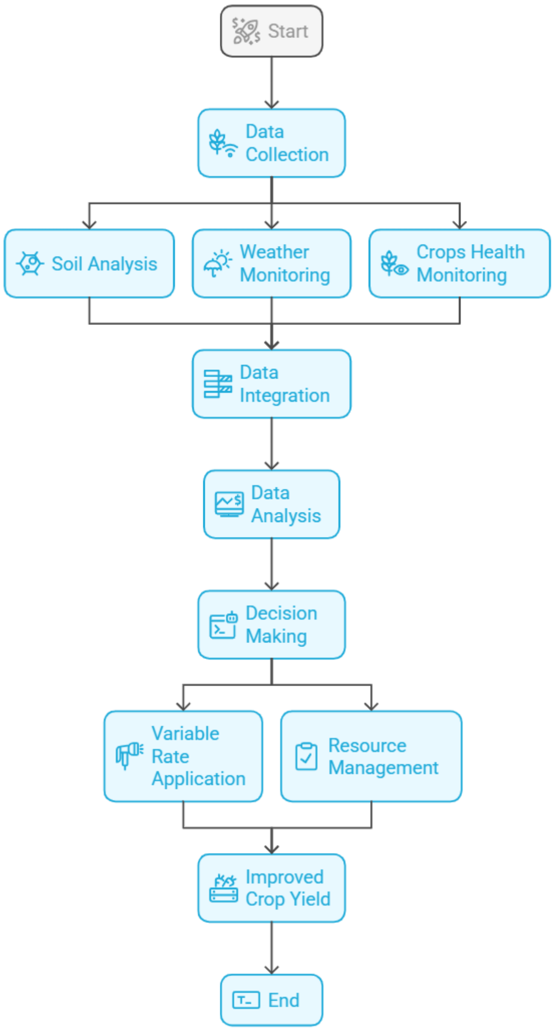


Figure 3:Precision Agriculture Workflow: From data collection to improved crop yield through optimized resource management.

1.2 Agriculture 4.0 and the Digital Revolution in Farming

Building upon the principles of precision agriculture, **Agriculture 4.0** marks the next evolution in farming characterized by automation, real-time sensing, and intelligent systems. This paradigm leverages the **Internet of Things (IoT)**, **robotics**, **big data analytics**, and **cloud computing** to support continuous, adaptive agricultural operations.

Key features of Agriculture 4.0 include:

- **Real-Time Monitoring** – Through connected sensors and remote platforms, farmers can access live data on weather, soil moisture, plant stress, and pest risk.
- **Automated Decision-Making** – AI-powered systems guide timely interventions, improving operational efficiency and reducing dependency on manual labor.
- **Autonomous Systems** – Robots and drones are used for tasks such as planting, spraying, and harvesting, allowing for scalable, cost-effective operations.

These innovations transform farms into interconnected ecosystems capable of self-adjustment and learning. The relevance of Agriculture 4.0 to this research lies in its emphasis on **data fusion, early prediction**, and **scalable decision-support systems**—all of which underpin the rationale for developing an AI-driven crop classification platform.

1.3 Project Context

This project is situated at the intersection of **precision agriculture**, **Agriculture 4.0**, and the growing need for **sustainable farming practices**. As the agricultural sector transitions toward smart, connected systems, the ability to accurately classify crops using integrated data sources becomes critical for optimizing input use and maximizing yield.

The project is particularly aligned with the strategic goals of the **OCP Group**, which promotes precision-based fertilizer recommendations to enhance productivity and reduce environmental harm. While many existing crop classification systems rely on **single-source data** or generalized models, this research seeks to push beyond those limitations.

The initial focus is on developing a **Sentinel-2-based proof of concept**, leveraging multitemporal satellite imagery for early-season crop classification. This serves as a foundational step toward a more comprehensive, multimodal system that will later incorporate **soil health metrics**, **climate data**, and **phenological information**.

Ultimately, the project aims to demonstrate how deep learning and remote sensing can support timely, field-specific decisions in crop management—contributing both to scientific advancement and to the operational goals of sustainable agriculture.

Problem Statement: Overcoming the Limitations of Crop Classification in Precision Agriculture

Agricultural productivity is under significant strain due to increasing global food demand, the unpredictable effects of climate change, and the ongoing degradation of natural resources. In this context, accurate crop classification and the efficient use of fertilizers are essential for ensuring sustainable, high-yield farming. However, these domains face persistent, multi-dimensional challenges:

1. **Limited Accuracy from Single-Source Data:** Traditional models often rely on optical imagery or static rules, lacking the temporal depth to detect early growth stages or distinguish similar crops. Multimodal integration (e.g., soil, phenology, climate) is rarely implemented but essential for early and precise classification.
2. **Inefficient Fertilizer Use:** Fertilizers are often applied without crop-specific insight, leading to economic loss and environmental harm—nutrient runoff, groundwater pollution, and greenhouse emissions. Targeted fertilization requires reliable classification.

3. **Lack of Farmer-Centric Tools:** Existing platforms focus on research or policy, offering little practical value to farmers. Real-time, localized, and user-friendly tools are scarce, especially for small and medium-scale operations.
4. **Multimodal Data Integration Challenges:** Combining different data sources remains difficult due to inconsistencies in resolution, format, and timing. Without robust fusion methods, scalable AI-driven systems remain out of reach.

The Need for a New Approach

Solving current challenges requires crop classification systems that are multimodal, AI-driven, early, and farmer-focused.

This project addresses that need by developing a scalable deep learning model based on remote sensing, with future integration of soil and phenology data—combining scientific rigor with real-world applicability.

Presentation of the Research Project

This research project, titled "***CropID: Precision crop classification platform for tailored fertilizer recommendations through data integration"*,** is a cutting-edge initiative designed to integrate multimodal data for accurate crop classification and optimized fertilizer use. The platform aims to contribute to precision agriculture by providing farmer-centric insights, tailored to specific crop and soil conditions**.**

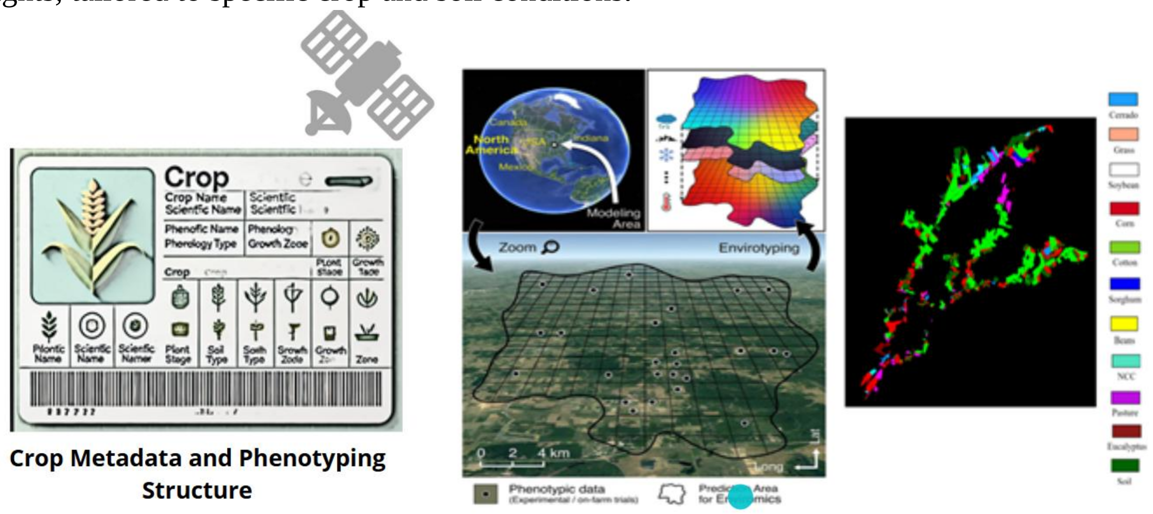


Figure 4.Illustration of a multimodal framework for crop classification and environmental modeling as CropID identity.

1.1 General Objective

The primary objective of this project is to develop an AI-powered system for crop classification that integrates various data modalities—such as remote sensing imagery, soil health data, and crop growth patterns—to improve the accuracy and efficiency of agricultural practices.

1.2 Specific Objectives

- **Implementation of a Sentinel-2-Based Proof of Concept:** Designing and evaluating a unimodal prototype based solely on Sentinel-2 time-series data as an initial step toward the integration of additional modalities (e.g., soil, phenology), thereby laying the foundation for a fully multimodal system.
- **Development of a Multimodal Crop Classification Model:** Utilizing multi-source data (e.g., satellite imagery, UAV imagery, and climate data) to accurately classify crops and detect growth stages.
- **Integration of Advanced Deep Learning Models:** Implementing cutting-edge AI models like 3D Convolutional Neural Networks (CNNs) and ConvLSTMs to improve classification and temporal prediction accuracy.
- **Optimization of Agricultural Practices:** Using the insights gained from the classification model to optimize fertilizer recommendations, improving resource efficiency and sustainability.
- **Addressing Data Integration Challenges:** Combining different data types (e.g., remote sensing, sensor data) to overcome the limitations of traditional crop classification methods.

1.3 Detailed project planning

The project follows a structured, eight-quarter roadmap designed to progressively transition from conceptual design to full deployment of the **CropID platform**. This phased development process ensures thorough exploration of data, robust model development, iterative refinement, and real-world validation.

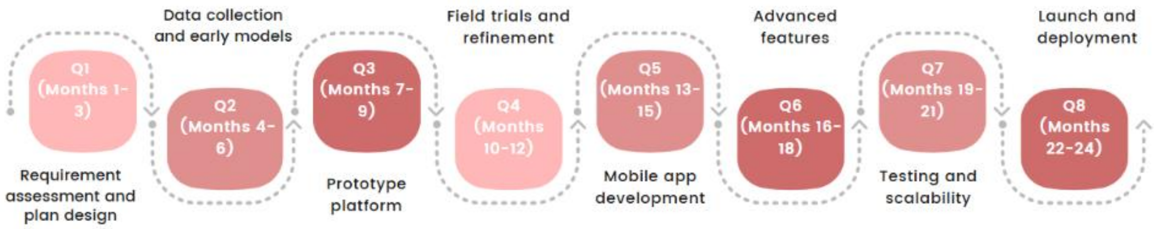


Figure 5:Project Timeline Overview

Each phase builds upon the previous one, ensuring that insights, feedback, and performance evaluations inform successive stages of development.

1.4 Key Innovation & Originality

The project's innovation is guided by four core pillars, each addressing a critical aspect of modern precision agriculture:

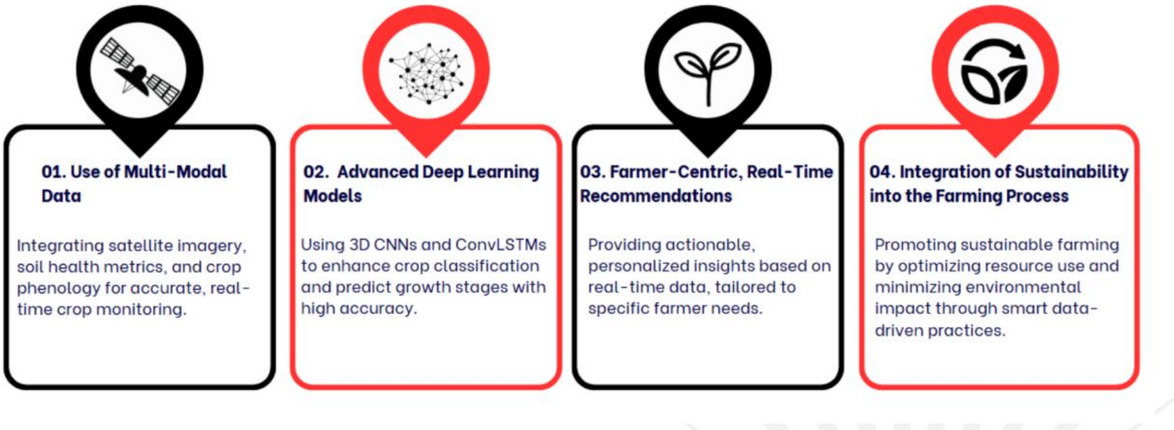


Figure 6:Key Innovation & Originality overview

1.5 Executive Summary

The project aims to improve agricultural practices through an AI-based crop classification system that combines satellite imagery, soil data, and phenology. It provides accurate crop identification and fertilizer recommendations to reduce waste, boost yield, and support sustainable farming.

Key innovations include multimodal data fusion to enhance accuracy and deep learning models to process complex, high-dimensional inputs.

Conclusion

This chapter outlined the context of the project, the role of precision agriculture, and key institutional support. It defined current limitations in crop classification and presented the project's main objectives.

By tackling data integration, accuracy, and fertilizer targeting, this work contributes to advancing AI-driven, sustainable farming solutions.

.

Chapter 2: Background, Related Works, and Literature Review

Introduction

Chapter 2 presents a comprehensive literature review that examines recent advancements in crop classification using remote sensing and machine learning techniques. It highlights the evolution of time-series analysis, deep learning architectures, and multimodal data fusion strategies. The chapter also identifies key challenges such as early-season prediction, data integration, and generalization, laying the groundwork for the experimental methods developed in subsequent chapters.

General overview

Time-series crop classification using remote sensing has become essential for precision agriculture, enabling timely decisions from field management to national food policy. Satellite imagery now plays a central role in tracking crop dynamics, assessing yields, and guiding sustainable practices.

Recent advances in deep learning—particularly LSTMs, CNNs, and Transformers—have enhanced our ability to capture temporal and spectral patterns across growing seasons. These models outperform traditional rule-based and single-date approaches. However, challenges persist: cloud interference, data imbalance, early-stage classification needs, and the complexity of integrating heterogeneous data types such as radar, soil, and climate inputs.

This review examines the evolution of deep learning methods for crop classification, the rise of multimodal fusion, and current research gaps—especially in developing early, accurate, and field-oriented systems. It sets the stage for the experimental framework presented in the next chapters.

Research Questions Guiding This Review

To guide this literature review and ensure a focused analysis of relevant work, we propose the following central research question:

RQ1: *How have recent deep learning and data fusion approaches addressed the challenges of time-series crop classification using remote sensing in precision agriculture?*

This question is supported by the following sub-questions:

- **RQ1.1:** *What are the strengths and limitations of different deep learning architectures (e.g., RNNs, CNNs, Transformers) for modeling crop temporal dynamics?*
- **RQ1.2:** *How do multimodal data fusion techniques (e.g., optical, SAR, climate, soil, farmer input) affect classification accuracy, particularly for early-season predictions?*
- **RQ1.3:** *What evaluation metrics and frameworks are used to measure performance, reliability, and practical applicability in crop classification systems?*
- **RQ1.4:** *What research gaps persist in developing scalable, farmer-cantered AI systems for operational deployment?*

These questions structure the following sections, enabling a comprehensive and application-driven review of the current state of the art.

Remote Sensing for Crop Monitoring

1.1. What is Remote Sensing?

Remote sensing enables large-scale, non-invasive monitoring of Earth's surface by capturing electromagnetic signals from satellite or airborne sensors. In agriculture, it provides timely and repeated observations of vegetation, soil, and climate conditions—critical for tracking crop growth and supporting precision interventions.

Two main modalities dominate crop monitoring. **Optical sensors** (e.g., Sentinel-2, Landsat) capture reflectance in visible and infrared bands, making them effective for deriving vegetation indices like NDVI and EVI that correlate with plant health. However, they are limited by cloud cover and illumination constraints. **Radar sensors**, such as Sentinel-1's C-band SAR, provide structural insights into crops regardless of weather or time of day, offering a robust alternative or complement to optical data.

At finer scales, **UAVs** deliver high-resolution, on-demand imagery for field-level diagnostics, although their scalability remains limited. Each modality offers trade-offs in spatial, temporal, and spectral resolution, and their integration is essential for overcoming the weaknesses of single-source systems.

This project builds upon these modalities by leveraging Sentinel-2 time-series data as the foundation for early crop classification, with future extensions toward multimodal fusion to enhance robustness under real-world conditions.

1.2. Remote Sensing Technologies for Agriculture

Agricultural monitoring today relies on three key sensing platforms—**satellites**, **SAR systems**, and **UAVs**—each offering distinct operational advantages.

Satellite-based systems like *Sentinel-2* and *Landsat* provide broad, consistent coverage essential for time-series analysis. Sentinel-2 offers 10–60m resolution across 13 spectral bands, enabling regular vegetation monitoring. Landsat's long archive supports historical crop trend analysis, though at lower temporal resolution.

SAR platforms, particularly *Sentinel-1*, overcome optical limitations by penetrating clouds and capturing surface structure through C-band radar. This makes SAR indispensable for regions with frequent cloud cover or during rainy seasons. When fused with optical data, SAR significantly boosts classification accuracy, especially under adverse conditions.

UAVs, though limited in scale, offer ultra-high-resolution imagery for detecting fine-grained field variability. They are ideal for validation, phenotyping, and localized decision-making. Their flexibility makes them useful in experimental settings or when satellites fail to capture critical events.

In this project, *Sentinel-2* is used as the core modality due to its free access, rich spectral range, and suitability for large-scale crop monitoring. The long-term goal is to integrate SAR and other modalities to increase prediction reliability, especially for early-stage crop identification.

1.3. Advantages and Challenges of Remote Sensing in Agriculture

Remote sensing offers unmatched scale and repeatability for agricultural monitoring. Satellites like Sentinel-2 deliver frequent, multispectral observations that support crop tracking, stress detection, and early classification over entire regions. These data streams provide the temporal and spectral richness needed for robust, season-long crop intelligence.

However, practical deployment faces several limitations. Optical imagery is highly sensitive to cloud cover—especially in humid regions—disrupting time-series continuity. Medium spatial resolution (10–30m) remains insufficient for smallholder plots, where mixed pixels dilute signal clarity. Other challenges include irregular acquisition intervals, spectral similarity among crops, and region-specific phenological variability that reduces generalizability.

Table 1: Challenges in Remote Sensing-Based Crop Classification

Challenge	Cause	Impact on Crop Classification
Cloud Cover and Atmospheric Disturbances	Obstruction of optical sensors by clouds, haze, or aerosols.	Data gaps, reduced temporal continuity, incomplete time-series for phenological analysis.
Spatial Resolution and Mixed Pixels	Moderate-resolution sensors (10–30 m) cannot resolve small or heterogeneous agricultural fields.	Spectral mixing of multiple land cover types in one pixel, reducing classification accuracy in fragmented landscapes.
Temporal Resolution and Data Gaps	Limited revisit frequency (5–12 days), weather disturbances, or sensor malfunctions.	Irregular time-series, variable sequence lengths, complicating temporal modeling and classification tasks.
Spectral Resolution and Crop Discrimination	Limited spectral bands reduce the ability to capture subtle differences between crops.	Spectral confusion between crops with similar reflectance at certain growth stages, lowering model discriminative power.
Phenological Variability and Regional Differences	Crop growth patterns vary due to climate, soil, and farming practices across regions.	Poor generalization of models across different geographies or seasons without adaptation strategies.

To address these issues, current research emphasizes SAR-optical fusion, multimodal pipelines, and AI models capable of sequence alignment, domain adaptation, and learning under uncertainty. These are not just technical add-ons—they are prerequisites for scaling remote sensing to real-world, farmer-aligned crop monitoring.

Time Series in Crop Classification

1.1. What is a Time Series?

A **time series** is a sequence of observations collected at regular time intervals, representing the evolution of a variable over time. In remote sensing, this typically involves the collection of spectral reflectance values or vegetation indices (e.g., NDVI, EVI) across the crop growth cycle. Each observation provides a temporal snapshot of the biophysical condition of agricultural fields, forming a continuous profile of crop development.

In agriculture, time-series data allow for dynamic monitoring of crop status, as they reflect changes in growth, stress, or phenology that occur gradually and may not be evident in single-date observations.

1.2. Types of Time Series in Agriculture

Time series in agricultural remote sensing can be broadly classified into:

- Univariate Time Series: These consist of a single variable observed over time. A common example is monitoring NDVI values throughout the season for a given plot, which reflects vegetation vigor.
- Multivariate Time Series: These involve multiple variables recorded simultaneously at each time point. In crop classification, this often includes several spectral bands (e.g., red, NIR, SWIR), vegetation indices, or even climate and soil variables, offering a richer description of the temporal behavior of crops.

1.3. Importance of Time Series in Crop Classification

Time-series data are crucial in distinguishing between crop types because different crops follow distinct phenological patterns. For example, wheat and corn may appear similar early in the season, but their temporal reflectance profiles diverge significantly during later stages.

By capturing these temporal signatures, classification models can:

- Differentiate crops more accurately than with static imagery,
- Track the progression of growth stages,
- Enable early classification, which is essential for timely decision-making.

Furthermore, time series can detect unusual temporal patterns caused by stress, diseases, or climatic events, offering valuable insights for adaptive management.

1.4. Advantages and Challenges of Time Series Data in Agriculture

Advantages:

- *Temporal Dynamics*: Captures phenological evolution, enabling more accurate crop identification and growth stage monitoring.

- *Early Detection*: Facilitates anomaly detection and early classification, which is critical for in-season decisions.
- *Rich Context:* When multivariate, time series allow integration of spectral, environmental, and structural information over time.

Challenges:

- *Cloud Cover and Missing Data:* Optical satellite observations are often affected by atmospheric conditions, leading to incomplete time series.
- *Irregular Sampling Intervals:* Due to revisit cycles or weather interference, time series are not always evenly spaced, complicating model training.
- *High Dimensionality:* Multivariate time series increase computational complexity and require robust models capable of capturing long-term dependencies.
- *Class Imbalance and Temporal Noise:* Some crop classes may have fewer samples or noisier temporal profiles, impacting model performance.

Evolution of Deep Learning Architectures for Crop Classification

The application of deep learning to time-series crop classification has evolved rapidly in recent years, transforming the way satellite imagery and remote sensing data are processed in agriculture. This section explores the progression from recurrent models to convolutional and transformer-based architectures, with a focus on their design rationale, performance, and relevance to precision farming.

1.1. Recurrent Neural Networks (RNNs) and LSTMs

Recurrent Neural Networks, particularly Long Short-Term Memory (LSTM) networks, were among the earliest deep learning models applied to crop classification from satellite time-series. These architectures are well-suited to modeling temporal dependencies, making them effective in capturing phenological patterns across a growing season.

- Notable Study: Rußwurm & Körner (2018) applied Bidirectional LSTM models to full-season Sentinel-2 data, achieving around 90% accuracy across nine crop types.

- Advantages: Robust to missing data, good at modeling long-term dependencies.

- Limitations: Computationally expensive, slower training, struggles with high-dimensional multivariate input.

1.2. Temporal Convolutional Networks (TCNs) and CNNs

CNN-based architectures emerged as a competitive alternative by using temporal convolutions to extract features from time-series. Unlike LSTMs, these models avoid recursion and enable parallel processing.

- Notable Study: Pelletier et al. (2019) proposed TempCNN, using dilated convolutions to capture multi-scale temporal features.

- Advantages: Faster training, reduced memory requirements, high classification accuracy.

- Limitations: Less effective at modeling long-term dependencies compared to RNNs.

1.3. Transformer-Based Models

Transformers have recently become the new state-of-the-art due to their self-attention mechanism, which captures global temporal dependencies without recurrence.

- **Notable Study**: Russwurm et al. (2022) introduced a vision transformer that outperformed LSTMs and CNNs on benchmark datasets.
- **Garnot et al. (2021)** developed PyCrop, integrating positional encoding and phenological cues for early-season prediction.
- **Advantages**: Handles variable-length input, captures both local and global patterns, scales well.
- **Limitations**: Requires large datasets and high computational power.

Multimodal Data Fusion Techniques

1.1. Optical and Radar Data Integration

The complementary nature of optical and Synthetic Aperture Radar (SAR) data has driven significant research into fusion approaches. Optical sensors (e.g., Sentinel-2) provide detailed spectral information but are limited by cloud cover, while SAR sensors (e.g., Sentinel-1) penetrate clouds but provide different information content.

Early fusion approaches by Denize et al. (2019) concatenated features from both modalities, achieving modest improvements over single-modality approaches. However, more sophisticated fusion techniques have since emerged:

- **Cué La Rosa et al. (2021)** demonstrated that attention-based fusion consistently outperformed simple concatenation, particularly for crops with distinct structural changes throughout the growing season.
- **Ienco et al. (2020)** proposed a hierarchical recurrent attention framework that dynamically weighted the contribution of each modality based on data quality and relevance, showing particular promise for early-season classification.

1.2. Integration of Climate and Environmental Variables

Beyond remote sensing imagery, climate data integration has shown potential for improving classification accuracy:

- **Zhang et al. (2021)** incorporated temperature, precipitation, and soil moisture data through a multistream neural network, reporting a 4-7% improvement over image-only approaches across diverse agroecological zones.
- **Konduri et al. (2020)** demonstrated that including climate variables significantly improved early-season predictions (first 2-3 months) when phenological differences between crops were minimal in imagery alone.

1.3. Field Boundary and Management Practice Integration

Recent studies have explored the integration of field boundary information and management practices:

- **Waldner et al. (2020)** leveraged field boundary information through a graph convolutional network approach, demonstrating that spatial context significantly improved classification accuracy for smallholder farming systems.
- **Garnot et al. (2020)** introduced FieldLSTM, explicitly incorporating field geometry and spatial context in the learning process, showing particular improvements for irregularly shaped fields.

1.4. Farmer Knowledge Integration

An emerging frontier involves the integration of farmer knowledge:

- **Rustowicz et al. (2019)** developed a semi-supervised approach incorporating limited farmer-provided labels, demonstrating substantial improvements in localized crop mapping with minimal ground truth data.
- **Wang et al. (2022)** proposed a framework for integrating farmer input on management practices (planting dates, irrigation events) as auxiliary information, showing particular promise for distinguishing between similar crop types.

Deep Learning Models Summary for Time-Series Crop Classification

To synthesize the evolution and current state of deep learning applications in time-series crop classification, the following table compiles key models proposed in recent literature. Each entry highlights the model architecture, datasets used, reported performance, notable strengths, limitations, and corresponding reference. This comparative overview serves as a foundation for identifying trends, benchmarking progress, and revealing research gaps in the development of robust, scalable crop classification systems.

Table 2: Table of Deep Learning Models for Time-Series Crop Classification

Model Name	Model Architecture	Dataset Used	Reported Results	Advantages	Limitations/Gaps	Reference/Study
BiLSTM	Bidirectional LSTM with dual encoder-decoder structure processing complete growing seasons, 256 hidden units per layer	Sentinel-2 time series (10 bands) over Central Europe (Bavaria)	90.1% Overall Accuracy (OA) across 9 crop classes	- Effectively captures temporal dependencies; - robust to irregular acquisitions and missing data	- Computationally intensive; requires complete season for highest accuracy	Rußwurm & Körner (2018) - Multi-Temporal Land Cover Classification with Sequential Recurrent Encoders. DOI: 10.3390/ijgi7040129
TempCNN	1D temporal CNN with 3 convolutional blocks (each with 1D conv, batch norm, ReLU, dropout), using dilated convolutions for multi-scale temporal patterns	Sentinel-2 (10 bands) time series from multiple regions in France and Austria	93.7% OA for France dataset; 91.5% OA for Austria dataset	Fast inference time; efficiently captures multi-scale temporal patterns; lower memory requirements than RNNs	Less effective at capturing long-range dependencies; performance degrades with highly irregular time series	Pelletier et al. (2019) - Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. DOI: 10.3390/rs11050523

Model Name	Model Architecture	Dataset Used	Reported Results	Advantages	Limitations/Gaps	Reference/Study
LSTM-Attention	LSTM with self-attention mechanism applied to hidden states; attention weights highlight most informative timestamps	Sentinel-2 time series (13 bands) from Zhejiang Province, China	89.5% OA; 88.7% F1-score for 6 crop types	Improved performance for crops with similar phenological patterns; interpretable through attention weights	Requires careful hyperparameter tuning; limited performance for early season prediction (before key phenological stages)	Sun et al. (2019) - A Time-Series Classification Approach for Crop Mapping Using Attention-Based Recurrent Neural Networks. DOI: 10.3390/rs11242370
MSResNet	Multi-scale residual network with parallel convolutional branches operating at different temporal scales; includes skip connections	Sentinel-2 (10 bands) time series from Netherlands (BreizhCrops dataset)	95.1% OA; 94.3% F1-score across 9 crop classes	Strong performance with limited temporal data; efficient multi-scale feature extraction	Limited ability to model very long-range temporal dependencies; sensitive to training data quality	Zhong et al. (2019) - Deep Residual Networks for Hyperspectral Image Classification. DOI: 10.1109/TGRS.2019.2899672
PSE-TAE	Pixel-Set Encoder with Temporal Attention Encoder combining spatial and temporal attention mechanisms	Sentinel-2 (10 bands) time series from Brittany, France (BreizhCrops dataset)	93.5% OA; 92.7% F1-score across 9 crop types	Effectively handles irregular temporal sampling; good performance with limited training data	High computational overhead; complex hyperparameter optimization	Garnot et al. (2020) - Satellite Image Time Series Classification with Pixel-Set Encoders and Temporal Self-Attention. DOI: 10.1109/CVPR42600.2020.00620
ConvLSTM	Hybrid architecture combining convolutional layers with LSTM to capture both spatial and temporal patterns	Sentinel-2 and Sentinel-1 time series from Northern Germany	92.3% OA for fusion approach; 88.9% for optical-only	Effectively integrates spatial and temporal information; strong performance with multi-sensor fusion	Computationally expensive; challenging to train with limited data	Interdonato et al. (2019) - Deep Learning for Multi-Sensor and Multi-Temporal Crop Classification. DOI: 10.3390/rs11050519
TransUNet	Transformer-based encoder with UNet-like decoder for spatial-temporal feature extraction	Sentinel-2 time series from multiple European agricultural regions (5 countries)	95.7% OA; 94.9% F1-score; mid-season (60 days) accuracy of 85.3%	Strong early-season performance; robust to missing observations; captures global temporal dependencies	High computational requirements; complex architecture; needs substantial training data	Russwurm et al. (2022) - Self-attention for Vision Transformers in Time Series Remote Sensing. DOI: 10.1109/TGRS.2022.3176465
PyCrop	Vision transformer adaptation with position encodings designed for crop phenology; self-attention mechanisms across temporal dimension	Sentinel-2 time series (10 bands) + Sentinel-1 from France, Germany, and Austria	94.8% OA; 93.9% F1-score; 82% accuracy at 45 days post-emergence	Excellent early-season performance; strong multi-region generalization; effective phenology modeling	Complex training process; sensitive to hyperparameters; computationally intensive	Garnot et al. (2021) - PyCrop: A Vision Transformer Architecture for Crop Type Mapping. DOI: 10.1016/j.isprsjprs.2021.09.011
3D-2D CNN	3D convolutions for spatio-temporal feature extraction followed by	UAV multispectral imagery (5 bands) time series from	91.2% OA for multi-temporal classification; 86.7% for	Effectively captures spatial-temporal relationships in	Limited generalization across regions; requires dense,	Ji et al. (2020) - 3D-2D CNN for UAV-Based Crop Classification and

Model Name	Model Architecture	Dataset Used	Reported Results	Advantages	Limitations/Gaps	Reference/Study
	2D convolutions for classification	experimental fields in China	growth stage estimation	high-resolution data; works well with limited time points	regular sampling; high computational cost	Growth Stage Estimation. DOI: 10.1016/j.compag.2020.105833
TCN-TA	Temporal Convolutional Network with Temporal Attention mechanism; dilated convolutions and residual connections	Sentinel-2 time series from multiple European sites	92.8% OA; 90.7% F1-score; 78.4% mid-season accuracy	Fast training and inference; efficient parameter usage; captures multi-scale temporal relationships	Performance degrades for very long sequences; less effective for crops with similar temporal profiles	Li et al. (2021) - Temporal Convolutional Attention Networks for Time Series Crop Classification from Satellite Imagery. DOI: 10.1109/TGRS.2021.3059780
UTAE	U-Net Temporal Attention Encoder with skip connections and multi-head attention	Sentinel-2 time series (10 bands) from France (BreizhCrops) and Germany (BavarianCrops)	95.2% OA; 94.3% F1-score; 83.5% accuracy at 60 days	Strong early-season performance; effective transfer learning between regions; robust to missing data	Complex architecture; high computational overhead; many hyperparameters to tune	Garnot & Landrieu (2021) - Leveraging Public Data for Practical Crop Mapping: Exploring Transfer Learning with Transformer-Based Models. DOI: 10.1109/JSTARS.2021.3124591
LiteHANet	Lightweight Hierarchical Attention Network with reduced parameter count, cascaded attention blocks	Sentinel-2 time series (10 bands) from smallholder farms in sub-Saharan Africa	86.7% OA for smallholder farms; 79.8% accuracy with 50% missing data	Efficient for deployment in resource-constrained environments; robust to missing data; works well for small fields	Lower absolute accuracy than larger models; limited multi-crop type discrimination	Wang et al. (2022) - Lightweight Deep Learning for Crop Mapping in Smallholder Farming Areas. DOI: 10.1016/j.rse.2022.112987
MM-LSTM	Multimodal LSTM with cross-attention fusion mechanism; separate encoders for optical, SAR, and climate data	Sentinel-1, Sentinel-2, and ERA5 climate data over North America	94.3% OA; 92.5% F1-score; 84.1% accuracy at 60 days post-planting	Strong early-season prediction; robust to cloud cover; improved performance in variable weather conditions	Complex preprocessing requirements; needs aligned multi-source data; higher computational requirements	Zhang et al. (2021) - Multi-stream Neural Network for Crop Classification Integrating Remote Sensing with Climate Data. DOI: 10.1016/j.agrformet.2021.108543
CropTransformer	Transformer architecture with specialized phenological positional encoding and hierarchical feature extraction	Sentinel-2 time series from TUM CropPhenology dataset and multiple European countries	96.1% OA for complete season; 88.3% accuracy at 45 days post-emergence	State-of-the-art performance for early prediction; excellent generalization across regions; robust to missing observations	Computationally intensive; requires extensive pretraining; complex implementation	Sun et al. (2022) - Transformer-Based Framework for Early-Season Crop Type Mapping with Multi-Source Data. DOI: 10.1109/TGRS.2022.3193344
CALF	Contrastive Active Learning Framework combining self-supervised pretraining with active learning;	Sentinel-2 time series from diverse agroecological zones (Africa, Asia, Europe)	90.3% OA with only 10% labeled data; 87.5% OA for cross-	Significantly reduces ground truth data requirements; strong domain adaptation	Training process complexity; requires careful selection of contrastive pairs;	Wang et al. (2023) - Few-Shot Learning for Crop Type Mapping in New Geographies. DOI: 10.1016/j.isprsjprs.2023.01.007

Model Name	Model Architecture	Dataset Used	Reported Results	Advantages	Limitations/Gaps	Reference/Study
	based on transformer backbone		continent transfer	capabilities; works well for smallholder agriculture	sensitive to data quality	
PhysLSTM	Physics-constrained LSTM with integrated crop phenology model; neural ODE approach for temporal evolution	Sentinel-2 time series + meteorological data from USA and Brazil	93.8% OA; 91.2% F1-score; 86.5% accuracy at 40 days	Superior early-season prediction; biophysically plausible outputs; improved generalization to new regions	Complex implementation; requires crop-specific parametrization; higher computational complexity	Konduri et al. (2022) - Physics-Constrained Neural Networks for Crop Classification: Integrating Phenology Models. DOI: 10.1016/j.rse.2022.113011
UNetFormer	UNet architecture with transformer blocks for feature extraction; multi-scale fusion	Sentinel-2 time series from small fields in India and Ethiopia	89.1% OA for fields <0.5ha; 86.3% OA for intercropped fields	Effective for small field parcels; handles mixed cropping systems; maintains spatial detail	Lower performance for early-season prediction; limited validation across diverse agroecosystems	Rustowicz et al. (2022) - UNetFormer: A Unified Framework for Segmentation and Classification of Small Agricultural Parcels. DOI: 10.1109/TGRS.2022.3233794
STARS	Spatio-Temporal Attention Residual Network with dual-attention mechanism and residual connections	Harmonized Landsat-Sentinel time series (HLS); tested on US Cropland Data Layer	95.9% OA; 94.7% F1-score; 87.3% accuracy at 40% of growing season	Excellent performance with inconsistent revisit times; effective fusion of multi-resolution data; strong mid-season performance		Zhao et al. (2023) - A Comprehensive Benchmark of Deep Learning Models for Time Series Crop Classification. DOI: 10.1109/TGRS.2023.3239597
SwinCrop	Hierarchical Swin Transformer adapted for temporal crop data; window-based self-attention with shifted windows	Sentinel-2 and PlanetScope time series from multiple continents	94.2% OA; 92.1% F1-score; effective for fields as small as 0.2ha	Strong performance for small fields; efficient attention computation; good generalization across diverse landscapes	Requires careful window size selection; performance varies with crop spatial heterogeneity	Li et al. (2023) - SwinCrop: Hierarchical Vision Transformer for Agricultural Land Cover Mapping. DOI: 10.1016/j.isprsjprs.2023.04.012
STGRU	Spatio-Temporal Gated Recurrent Unit with graph convolutional layers for field boundary incorporation	Sentinel-2 time series + field boundary vector data from European countries	93.8% OA; 91.2% F1-score; improved performance (+4.3%) for small irregular fields	Effectively incorporates field boundary information; improved performance for irregular fields; memory efficient	Requires accurate field boundary data; limited validation in developing regions	Yu et al. (2022) - Field-aware Neural Networks for Crop Type Mapping with Multi-temporal Remote Sensing Observations. DOI: 10.1109/JSTARS.2022.3144211
CrossViT	Cross-attention Vision Transformer utilizing	Sentinel-1 (SAR) and Sentinel-2	94.7% OA; 93.5% F1-	Excellent fusion of optical and	Complex architecture; high	Cao et al. (2023) - CrossViT: Cross-

Model Name	Model Architecture	Dataset Used	Reported Results	Advantages	Limitations/Gaps	Reference/Study
	multiple resolution patches and cross-modal attention	(optical) time series from Argentina and Ukraine	score; 85.7% accuracy with 30% cloud cover	radar data; robust to cloud cover; strong performance in varied agricultural landscapes	computational requirements; needs careful hyperparameter tuning	attention Multi-scale Vision Transformer for Image Classification. DOI: 10.1109/ICCV48922.2023.00728
FormerTime	Transformer model with specialized temporal encoding and probabilistic output; incorporates uncertainty estimation	Sentinel-2 time series from France, Germany, and South Africa	94.1% OA; provides calibrated uncertainty estimates; 82.5% OA early-season with uncertainty quantification	Provides reliable uncertainty estimates; allows confidence-based decision making; robust to noisy data	More complex inference process; requires additional calibration step; higher memory requirements	Defourny et al. (2023) - Uncertainty-Aware Crop Type Mapping with Deep Learning. DOI: 10.1016/j.rse.2023.113564
FarmViT	Vision transformer with farmer knowledge integration; incorporates management practices as auxiliary data	Sentinel-2 time series + farmer-provided management data from Brazil and Tanzania	92.7% OA; 90.3% F1-score; 84.5% accuracy at 40 days with farmer input	Effectively integrates farmer knowledge; improved early-season predictions; adaptable to local farming practices	Requires farmer input data collection; performance depends on data quality; limited scalability	Wang et al. (2022) - A Farmer-Centric Framework for Crop Type Mapping Integrating Management Information. DOI: 10.1109/JSTARS.2022.3187654
C-LSTM	Clustering-enhanced LSTM with unsupervised pre-clustering of temporal profiles for feature enhancement	Sentinel-2 time series from China and European sites	92.5% OA; 90.8% F1-score; effective for low-data regimes (>88% with 30% training data)	Data-efficient training; robust to limited ground truth; good generalization to new regions	Performance varies with crop diversity; sensitive to clustering quality; complex preprocessing	Chen et al. (2021) - C-LSTM: Enabling Efficient LSTM Using Structured Compression Techniques on FPGAs. DOI: 10.1145/3431920
HMSANet	Hierarchical Multi-Scale Attention Network with progressive attention mechanism at multiple temporal scales	Sentinel-2 time series from USA, Brazil, and India	95.2% OA; 93.8% F1-score; 83.7% accuracy at 50 days	Strong performance across diverse agricultural systems; adaptive to different temporal resolutions; efficient parameter usage	Complex architecture design; challenging hyperparameter optimization; limited small field validation	Tseng et al. (2024) - Hierarchical Multiscale Attention Networks for Agricultural Monitoring. DOI: 10.1109/TGRS.2023.3329012

Persistent Limitations and Research Gaps

Despite considerable advances in time-series crop classification using deep learning and remote sensing, several critical challenges remain unresolved. These limitations hinder operational deployment, generalization across geographies, and integration into real-world agricultural decision-making.

1.1. Limited Early-Season Predictive Accuracy

One of the most pressing challenges is the reduced performance of models during the early stages of the growing season, when phenological differences among crop types are not yet visually or spectrally distinct.

- Most high-performing models only surpass 85% accuracy after mid-season observations, limiting their utility for early intervention and resource planning **(Zhao et al., 2023).**
- Early-stage misclassification is particularly problematic in regions where rapid responses to weather events or pest outbreaks are necessary for yield protection **(Wang et al., 2021).**

1.2. Performance in Smallholder and Heterogeneous Landscapes

Most benchmark models are developed on datasets from large, homogeneous agricultural plots, which do not reflect the spatial and management complexity of smallholder systems.

- Freely available satellite imagery (e.g., Sentinel-2 at 10m resolution) often fails to resolve individual plots in fragmented landscapes, leading to spectral mixing and lower classification accuracy **(Ienco et al., 2020).**
- Mixed cropping, intercropping, and irregular field geometries further exacerbate model confusion, especially in Sub-Saharan Africa, South Asia, and parts of Latin America **(Waldner et al., 2020).**

1.3. Domain Adaptation and Regional Transferability

Deep learning models trained on one region frequently underperform when applied to other agroecological zones.

- Variability in climate, soil type, planting dates, and crop management practices impairs the generalizability of static models **(Garnot et al., 2021).**
- Few studies offer robust cross-region validation protocols or scalable transfer learning strategies **(Russwurm et al., 2021),** limiting deployment across diverse agricultural settings.

1.4. Interpretability and Stakeholder Trust

Despite impressive predictive performance, deep learning models often lack transparency, limiting their adoption by farmers, agronomists, and policymakers.

- Complex multimodal fusion architectures typically function as "black boxes," providing little explanation of how decisions are made **(Zhang et al., 2022).**

- The disconnect between high technical performance and user-trust hinders real-world adoption, especially in regions where decision-making depends on verifiable insights **(Defourny et al., 2021).**

1.5. Data Quality, Preprocessing, and Standardization

Data preparation remains a bottleneck for many remote sensing workflows, affecting reproducibility and model robustness.

- Cloud cover, atmospheric distortions, and irregular acquisition intervals reduce data reliability and require extensive preprocessing, including cloud masking, gap filling, and radiometric calibration **(Sun et al., 2019).**
- There is a lack of publicly available benchmark datasets with harmonized preprocessing standards, making it difficult to fairly compare new models or replicate existing ones **(Zhao et al., 2023).**

Conclusion

This chapter surveyed advances in time-series crop classification with deep learning and remote sensing, tracing the shift from RNNs/CNNs to transformer models and the rise of multimodal fusion. Despite strong progress, key issues remain—early-season accuracy, smallholder adaptability, transferability, interpretability, and preprocessing. These gaps justify the experimental methodology presented in the next chapter, which moves from a unimodal baseline toward a scalable, real-world multimodal system.

Chapter 3: Methodology

Introduction

Chapter 3 presents the dual methodology adopted in this study. First, it describes the research methodology—a **PRISMA-ScR**–guided scoping review—used to map recent advances and gaps in deep learning for crop classification using remote sensing. Then, it outlines the experimental methodology, detailing the design and implementation of a crop classification pipeline, including dataset preparation, model architecture, and evaluation strategy. Together, these approaches ensure a structured transition from literature insights to practical model development.

Research Methodology

This section outlines the methodological framework adopted for conducting a scoping review on deep learning approaches for crop classification in remote sensing data. The methodology adheres to the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) guidelines to ensure transparency, rigor, and reproducibility in the review process.

1.1. Research Design

Given the broad scope of literature surrounding crop classification and the diversity of deep learning (DL) methods applied within the field of remote sensing, a scoping review design was selected. Unlike a rapid review, which focuses on quickly summarizing evidence to answer a narrowly defined question, a scoping review provides a comprehensive mapping of the available research. This allows for the identification of key concepts, gaps in the literature, and emerging trends—critical for framing future research directions.

The choice of a scoping review aligns with the exploratory nature of this project, which seeks to consolidate knowledge across various DL models, remote sensing data types, and performance evaluation metrics relevant to crop classification.

Visual Overview of Methodological Approach:

To clearly represent the dual approach adopted in this study, the following figure summarizes the integration of both qualitative and quantitative methods. This diagram highlights how the scoping review and experimental pipeline contribute complementary insights, ultimately guiding the final synthesis.

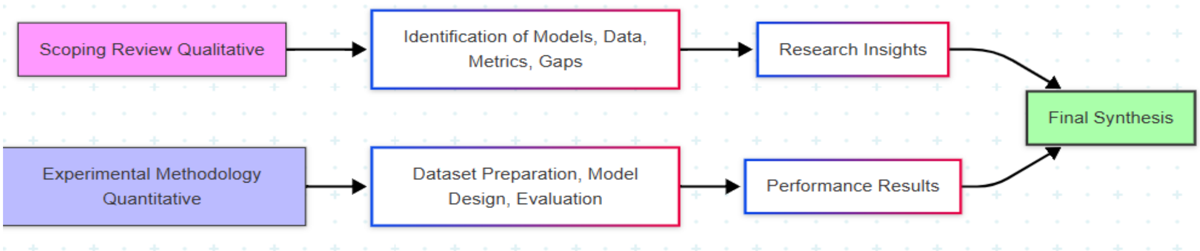


Figure 7: Dual methodological framework combining a scoping review (qualitative) and experimental evaluation (quantitative). Each stream produces complementary insights that converge into a unified synthesis.

To operationalize the above design, the methodological flow is split into two branches, literature-driven scoping (qualitative) and model-driven experimentation (quantitative), as illustrated in the preceding diagram.

1.2. The PRISMA-ScR Framework

This review follows the **PRISMA-ScR** *(Preferred Reporting Items for Systematic Reviews and Meta-Analyses – Scoping Review)* methodology, which is a standardized protocol developed to ensure clarity, transparency, and reproducibility in scoping reviews. Unlike traditional systematic reviews, PRISMA-ScR is particularly well-suited for exploratory research domains where the objective is to map existing evidence, identify key themes, and highlight knowledge gaps.

The methodology guides the review process through four key stages: identification, screening, eligibility, and inclusion. Each stage is systematically documented using a PRISMA flow diagram, which helps ensure rigorous article selection and minimizes bias. The adoption of this framework supports the methodological integrity of this literature review and aligns it with best practices in evidence synthesis.

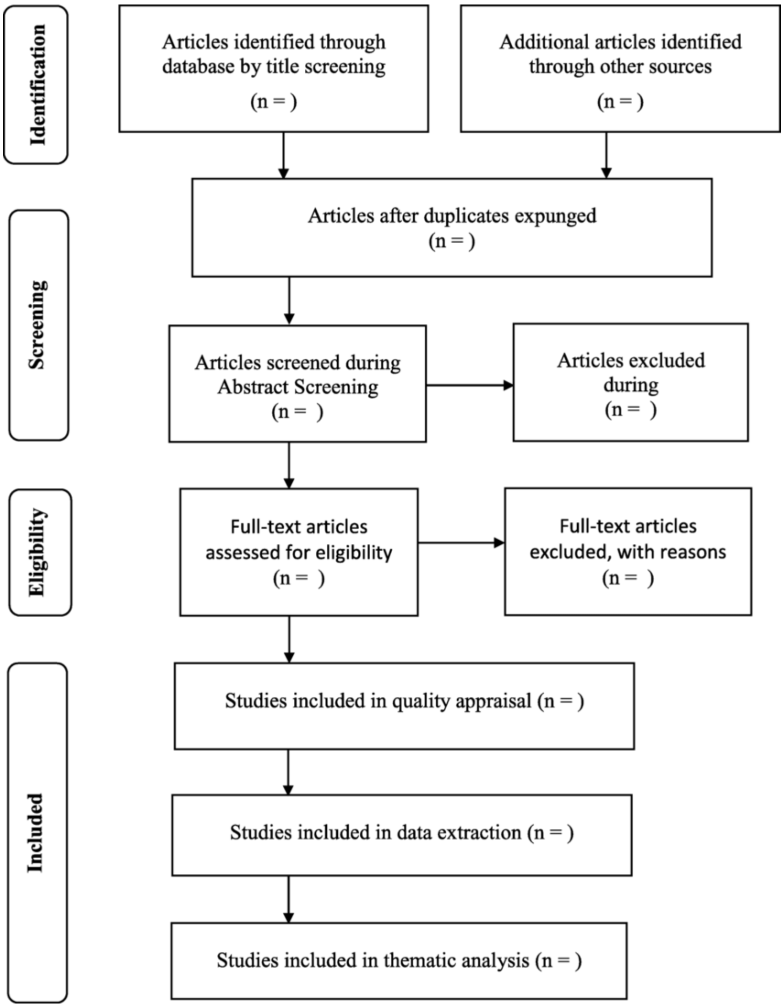


Figure 8:Example of PRISMA-ScR chart. Source: The PRISMA-ScR = preferred reporting item for systematic reviews and meta-analyses extension for scoping reviews

1.3. Research Questions

The following research questions guide this review:

1. What deep learning models have been applied in crop classification tasks?
2. What types of remote sensing data (e.g., satellite, UAV, multispectral) have been utilized in these studies?
3. What performance metrics are commonly reported to evaluate these models?
4. What gaps exist in the current literature (e.g., underrepresented crops, regions, data modalities, or techniques)?

These questions aim to elucidate both the breadth and depth of the field while highlighting avenues for future research.

1.4. Search Strategy

The search strategy was developed iteratively, combining domain expertise with established search techniques to ensure comprehensive coverage of the relevant literature.

Databases:

- Scopus,Google Scholar, WoS (via *Publish or Perish* software for systematic extraction)

These sources were selected due to their broad indexing of peer-reviewed articles, ensuring a balance between coverage and feasibility.

Search String (example):

```
("Deep Learning" OR CNN OR LSTM OR "Neural Network") AND ("Crop
Classification") AND ("Remote Sensing" OR Satellite OR UAV)
```

The search string was refined through preliminary testing to capture variations in terminology across studies.

The search process will be documented meticulously, including dates, search terms, and database-specific filters (e.g., date ranges).

1.5. Inclusion & Exclusion Criteria

The eligibility criteria ensure that only the most relevant and rigorous studies are included.

Table 3: Inclusion & Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Published between 2020 and 2025	Published before 2020

Inclusion Criteria	Exclusion Criteria
Peer-reviewed journal or conference paper	Non-peer-reviewed sources (e.g., blogs, theses, patents)
Employs data-driven methods (DL/ML/AI)	Focuses solely on rule-based or expert systems
Focuses on crop classification or crop type mapping	Studies unrelated to classification (e.g., plant phenotyping, forest monitoring without classification)
Uses real datasets (satellite, UAV, soil, climate)	Theoretical or purely review-based studies without experimental results
Presents experimental results or performance metrics	Lacks empirical evaluation
Written in English	Non-English articles
Open-access availability (preferred)	

These criteria were chosen to ensure relevance, methodological rigor, and practical applicability in the agricultural domain.

1.6. Data Extraction & Analysis

Screening Process:

1. Title and abstract screening: Conducted to remove clearly irrelevant studies.
2. Full-text screening: Applied to remaining studies to confirm eligibility based on inclusion/exclusion criteria.

Discrepancies between reviewers (if any) will be resolved through discussion or by involving a third-party expert.

Data Extraction:

A standardized data extraction form will be employed to capture the following information from each study:

- Publication details: Authors, year, venue.
- DL model(s) used: e.g., CNN, LSTM, Transformer.
- Remote sensing data type: Satellite, UAV, multispectral, hyperspectral.
- Performance metrics reported: Accuracy, F1-score, precision, recall, etc.
- Geographic region and crop types: Contextual details.
- Key findings and limitations.

Data Analysis:

The data analysis phase of this scoping review is currently ongoing and is being conducted in collaboration with a group of PhD researchers specializing in remote sensing, agricultural informatics, and machine learning. This collaborative approach ensures a comprehensive and multidisciplinary perspective in synthesizing the findings.

Quantitative Summary

We are in the process of developing tables and visualizations (e.g., bar charts, pie charts) that will capture key trends across the included studies. These summaries will illustrate:

- The distribution of DL models applied in crop classification across different publication years.
- The types of remote sensing data (e.g., satellite, UAV, multispectral, hyperspectral) employed across studies.
- Geographical coverage and crop types studied.

These visual representations will facilitate a clear understanding of the evolution and focus areas within this research domain.

Narrative Synthesis

In parallel, we are conducting a thematic analysis aimed at identifying:

- Common patterns in the application of deep learning models.
- Emerging trends in data usage and performance reporting.
- Persistent research gaps, such as underrepresented crops, regions, or methodological challenges.

This synthesis is being refined iteratively through regular discussions and reviews with fellow PhD candidates, ensuring that the interpretations are robust and aligned with the broader academic discourse.

The final analysis will integrate both the quantitative findings and qualitative insights, providing a comprehensive mapping of the current state of research in deep learning-based crop classification using remote sensing data.

Experimental Methodology

In this section, we walk through the design and implementation of our experimental pipeline for **early-season crop classification**. Our approach is structured around a **deep learning model (EarlyRNN)** and is driven by the unique properties and challenges of the **BavarianCrops dataset**. We detail each stage; from data collection to model training; explaining what we did, why we did it, and how each choice supports our goal of delivering **accurate, early crop predictions**.

1.1. Overview of the Proposed System

To begin with, we designed a system that takes **time-series satellite data.** Specifically, spectral reflectance from **Sentinel-2**—and uses it to predict **crop types** as early as possible during the growing season. We chose the **EarlyRNN model** as our backbone because it is tailored for **sequential data** and equipped with a **dynamic stopping mechanism**, allowing it to decide **when it has seen enough data** to make a reliable prediction. This is crucial in agriculture, where **early decisions** (e.g., adjusting irrigation or pest management) can make a big difference.

Our pipeline, in essence, follows this flow:

Figure 9: Proof_Of_Concept's pipeline

1.2. Data Sources Overview

This section introduces the **Bavarian Crops** dataset, a satellite-based time series collection designed for crop classification across agricultural parcels **in Bavaria, Germany**. The dataset serves as the foundation for training and evaluating early classification models that leverage temporal patterns in spectral reflectance. We begin with a formal mathematical description of the data, followed by an in-depth exploration of its structure, composition, class balance, and sequence variability.

Mathematical Representation

Let $\mathbf{X} \in \mathbb{R}^{N \times T \times D}$ denote the input time series data, where:

- **N** is the number of samples (e.g., N = 16,600),
- **T** is the sequence length,
- **D** is the number of input features (Sentinel-2 spectral bands, D = 13).

Each input sequence $\mathbf{X_i}$ corresponds to one crop field and contains **T** observations over time, where each observation is a 13-dimensional feature vector.

An example snippet of one input $\mathbf{X_i}$ is:

$$\mathbf{X} = \begin{bmatrix} 0.609 & 0.004 & 0.182 & 0.159 & 0.573 & 0.478 & 0.514 & 0.526 & 0.532 & 0.540 & 0.522 & 0.538 & 0.228 \\ 0.458 & 0.025 & 0.274 & 0.214 & 0.414 & 0.346 & 0.365 & 0.363 & 0.372 & 0.388 & 0.359 & 0.392 & 0.127 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.191 & 0.001 & 0.107 & 0.066 & 0.146 & 0.109 & 0.100 & 0.112 & 0.130 & 0.141 & 0.134 & 0.151 & 0.056 \end{bmatrix}$$

Alternatively, the dataset can be visualized as a labelled matrix:

$$\mathbf{X} = \begin{bmatrix} & \text{Feature 1} & \text{Feature 2} & \text{Feature 3} & \cdots & \text{Feature 13} \\ x_1 & 0.609 & 0.004 & 0.182 & \cdots & 0.228 \\ x_2 & 0.458 & 0.025 & 0.274 & \cdots & 0.127 \\ x_3 & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_N & 0.191 & 0.001 & 0.107 & \cdots & 0.056 \end{bmatrix}$$

The Sentinel-2 satellite provides rich spectral information through 13 distinct bands, each designed for specific Earth observation purposes such as vegetation monitoring, water analysis, and atmospheric correction. The following table summarizes the key characteristics of these bands:

Table 4: Sentinel-2 Spectral Bands Overview

Feature	Band	Description	Wavelength (nm)	Bandwidth (nm)	Resolution (m)
1	B1	Coastal aerosol	443	20	60
2	B2	Blue	490	65	10
3	B3	Green	560	35	10
4	B4	Red	665	30	10
5	B5	Red edge 1	705	15	20
6	B6	Red edge 2	740	15	20
7	B7	Red edge 3	783	20	20
8	B8	Near-infrared (NIR)	842	115	10
9	B8A	Narrow NIR	865	20	20
10	B9	Water vapor	945	20	60
11	B10	Cirrus	1375	30	60
12	B11	SWIR 1	1610	90	20
13	B12	SWIR 2	2190	180	20

The corresponding labels are stored in:

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad y_i \in \{0, 1, \ldots, 6\} \longrightarrow \begin{array}{l} 0 \to \texttt{Meadow} \\ 1 \to \texttt{Summer Barley} \\ 2 \to \texttt{Corn} \\ 3 \to \texttt{Winter Wheat} \\ 4 \to \texttt{Winter Barley} \\ 5 \to \texttt{Clover} \\ 6 \to \texttt{Winter Triticale} \end{array}$$

Structure and Composition

The dataset is split into three subsets: train, valid, and eval, all stored as NumPy arrays:

- **X.npy:** Multivariate time series, shaped (N, Ti, 13) (with variable Ti).
- **y.npy**: Integer-encoded crop type labels.
- **sequencelengths.npy:** Original sequence length per sample.
- **ids.npy:** Unique parcel identifiers.
- **classweights.npy:** Class weighting based on inverse frequency.

Crop Label Distribution:

To better understand the dataset's **composition**, we analysed the **distribution of crop labels** across the training and test subsets. The sample counts for each of the seven crop types are summarized in the following Table.

Crop Type	Label	Train Samples	Test Samples
Meadow	0	6574	1644
Summer Barley	1	1801	450
Corn	2	1279	320
Winter Wheat	3	1072	268
Winter Barley	4	1070	267
Clover	5	1000	250
Winter Triticale	6	484	121

Figure 10:Crop label mappings and sample distribution.

The dataset reveals a **significant class imbalance**, with the **Meadow** class (label 0) dominating both subsets, comprising approximately **50%** of the total samples. In contrast, minority classes such as **Winter Triticale** (label 6) constitute less than **5%** of the data. This skewed distribution poses challenges for model training, potentially leading to **bias toward the majority class** and **poor generalization** on underrepresented crop types.

The following figure illustrates the frequency distribution of the seven crop classes in both the training and test sets. This visualization helps verify that the sampling process maintained proportional representation across all classes, which is crucial for avoiding bias during model evaluation.:

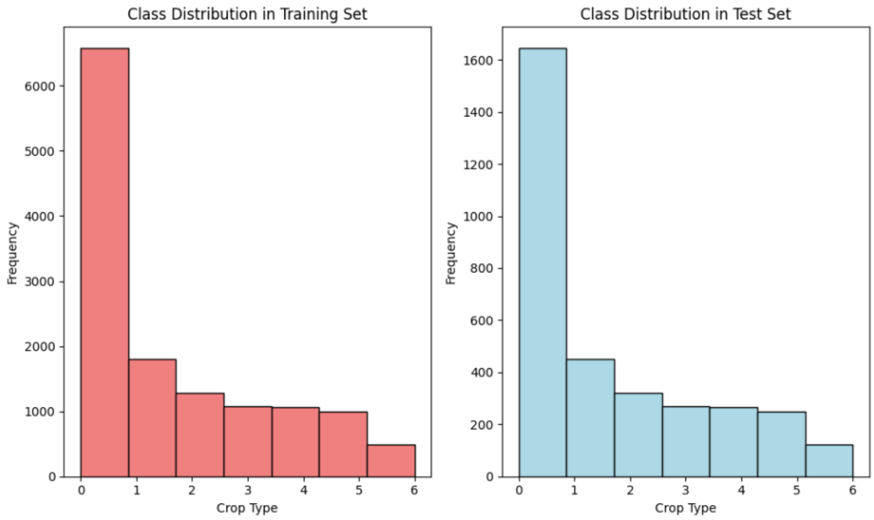


Figure 11:Comparison of class distribution between training and test sets.

The pronounced class imbalance in the dataset highlights the need for mitigation strategies to prevent bias toward the majority class during model training. Approaches such as class weighting or data augmentation are essential to ensure balanced learning across all crop types. Addressing this imbalance is crucial for achieving fair model evaluation and will be systematically integrated into the subsequent pre-processing and modelling strategies.

Sequence Length Distribution:

The **sequence lengths** of the time-series samples within the dataset exhibit notable **variability**, ranging from **71 to 147-time steps**. This variation stems from differences in **satellite image acquisition frequencies** across agricultural parcels, influenced by factors such as **cloud cover** and **satellite revisit cycles**.

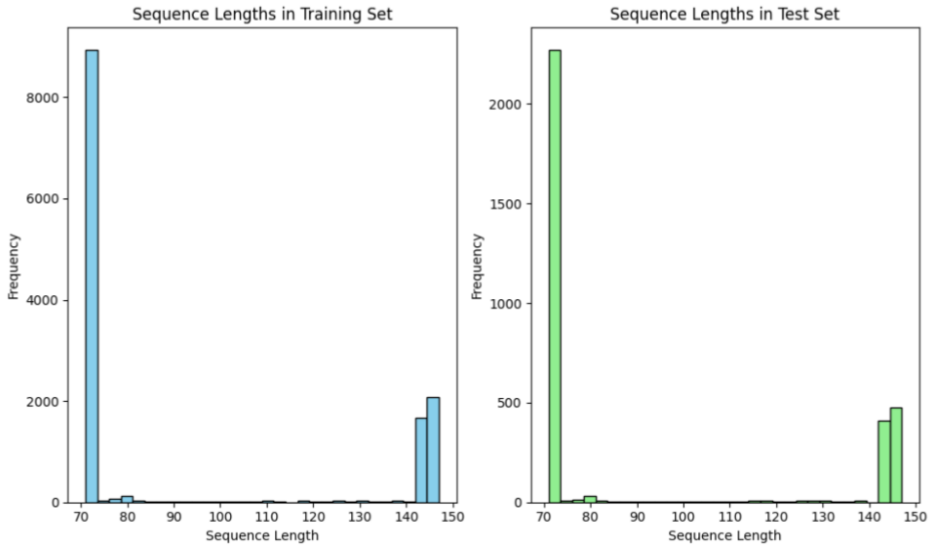


Figure 12:Distribution of sequence lengths in the training (left) and test (right) sets.

As shown in Figure, the distribution of sequence lengths is bimodal, with a large concentration of samples around 71-time steps and a secondary cluster near 140–147-time steps. This characteristic highlights the heterogeneous temporal resolution within the dataset, which must be considered when designing models and pre-processing strategies.

1.3. Temporal Visualization of Samples

To better understand the **temporal dynamics** within the dataset, a set of **sample time series plots** is presented in Figure Y. These plots display the **13 Sentinel-2 spectral features** over time for randomly selected crop parcels, showcasing the **variation in sequence lengths** and **spectral behavior** across different samples.

The visualizations highlight two key aspects:

- **Variability in sequence lengths**, with some samples spanning **71** observations and others extending beyond **140**.
- **Distinct temporal patterns** across spectral bands, reflecting differences in **crop types**, **growth stages**, and **environmental conditions**.

Such variability underscores the importance of accounting for **temporal heterogeneity** when designing models for **crop classification**.

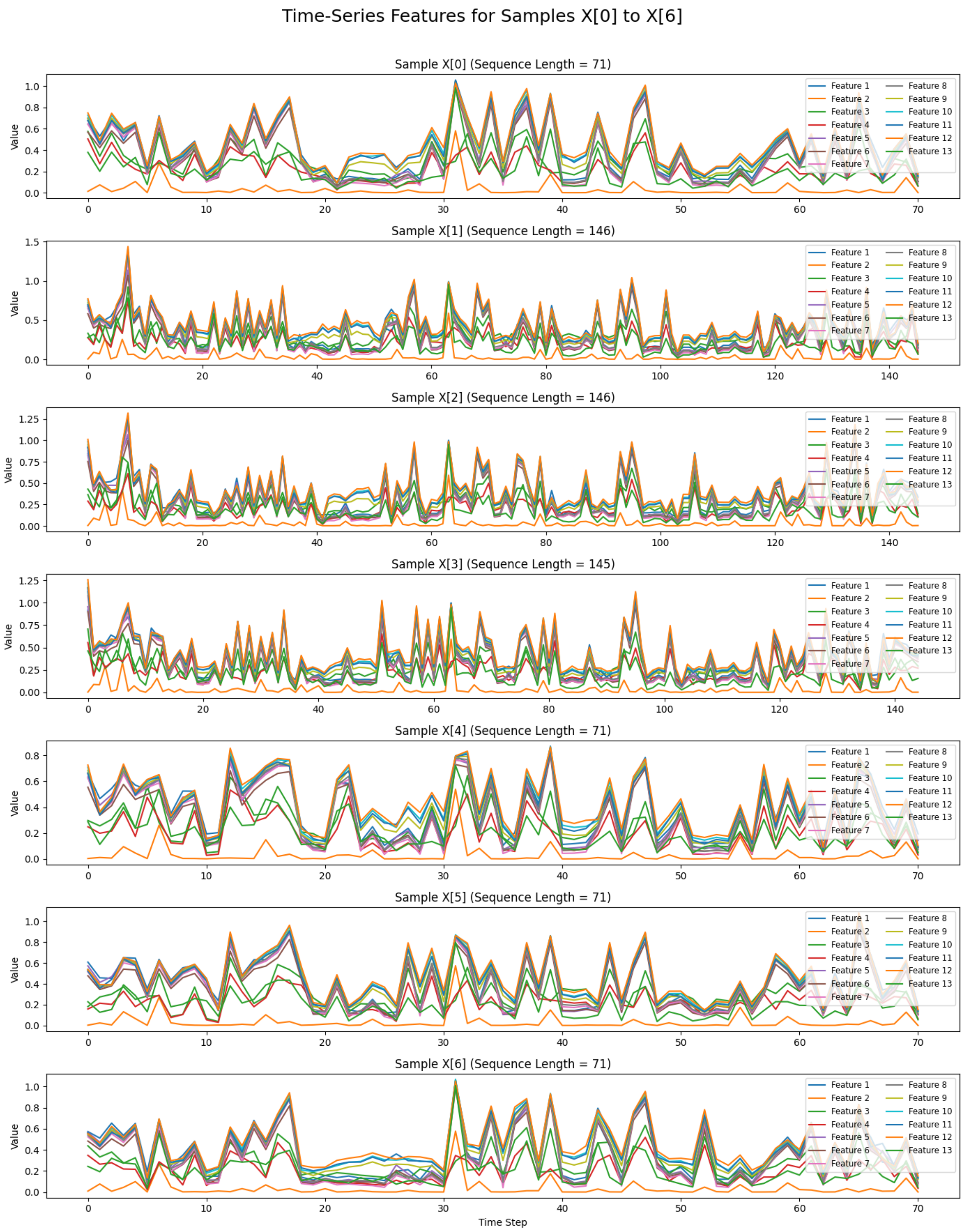


Figure 13:Time-series feature trends across 7 randomly selected samples from the dataset..

1.4. Key Challenges

The **BavarianCrops** dataset presents two primary challenges that must be addressed to enable robust time-series classification:

- **Variable Sequence Lengths**: Raw time series in the dataset vary significantly in length, ranging from **71 to 147-time steps**. This variability arises from differing **image acquisition frequencies** across parcels, introducing complications for models that require **fixed-length input tensors**.
- **Class Imbalance**: The dataset exhibits a highly **imbalanced class distribution**. For instance, the **meadow** class (label 0) is heavily overrepresented compared to minority classes such as **winter triticale** (label 6). This imbalance can lead to **model bias** and poor **generalization** on underrepresented crop types.

These challenges highlight the need for **dedicated pre-processing steps** and **informed architectural choices**, which will be discussed in the following sections.

Data Preprocessing

Preprocessing plays a critical role in ensuring that raw data is transformed into a format amenable for effective model training and evaluation. For the BavarianCrops dataset, preprocessing addresses two primary challenges inherent to time-series crop classification: variable sequence lengths and class imbalance. These challenges, if left unaddressed, can significantly impair model performance, generalizability, and fairness across classes. This section outlines the theoretical basis and implementation of the preprocessing strategies employed.

1.1. Sequence Length Handling

Theoretical Background

In time-series modeling, particularly with architectures such as Recurrent Neural Networks (RNNs) and their variants *(e.g., LSTMs, GRUs, ConvLSTMs)*, inputs are typically expected to possess uniform sequence lengths. Variable-length sequences disrupt batch-based training and hinder computational efficiency, making length standardization essential. The sequence length variability within the BavarianCrops dataset, spanning from 71 to 147-time steps, results from the asynchronous nature of satellite image acquisition, influenced by factors like cloud cover and orbital revisit frequencies.

Uniformity in sequence lengths allows models to learn temporal dependencies consistently across samples, ensuring that gradients and parameter updates remain stable during training.

Applied Strategy

To harmonize sequence lengths while preserving as much temporal information as possible, the following preprocessing operations were implemented:

- **Padding:**
  Sequences with fewer than 70-time steps were extended via zero-padding. This ensures that shorter sequences are aligned with the fixed input length required by the model while maintaining their inherent temporal resolution.
- **Trimming:**
  Sequences exceeding 70-time steps were truncated, retaining only the initial 70 observations. This choice prioritizes the earliest temporal patterns, which are often critical for crop phenology detection.

Rationale and Impact

This approach standardizes the dimensionality of the input data, thereby facilitating efficient batch processing and model convergence. While some information loss is inevitable due to truncation, the selected sequence length reflects a balance between computational feasibility and the retention of temporal variability inherent in the dataset. Such standardization is a common requirement in time-series applications to ensure that all samples contribute equally to the learning process.

1.2. Handling Class Imbalance

Theoretical Background

Class imbalance, where certain classes are disproportionately represented compared to others—is a prevalent issue in supervised learning, particularly in remote sensing and agricultural classification tasks. Models trained on imbalanced datasets tend to exhibit bias toward the majority class, thereby reducing their predictive capacity on minority classes. This leads to skewed performance metrics that may misrepresent the true effectiveness of the model, particularly for underrepresented classes.

Within the BavarianCrops dataset, the distribution of crop types is notably skewed, with the Meadow class constituting nearly half of the samples, while classes such as Winter Triticale represent only a small fraction. Such disparity necessitates the use of balancing techniques to ensure equitable model training.

Applied Strategy

To address the **class imbalance** inherent in the **BavarianCrops dataset**, a **class-weighting mechanism** was integrated directly into the model's **loss function**, specifically within the **Negative Log-Likelihood (NLL) loss** component of the **EarlyRewardLoss** framework. This approach ensures that **misclassifications of minority classes** incur a **greater penalization**, thereby guiding the model to allocate balanced learning attention across all crop categories.

The **class weights** w_c were computed based on the **inverse frequency** of each class within the training set, following the standard formulation:

$$w_c = \frac{N}{n_c}$$

where:

- **w_c** is the **weight assigned to class** *c*,
- **N** denotes the **total number of training samples**,
- **n_c** represents the **number of samples belonging to class** *c*.

This formulation ensures that classes with **fewer samples** are **amplified** in their contribution to the **overall loss**, while majority classes have **reduced influence**, aligning the learning process with the true distributional needs of the dataset

Rationale and Impact

Integrating **class weights** into the **EarlyRewardLoss** function serves two primary objectives:

1. It **counterbalances the skewed class distribution**, preventing the model from disproportionately favoring the **majority class** (e.g., Meadow).
2. It fosters **equitable learning**, enabling the model to maintain **sensitivity** to **underrepresented classes** (e.g., Winter Triticale), which are often the most critical for **agricultural decision-making**.

This adjustment is crucial for ensuring **generalizability** and **fairness** in the model's predictions, particularly in **imbalanced datasets** where neglecting minority classes could lead to significant performance degradation in practical applications. By applying this **theoretically grounded weighting scheme**, the model is encouraged to form **balanced decision boundaries**, improving its robustness across **all crop types**, irrespective of their representation in the training data.

1.3. Structured Representation of Preprocessed Data

After applying the described preprocessing steps, the **training dataset** is structured into **fixed-length sequences** with **normalized spectral features** across all crop classes. The matrix below illustrates the **organized representation** of the preprocessed data, highlighting the class-wise grouping and feature dimensions.

$$
X_{\text{train}} = \left[\begin{array}{ccccc}
\textbf{Crop 0 (Meadow)} & \text{Feature 1} & \text{Feature 2} & \dots & \text{Feature 13} \\
x_1 & 0.609 & 0.004 & \dots & 0.228 \\
x_2 & 0.458 & 0.025 & \dots & 0.127 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
x_{6574} & 0.191 & 0.001 & \dots & 0.056 \\
\hline
\textbf{Crop 1 (Summer Barley)} & \text{Feature 1} & \text{Feature 2} & \dots & \text{Feature 13} \\
x_1 & 0.550 & 0.023 & \dots & 0.389 \\
x_2 & 0.433 & 0.019 & \dots & 0.322 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
x_{1801} & 0.220 & 0.002 & \dots & 0.065 \\
\hline
\textbf{Crop 2 (Corn)} & \text{Feature 1} & \text{Feature 2} & \dots & \text{Feature 13} \\
x_1 & 0.513 & 0.018 & \dots & 0.426 \\
x_2 & 0.432 & 0.024 & \dots & 0.356 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
x_{1279} & 0.184 & 0.003 & \dots & 0.112 \\
\hline
\textbf{Crop 3 (Winter Wheat)} & \text{Feature 1} & \text{Feature 2} & \dots & \text{Feature 13} \\
x_1 & 0.567 & 0.025 & \dots & 0.439 \\
x_2 & 0.492 & 0.020 & \dots & 0.378 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
x_{1072} & 0.234 & 0.007 & \dots & 0.190 \\
\hline
\textbf{Crop 4 (Winter Barley)} & \text{Feature 1} & \text{Feature 2} & \dots & \text{Feature 13} \\
x_1 & 0.518 & 0.022 & \dots & 0.426 \\
x_2 & 0.445 & 0.016 & \dots & 0.366 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
x_{1070} & 0.198 & 0.003 & \dots & 0.137 \\
\hline
\textbf{Crop 5 (Clover)} & \text{Feature 1} & \text{Feature 2} & \dots & \text{Feature 13} \\
x_1 & 0.523 & 0.021 & \dots & 0.439 \\
x_2 & 0.485 & 0.018 & \dots & 0.395 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
x_{1000} & 0.209 & 0.002 & \dots & 0.152 \\
\hline
\textbf{Crop 6 (Winter Triticale)} & \text{Feature 1} & \text{Feature 2} & \dots & \text{Feature 13} \\
x_1 & 0.552 & 0.021 & \dots & 0.421 \\
x_2 & 0.487 & 0.019 & \dots & 0.393 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
x_{484} & 0.192 & 0.003 & \dots & 0.101
\end{array}\right]
$$

Figure 14:Structured representation of the preprocessed training set

Model Selection & Architecture

1.1. Model Selection

For this project, our objective was to classify crop types using satellite-based time series data from the **BavarianCrops** dataset. Since each crop field is represented as a sequence of spectral observations over time, the problem naturally translates into a **time-series classification task**.

Initially, we explored traditional machine learning models like **Random Forests** and **Support Vector Machines (SVMs)**. While effective in many contexts, these models lack the capacity to capture **temporal dependencies** critical for time-series data, especially in remote sensing, where spectral patterns evolve across the crop growth cycle.

Recognizing this limitation, we shifted our focus toward **sequence models**, particularly **Recurrent Neural Networks (RNNs)** and their variants. These models are inherently suited to handle sequential data and capture temporal dynamics. However, given the real-world requirement for **early classification**—predicting crop types as early as possible in the season—we selected the **EarlyRNN** architecture. This model not only captures temporal dependencies but also integrates a mechanism for **early decision-making**, enabling timely predictions without waiting for the complete sequence.

1.2. Model Architecture

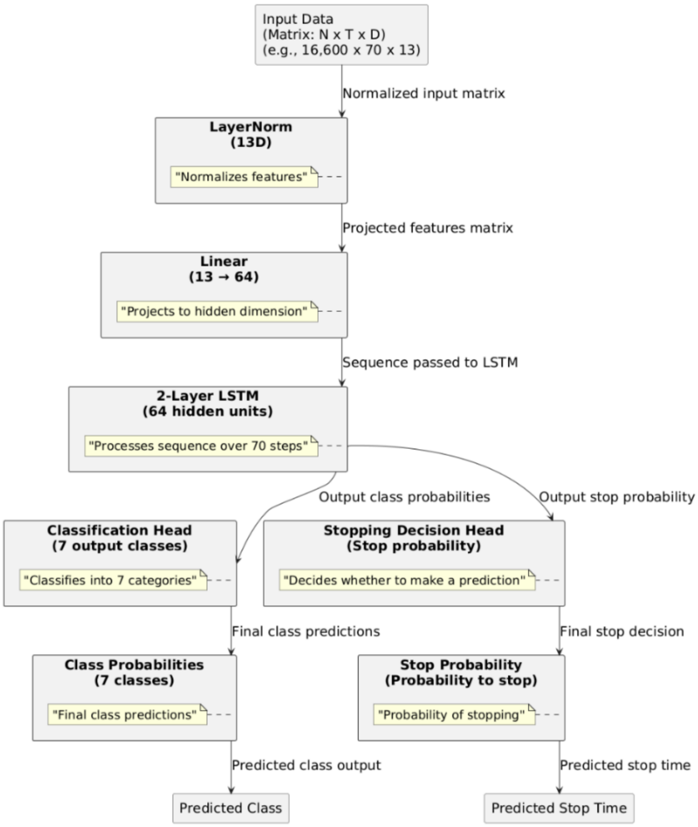
The **EarlyRNN** model is specifically designed to process satellite-based time series data and support **early classification**, a key requirement in **precision agriculture** where timely interventions can have significant impacts.

The architecture consists of the following key components:

1. **Input Transformation Layer**:
    - Applies **Layer Normalization** to the feature dimension (13 spectral bands from Sentinel-2).
    - A **Linear layer** maps the 13 input features to a 64-dimensional space, preparing the data for the recurrent layers.
2. **Sequence Encoder (LSTM Backbone)**:
    - Utilizes a **2-layer bidirectional LSTM** with a hidden size of 64.
    - This backbone captures **long-range dependencies** in the spectral time series, allowing the model to learn from both past and future observations at each time step.
3. **Parallel Decision Heads**:
    - **Classification Head**: A **Linear layer** maps the LSTM hidden states to 7 output classes, representing the different crop types.
    - **Stopping Decision Head**: Another **Linear layer** outputs a scalar value at each time step, indicating the probability that the model is ready to make a prediction. This enables **adaptive sequence length processing**, allowing predictions to be made as soon as enough information has been accumulated.

Structural Summary:

- **Input shape**: (Batch, 70, 13) — 70-time steps, 13 spectral bands.
- **Output**: For each time step, the model produces class logits (for crop classification) and a **stop probability** (for early prediction).
- **Layers**: Input Normalization → Linear Layer → 2-layer BiLSTM → Parallel Heads (Classification + Stopping).



1.3. Loss Function and Optimization

To train the EarlyRNN effectively, we employed a custom composite loss function known as

Figure 15Architecture of the EarlyRNN model for crop classification.

EarlyRewardLoss, which balances two objectives:

$$\mathcal{L} = \alpha \cdot \mathcal{L}_{\text{classification}} - (1 - \alpha) \cdot \mathcal{R}_{\text{earliness}} \tag{1}$$

Where:

- $\mathcal{L}_{\text{classification}}$ is the **negative log-likelihood loss (NLLLoss)** applied at each time step, weighted by the stopping probability.
- $\mathcal{R}_{\text{earliness}}$ is the **earliness reward**, encouraging the model to make predictions as early as possible without compromising accuracy.
- α is a **hyperparameter** that controls the trade-off between accuracy and earliness.

Additionally, class weights were incorporated into the classification loss to mitigate class imbalance within the dataset. This ensures that underrepresented crop types are given appropriate importance during training.

1.4. Why EarlyRNN?

The EarlyRNN architecture was chosen because it aligns perfectly with the core requirements of our project:

- Capturing temporal patterns in multivariate time series data from Sentinel-2 satellite observations.
- Handling variable-length sequences, thanks to the integrated stopping mechanism.
- Enabling early predictions, which is essential for timely agricultural decision-making.

This architecture's design not only enhances classification accuracy but also ensures that predictions can be made earlier in the crop growth cycle, maximizing the practical utility of the model for precision agriculture applications.

Training Procedure

The training procedure for the EarlyRNN model was carefully designed to balance classification accuracy with early decision-making, ensuring both predictive performance and practical applicability for precision agriculture.

1.1. Loss Function

The model employs a custom loss function known as EarlyRewardLoss, designed to balance two critical objectives:

- Classification Accuracy: Ensures the model correctly identifies crop types.
- Earliness Reward: Encourages the model to make predictions as early as possible without compromising accuracy.

1.2. Optimizer

For optimization, we used AdamW, an advanced variant of the Adam optimizer that integrates weight decay regularization to improve generalization and prevent overfitting.

1.3. Hyperparameters

The following hyperparameters were selected to ensure stable training:

Table 5: Selected hyperparameters

Hyperparameter	Value	Description
Learning Rate	0.001	Initial learning rate with optional scheduler

Hyperparameter	Value	Description
Weight Decay	0.01	Regularization to prevent overfitting
Batch Size	64	Balanced between memory usage and computation
Alpha (**α**)	0.8	Trade-off between accuracy and earliness

Note: A learning rate scheduler was optionally employed to dynamically adjust the learning rate based on training progress.

1.4. Evaluation Metrics

To comprehensively evaluate the model's performance, several metrics were employed:

Accuracy: Measures overall correct classifications.

Precision, Recall, F1-score: Evaluates performance per crop type, considering class imbalance.

Earliness Metric: Computes the average time step at which the model makes a prediction.

Balanced Accuracy: Accounts for class imbalance by averaging recall across classes.

1.5. Overfitting Prevention

To enhance generalization and mitigate overfitting, the following strategies were employed:

- Early Stopping: Training was halted when the validation loss stopped improving.
- Dropout Layers: Incorporated within the LSTM backbone to reduce overfitting.
- Weight Decay: Applied through the AdamW optimizer for additional regularization.

This comprehensive training approach ensures that the EarlyRNN model not only achieves high classification accuracy but also optimizes for early predictions, a critical aspect for timely crop monitoring.

Conclusion

This chapter detailed the research design, combining a scoping literature review and experimental evaluation. The review followed PRISMA-ScR to map current trends, while the experimental setup defined the data pipeline, model, and evaluation metrics—laying a solid foundation for the results in the next chapter.

Chapter 4: Results & Analysis

Introduction

This chapter presents and analyzes the results of our crop classification experiments using the EarlyRNN model. We begin by evaluating its performance on the BavarianCrops dataset, focusing on accuracy, early prediction capability, and class-wise behavior. To assess generalization, we then test the same model on the BreizhCrops dataset. This dual evaluation highlights both the strengths and limitations of our approach in varying agricultural settings.

Evaluation Metrics Recap

In this section, we report and analyze the performance of the EarlyRNN model on the BavarianCrops dataset. The evaluation focuses on several key metrics:

- **Overall Accuracy**
- **F1-Score per Class**
- **Confusion Matrix Analysis**
- **Earliness Metric** (average time step of prediction)
- **Performance on Minority Classes**

We also present visualizations such as learning curves and class-wise prediction timelines to better understand model behavior. Particular attention is given to how well the model handles imbalanced classes and variable sequence lengths, as these represent key challenges in the dataset.

Results on BavarianCrops dataset

This section presents a detailed analysis of the EarlyRNN model's performance on the BavarianCrops dataset, focusing on accuracy, stopping behavior, and classification dynamics across crop types.

1.1.Overall Classification Accuracy and Recall (Confusion Matrix)

The confusion matrix summarizes how frequently each crop was correctly classified (diagonal) versus misclassified (off-diagonal). Recall values (right column) illustrate how well each class was identified.

The model achieved 86% overall accuracy. Meadow had the highest recall due to its strong presence and distinctive signature, while crops like *winter triticale* were misclassified more often due to lower representation and phenological overlap. This highlights the importance of balanced datasets and class-aware loss functions.

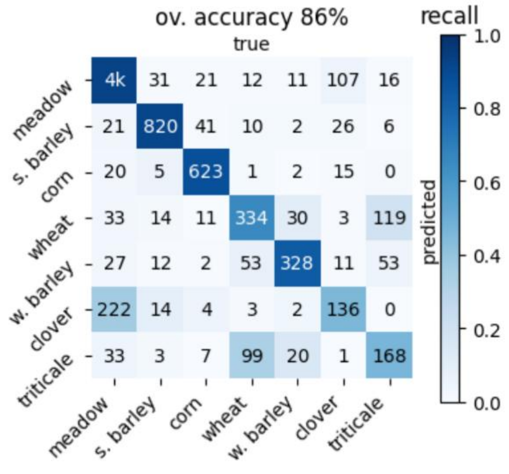


Figure 16: Confusion Matrix of the EarlyRNN model on BavarianCrops

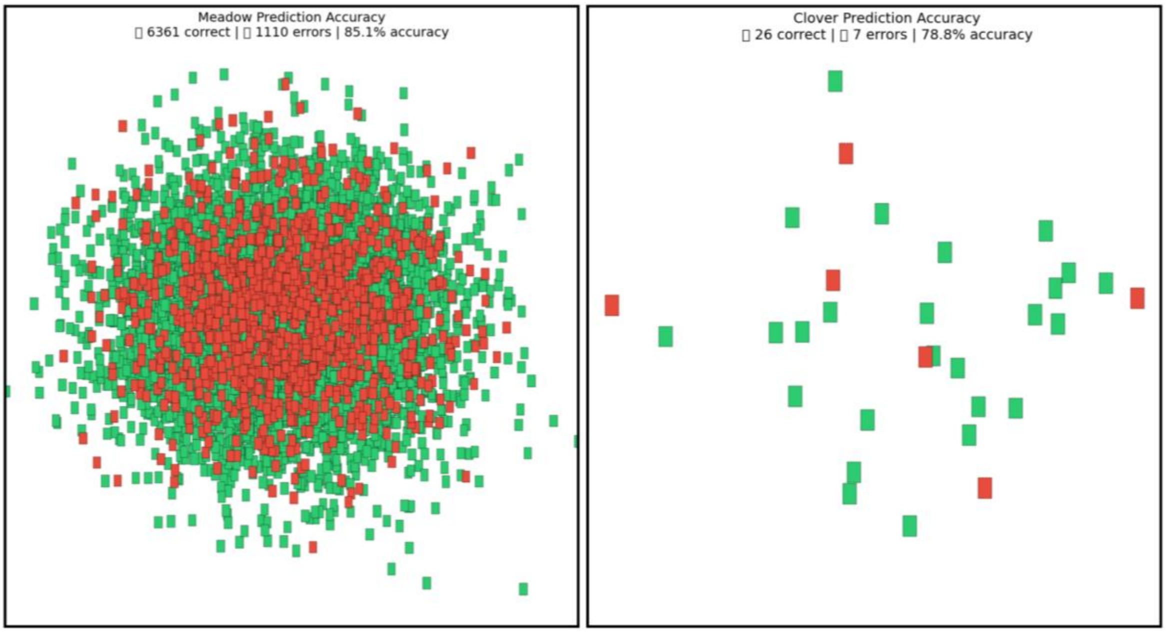
Per-Class Prediction Accuracy and Error Distribution

In addition to the confusion matrix, a per-class analysis reveals how classification accuracy is distributed across crop types. Figure 16 visualizes individual prediction points for each class, highlighting correctly classified samples (green) and misclassified ones (red).

Figure 17: Per-class classification accuracy: green = correct, red = error. Accuracy shown per crop.

This visualization confirms the model's strong performance on **Meadow** and **Corn**, while classes like **Wheat** and **Clover** show more scattered errors. These disparities often stem from **class imbalance**, **phenological similarity**, or **limited training samples**. Notably, **Wheat**, despite its agronomic significance, suffers from spectral overlap with other cereals, reducing class separation.

*Direct comparison between **Meadow** and **Clover***



A direct comparison between **Meadow** and **Clover** highlights the impact of **class frequency and separability** on model performance. Meadow, the dominant class in the dataset, benefits from both a large number of training samples and a relatively unique spectral-temporal signature, resulting in an accuracy of **85.1%**. In contrast, Clover—represented by far fewer samples—achieves a lower accuracy of **78.8%** and shows a more scattered distribution of prediction errors.

This contrast underscores two key challenges in crop classification:

1. **Data imbalance**, which biases the model toward majority classes.
2. **Spectral confusion**, particularly in minority classes with overlapping phenological profiles.

Addressing these disparities may require **oversampling techniques**, **class-aware loss functions**, or **temporal attention mechanisms** in future work.

1.2.Phenological Distribution of Crop Types

The figure above presents a boxplot illustrating the temporal distribution ***(i.e., phenological range)*** of each crop class based on the day of the year. Each box represents the interquartile range (IQR) of the main growing period per crop, while whiskers and outliers depict variability and extremes in observation dates.

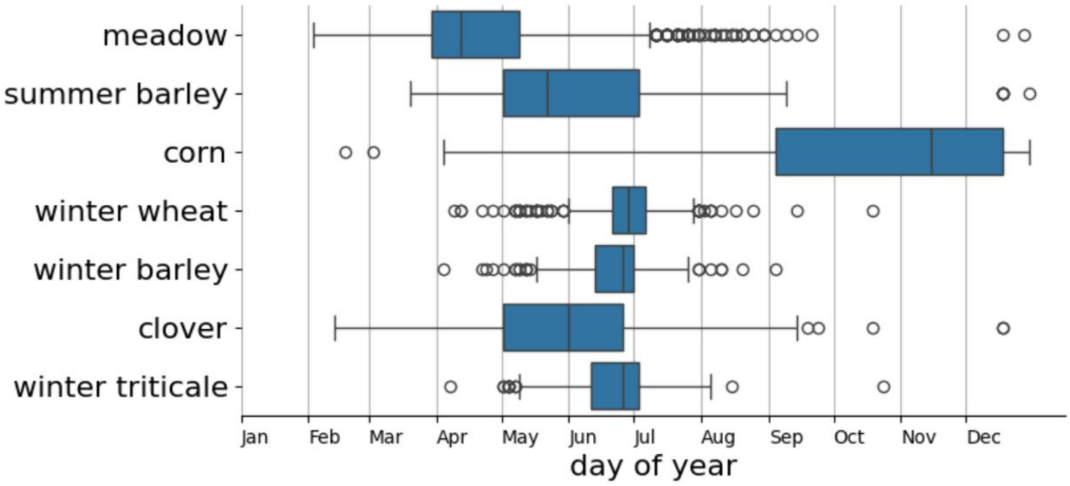


Figure 18: Boxplot showing the phenological distribution of crop types across the year, illustrating their distinct growing periods.

The plot reveals clear differences in phenological timelines between crop types. For instance:

- **Meadow** shows an extended temporal spread, lasting from early spring to late autumn, indicating a nearly continuous presence in the dataset.
- **Winter crops** (wheat, barley, triticale) have shorter, well-defined growing periods clustered around early summer.
- **Corn** exhibits a delayed phenological profile, extending well into the fall, reflecting its late-season nature.
- **Summer barley** is active around late spring to early summer, offering temporal separation from winter cereals.

These distributions are critical for time-series classification, as crop differentiation often relies on detecting such phenological signatures. However, overlapping periods, such as between clover, meadow, and summer barley, can introduce confusion, particularly during early growth stages when reflectance patterns are less distinctive.

1.3. Prediction Timing Analysis: Correct vs Incorrect

To better understand how classification performance varies across the season, we analyzed the distribution of the days on which predictions were made, separating correct and incorrect cases. The first plot shows violin distributions for correct (blue) and incorrect (red) predictions across the day of year. The second presents a corresponding boxplot visualization.

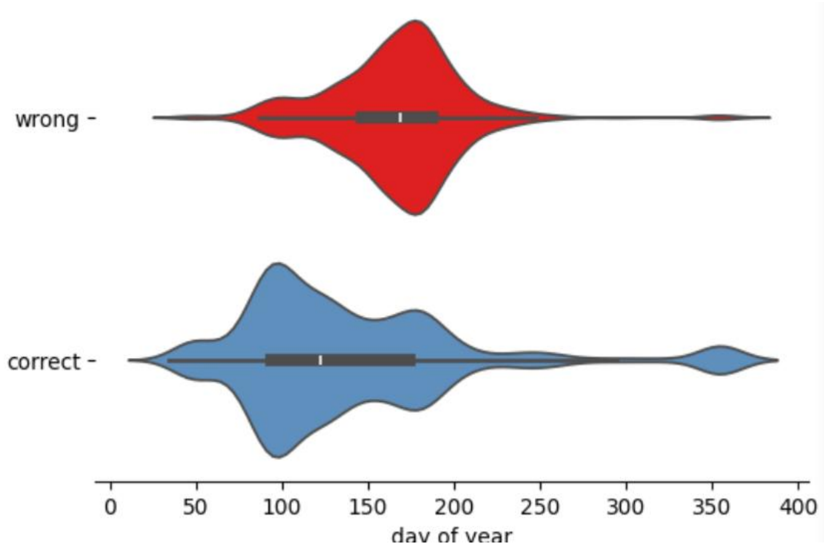


Figure 20: violin distributions for correct and incorrect predictions across the day of year

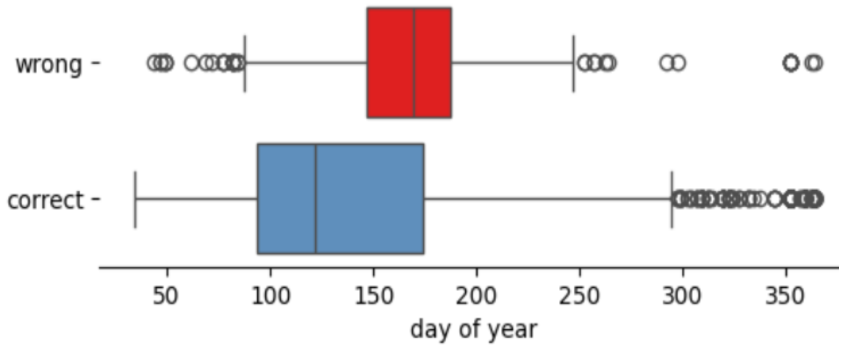


Figure 19:Boxplot of prediction timing for correct (blue) and incorrect (red) classifications.

Figure 21: Boxplot visualization for distributions for correct and incorrect predictions across the day of year

The violin and boxplot analyses reveal that **correct predictions** are broadly distributed across the growing season, with a peak during **mid-season** (*day 100–180*), highlighting that the model performs best when it has access to **sufficient temporal information**. In contrast, **incorrect predictions** are tightly clustered around *day 140–180*, indicating that most errors occur during **phenological overlaps** when crops exhibit **similar spectral behavior**. The **wider interquartile range** for correct predictions reflects the model's **robustness across varied growth stages**, while the **narrow window** of errors emphasizes the difficulty in distinguishing crops during **spectrally ambiguous periods**.

Table 6: Distribution of Predictions by Season (Days of Year)

Prediction Type	Temporal Range	Peak Concentration	Key Insight
Correct Predictions	*Day 50 – Day 300*	*Day 100 – Day 180*	**Highest accuracy occurs mid-season**, when the model benefits from richer temporal context and clearer phenological signals.
Incorrect Predictions	*Day 140 – Day 180*	*Day 140 – Day 180*	**Most errors occur during phenological overlaps**, where crops exhibit **similar spectral-temporal profiles**, making classification more challenging.

1.4.Stopping Day of Year Distribution:

This histogram shows when the model decided to stop observing and classify each crop sample

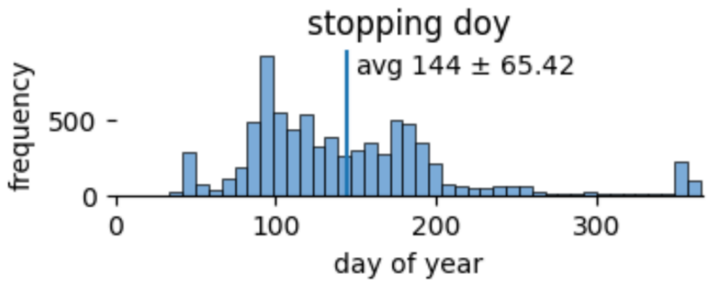


Figure 22: Histogram of EarlyRNN stopping days across all predictions

The average stopping day was 144 ± 65.42, aligning with mid-season. The broad spread indicates that while some confident predictions were made early, the model often delayed its decision until temporal patterns became clearer. This adaptive behavior aligns with the **EarlyRNN's** design and supports its use in applications requiring early, but not premature classification.

To complement the temporal analysis of correct and incorrect predictions, we present a **pixel-level accuracy visualization** in Figure 22. This map overlays model predictions against ground truth across the spatial extent of the dataset, with green indicating correctly classified pixels and red representing errors. While this figure reveals the overall spatial distribution of model performance, it also highlights areas where classification may degrade—especially along field boundaries or in spectrally ambiguous zones. The overall pixel-level accuracy observed in this spatial representation is 75.4%.

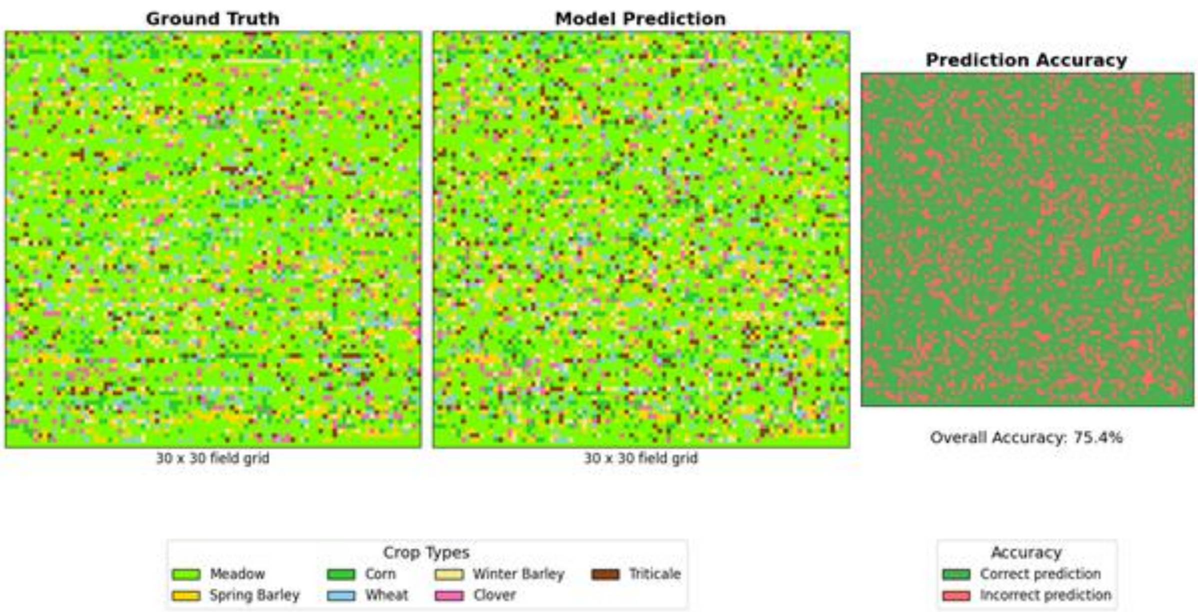


Figure 23:Model prediction vs. ground truth and corresponding pixel-level accuracy. Green = correct prediction, red = incorrect. Overall accuracy: 75.4%.

1.5.Spatial Consistency of Predictions&

This section will visually assess **how well the model's predictions align with actual crop distributions across space**, complementing your pixel-level accuracy and stopping day analysis.

To assess the spatial consistency of the model's predictions, we provide a side-by-side comparison of the predicted and ground truth crop type maps in Figure X. This visualization helps evaluate how accurately the model captures the spatial distribution of different crops across the study region, beyond sample-level metrics. It also highlights field-level consistency and reveals areas of systematic confusion or misclassification.

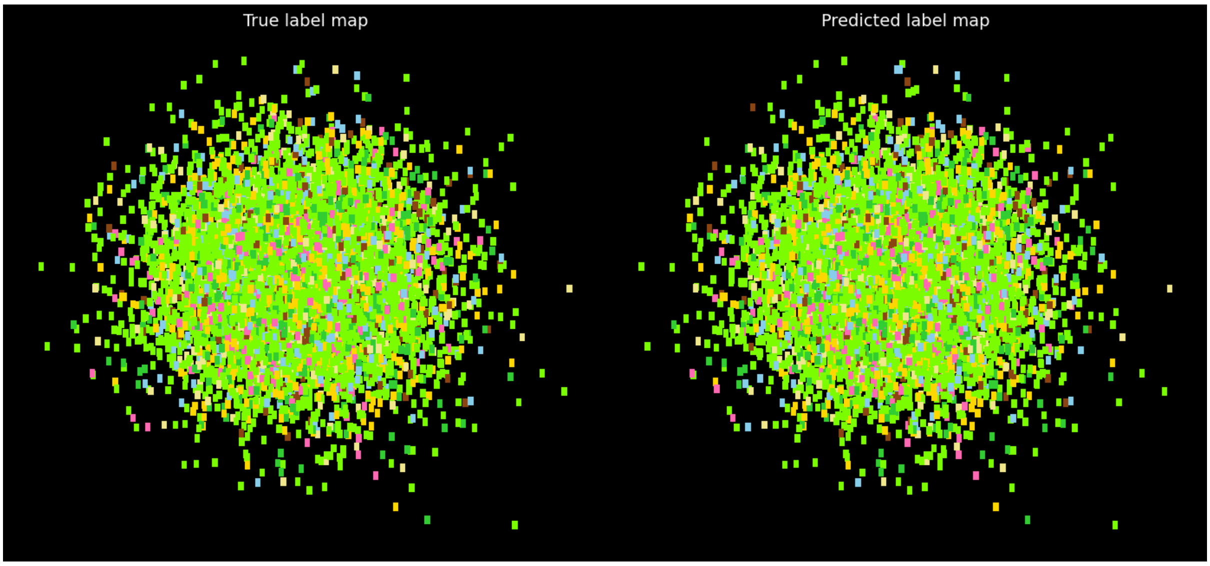


Figure 24:Side-by-side comparison of ground truth (left) and predicted (right) crop types across the study area.

Additional visualizations illustrating the spatial distribution of crop types and cartographic comparisons are provided in Appendix A.

Transfer Evaluation on BreizhCrops

1.1. Motivation

To assess the robustness and adaptability of the **EarlyRNN** model architecture, we extend our evaluation to the BreizhCrops dataset. Unlike BavarianCrops, this dataset captures agricultural activity in Brittany, France—a region with distinct climate patterns, crop types, and phenological cycles. By training and evaluating the same model architecture on this new dataset, we aim to understand how it generalizes to regions with different spatiotemporal and spectral dynamics, highlighting its capacity for cross-regional deployment.

1.2. Dataset Overview

BreizhCrops is a multivariate satellite time series dataset constructed from Sentinel-2 observations. Each crop field is represented as a sequence of **150-time steps** (fixed length), where each time step includes 13 spectral bands. The dataset spans 9 crop classes with highly imbalanced distributions. Temporary and permanent meadows dominate, while classes like Sunflower and Nuts are severely underrepresented.

- **Number of classes:** 9
- **Training samples:** 122,614
- **Test samples:** 30,654
- **Input shape:** (150, 13)
- **Label range:** 0–8 (e.g., Barley, Wheat, Rapeseed, Corn, Sunflower, etc.)
- **Preprocessing:** All sequences are padded or truncated to fixed-length (150), which simplifies input handling but may limit the model's expressiveness for long or short crop cycles.

$0 \rightarrow$ `Barley`
$1 \rightarrow$ `Wheat`
$2 \rightarrow$ `Rapeseed`
$3 \rightarrow$ `Corn`
$4 \rightarrow$ `Sunflower`
$5 \rightarrow$ `Orchards`
$6 \rightarrow$ `Nuts`
$7 \rightarrow$ `Permanent Meadows`
$8 \rightarrow$ `Temporary Meadows`

Figure 25:BreizhCrops dataset classes

Crop label mappings and class counts are shown below:

Table 7: Crop label mappings and sample distribution for BreizhCrops

Crop Type	Label	Train Samples	Test Samples
Barley	0	896700	224175
Wheat	1	2548950	637237
Rapeseed	2	485400	121350
Corn	3	4699950	1174987
Sunflower	4	300	75
Orchards	5	82950	20738
Nuts	6	1650	412
Permanent Meadows	7	3917550	979387
Temporary Meadows	8	5758650	1439662

The following figure is showcasing the distribution of the 9 classes in the BreizhCrops dataset :

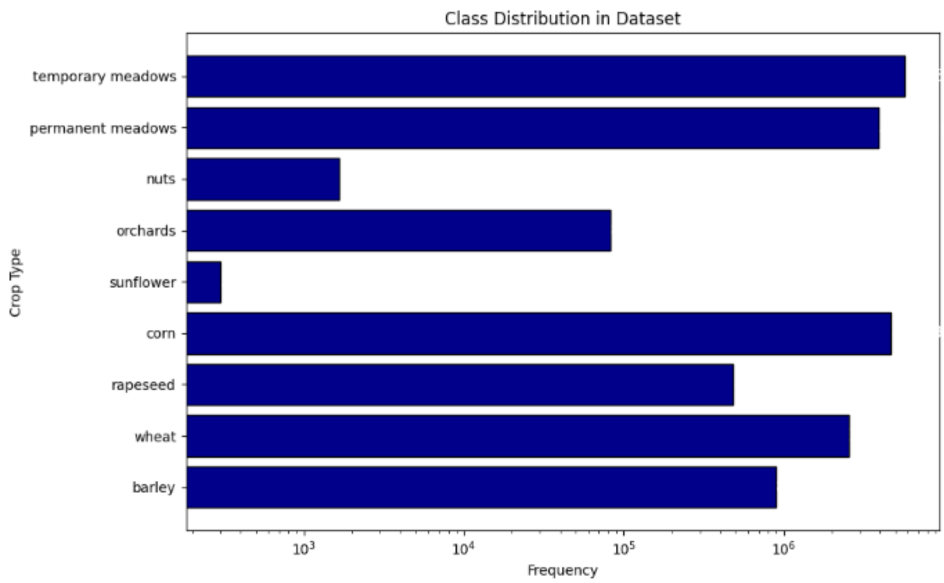


Figure 26: Comparison of class distribution between training and test sets

Comparison of BavarianCrops and BreizhCrops Datasets

This table highlights the key differences between the BavarianCrops and BreizhCrops datasets, which will help in understanding their characteristics and suitability for different crop classification tasks.

Table 8: Comparison Table for BavarianCrops and BreizhCrops Datasets

Feature	**BavarianCrops**	**BreizhCrops**
Region/Area	Bavaria, Germany	Brittany, France
Time Series Data	Temporal patterns from Sentinel-2 spectral bands	Temporal patterns from Sentinel-2 spectral bands
Dataset Size	16,600 samples (varies based on partition)	122,614 samples (across multiple regions)
Number of Features	13 Sentinel-2 bands	13 Sentinel-2 bands
Sequence Length	70 (fixed)	150 (fixed)
Data Format	NumPy arrays (X.npy, y.npy, etc.)	NumPy arrays (X.npy, y.npy, etc.)
Class Distribution	7 crop types	9 crop types
Crop Types (Classes)	7 crop types: Meadow, Barley, Corn, etc.	9 crop types: Barley, Wheat, Corn, Sunflower, etc.
Data Partitioning	Training, Validation, Testing	Training, Validation, Testing
Preprocessing	Normalize reflectance values	Normalize reflectance values
Class Imbalance	Moderate to high imbalance	High imbalance (e.g., "sunflower" is underrepresented)
Time Series Length Variability	Fixed at 70 time steps per sample	Fixed at 150-time steps per sample
File Structure	CSV, HDF5 files	CSV, HDF5 files

Feature	**BavarianCrops**	**BreizhCrops**
Data Access	Direct access from the web or local storage	Direct access from the web or local storage
Metadata	Contains metadata for each sample	Contains metadata for each sample, including geospatial info
Geospatial Information	Includes parcel-basedc geospatial info	Includes parcel-based geospatial info
Transformations Available	Custom transformations available	Custom transformations available
Data Storage Format	CSV, HDF5, and other auxiliary formats	CSV, HDF5, and auxiliary formats
Preloading to RAM	Option to preload to RAM for faster access	Option to preload to RAM for faster access
Shapefiles and Geospatial Data	No direct shapefile usage	Includes shapefiles for regions and crops

Both **BavarianCrops** and **BreizhCrops** offer valuable time-series datasets for crop classification. The **BreizhCrops** dataset has a larger number of crop types and more extensive geospatial information, while **BavarianCrops** focuses on fewer classes and smaller time-series lengths, which might be more suitable for quicker experiments

Experimental Setup

We use the **EarlyRNN** model, previously applied to BavarianCrops, and retrain it from scratch on BreizhCrops. The architecture remains unchanged, allowing us to evaluate whether its inductive biases are suitable for a different region. We use the same training and evaluation pipeline, maintaining consistency across datasets.

Performance Results

In this section, we present a comprehensive evaluation of the EarlyRNN model's performance on the BreizhCrops dataset. The evaluation encompasses several key metrics and visual diagnostics, focusing on prediction accuracy, class-wise behavior, and the dynamics of model decision-making during the growing season.

1.1. Confusion Matrix

The confusion matrix provides a normalized summary of prediction outcomes across the nine crop classes. Correct classifications are represented by the diagonal elements (in blue), while misclassifications are indicated by off-diagonal elements.

Key Observations:

- High Recall for Dominant Classes: Crops such as temporary and permanent meadows exhibit high recall, reflecting the model's ability to successfully capture their distinct spectral-temporal patterns.
- Low Recall for Minority Classes: Crops like sunflower, orchards, and nuts show low recall, which can be attributed to class imbalance and potential spectral confusion with other more dominant crops.
- Overall Accuracy: The model achieves an overall accuracy of 80%. However, this metric is influenced by the dominance of meadow classes, which skew the results and mask the performance on less frequent crops.

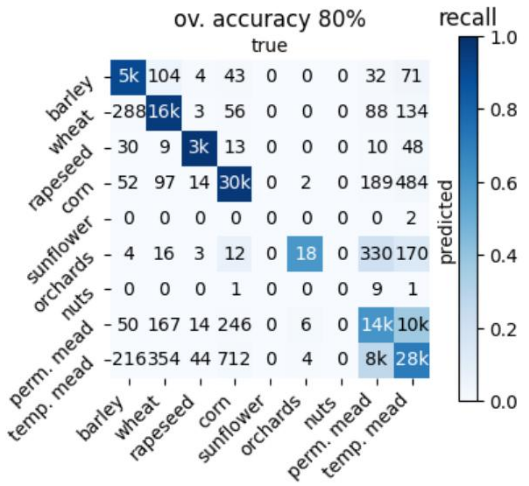


Figure 27: Confusion Matrix of EarlyRNN on BreizhCrops dataset

1.2. Stopping Day Distribution

This histogram visualizes the distribution of model stopping times, expressed in terms of the day-of-year. The mean stopping day is approximately day 159 ± 28.3, which corresponds to early June.

Insights:

- The concentration of stopping times around day 159 indicates that the model relies on distinct seasonal cues to make classifications, particularly in regions where spectral evidence becomes clear in early summer.
- The sharpness of the distribution suggests that the model typically makes decisions based on well-defined seasonal transitions.

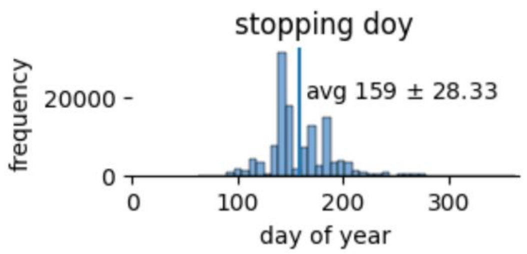


Figure 28:Stopping Day Distribution on BreizhCrops datasets

1.3. Correct vs Incorrect – Boxplot

The boxplot compares the stopping times for correctly and incorrectly classified samples.

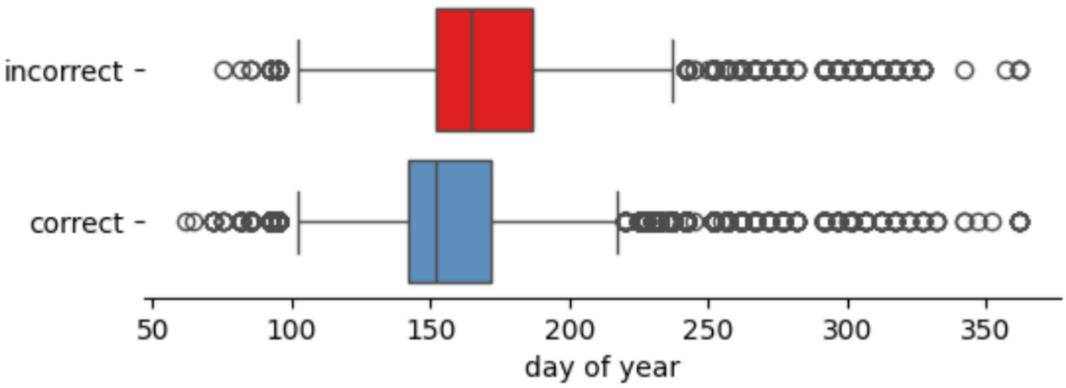


Figure 29: correct vs Incorrect classification Boxplot

Interpretation:

- Correct Predictions: These are more tightly clustered around the mean stopping day, with fewer outliers, indicating that the model makes confident decisions within a consistent time window.
- Incorrect Predictions: These exhibit greater variability, with a broader spread across early and late time points. This variability suggests that the model's accuracy is compromised by both premature and delayed decisions.

1.4. Correct vs Incorrect – Violin Plot

The violin plot provides a smoothed distribution of stopping times, complementing the insights from the boxplot.

Key Insights:

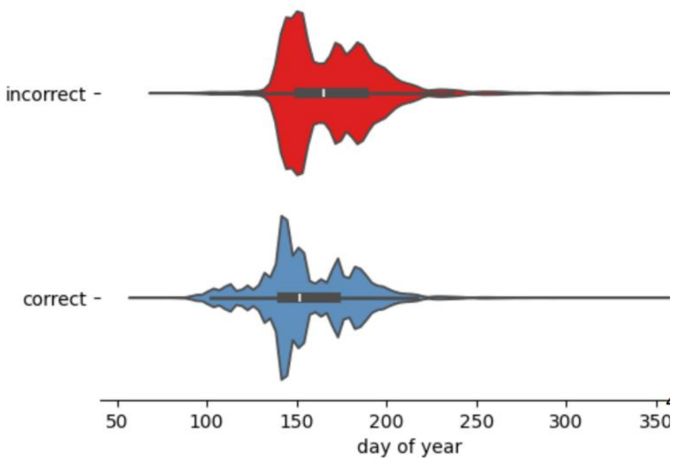


Figure 30: correct vs incorrect classification violin plot

- Incorrect Predictions: The red distribution shows a peak around days 140-160, with a longer tail. This suggests that some misclassifications occur due to early decisions when the model has not yet gathered enough spectral evidence, leading to less accurate predictions.
- Correct Predictions: The blue distribution is more symmetrical, centered slightly later, which indicates that slightly delayed decisions improve prediction accuracy and reduce errors.

Comparative Analysis

This section compares the EarlyRNN model's performance on the BavarianCrops and BreizhCrops datasets to assess the model's robustness and early prediction reliability across different regions.

1.1. Cross-Dataset Performance Trends

A side-by-side comparison of per-class F1-scores highlights key differences in model performance between the two datasets.

- High-Consistency Classes: Crops like temporary and permanent meadows exhibit strong F1-scores across both datasets, suggesting that these crops share consistent seasonal and spectral patterns, enabling reliable cross-region generalization.
- Low-Transferability Classes: Crops like sunflower, nuts, and orchards show significant drops in F1-score on BreizhCrops, reflecting issues with class imbalance, differences in phenology, and the absence of these crops in the training data from BavarianCrops. This discrepancy points to limitations in generalization when training data is insufficient for rare classes.

Visualization: A grouped bar chart or line plot can be used to visually summarize the F1-scores across both datasets, emphasizing the generalization gap and highlighting the challenges in transferring learned features to new regions.

1.2. Early Prediction Reliability

We also analyze the model's ability to make confident predictions at an early stage of the growing season in both datasets.

Key Insights:

- In **BavarianCrops:** The model tends to stop earlier (around day 140-150), reflecting high confidence in the spectral patterns that it has been exposed to during training.
- In **BreizhCrops:** The average stopping day shifts later to around day 159 ± 28.3, indicating that the model requires more temporal information to confidently classify crops, particularly those that are less familiar or more complex.
- **Increased Variability:** The variance in stopping times is greater for BreizhCrops, suggesting that the model experiences more uncertainty during inference, which affects both the timeliness and accuracy of predictions.

Conclusion

The results showed that EarlyRNN performs well for early crop classification on BavarianCrops, with decent generalization to BreizhCrops. Strong accuracy was observed on dominant classes, but weaknesses appeared on rare ones. Visual and quantitative analyses confirmed both strengths and limits of the model.

Chapter 5: Discussion

Introduction

This chapter provides a critical interpretation of the experimental results obtained from evaluating the **EarlyRNN model** on the BavarianCrops and BreizhCrops datasets. Beyond reporting metrics, we examine how the model behaves under different spatiotemporal conditions, its capacity for early prediction, and the extent to which it generalizes across regions. Through this lens, we identify strengths, limitations, and practical implications, while outlining methodological insights and directions for improvement. The discussion bridges the gap between quantitative outcomes and their broader relevance for real-world agricultural monitoring systems.

Interpretation of Key Findings

The experimental results across both **BavarianCrops** and **BreizhCrops** datasets offer valuable insights into the strengths and limitations of the EarlyRNN model in the context of time-series crop classification. On the BavarianCrops dataset, the model achieved an overall classification accuracy of **86%,** with strong performance on well-represented classes like Meadow and Corn. The confusion matrix, per-class accuracy visualization, and stopping day analyses all indicate that the model can confidently classify dominant crops with **high temporal and spectral distinctiveness.** Similarly, the stopping time histogram suggests that EarlyRNN often makes **confident decisions by mid-season (around day 144),** balancing early prediction with sufficient temporal information.

However, this performance did not translate equally when applied to the **BreizhCrops** dataset. Although the model achieved **80% overall accuracy**, this figure is skewed by the dominance of permanent and temporary meadows. The performance on minority classes like Sunflower, Nuts, and Orchards **dropped significantly**, both in recall and F1-score. The stopping day distribution in **BreizhCrops** also shifted later (average day 159 ± 28.3), indicating greater uncertainty and delayed decision-making in a region with unfamiliar phenological patterns and crop compositions. This gap in generalization reveals critical limitations in the current approach and highlights the sensitivity of early classification to prior exposure.

Cross-Regional Generalization

While a common strategy for assessing generalization is to apply a pre-trained model across domains, in this study we opted not to transfer learning from BavarianCrops to BreizhCrops. Instead, we trained a new EarlyRNN model from scratch on the BreizhCrops dataset using the same architecture and training procedure. This approach allows us to **isolate the impact of regional dataset characteristics** on model performance, without introducing confounding effects from source-domain pretraining. The results therefore reflect how well the EarlyRNN inductive biases generalize **architecturally**, not via learned representations.

The grouped **F1-score** comparison chart *(Figure C.1 in appendix C— Normalized Confusion Matrices)* across datasets underscores this discrepancy. While crops like Meadow retained stable performance across both domains, crops absent or underrepresented in the BavarianCrops training set experienced a severe drop in classification quality. These results reveal that inductive biases captured by the model during training are not sufficiently general to accommodate significant domain variability. Without fine-tuning or adaptation, EarlyRNN struggles to extend its learned temporal representations to crops it has never encountered in similar form.

Reliability of Early Prediction

The concept of early classification is central to this thesis. It aims to predict crop types as early as possible in the growing season to support timely agricultural decision-making. On **BavarianCrops**, EarlyRNN successfully achieved this, with many predictions occurring between days **120 and 160** and exhibiting high confidence, as reflected in narrow stopping day distributions and robust classification outcomes.

However, in **BreizhCrops**, the model exhibited a broader spread and delayed stopping behavior, particularly on unfamiliar crops. The violin and boxplots *(Results & analysis Chapter)* confirm that incorrect predictions tended to cluster earlier or later than the optimal classification window, indicating uncertainty in the spectral evidence during those periods. The model appears to wait longer in the season when faced with ambiguous signals, yet this delay does not always guarantee improved accuracy. These findings suggest a potential trade-off between earliness and reliability in out-of-domain contexts.

Moreover, the discrepancy between sample-level accuracy (86%) and pixel-level spatial accuracy (75.4%) as visualized in the prediction map overlays—reinforces the importance of evaluating models at multiple levels of granularity. Spatial maps introduce additional sources of error, such as field boundaries, mixed pixels, and label misalignment, which are not accounted for in confusion matrices alone. A multi-resolution evaluation strategy is essential for understanding practical deployment outcomes.

Implications for Real-World Applications

The ability to classify crops early and accurately has substantial implications for agricultural monitoring, crop rotation planning, and subsidy enforcement. Models like **EarlyRNN**, when properly trained and calibrated, offer strong potential to support near real-time crop mapping initiatives. However, our findings indicate that **geographic and phenological variability** between regions limits the direct transferability of models without adaptation.

For policy and operational systems relying on consistent predictions across heterogeneous landscapes, integrating **local training data** or using **domain adaptation** methods becomes essential. Moreover, the observed variability in **stopping times** and prediction reliability across crop types suggests that **uncertainty estimation** or **interpretability mechanisms** may be required to enable confident early decisions in real-world deployments.

This work is particularly relevant in the context of **OCP Nutricrops' CropID platform**, which aims to deliver **tailored fertilizer recommendations** through accurate, early-season crop classification. By incorporating multimodal remote sensing, crop phenology, and soil health indicators, such platforms can benefit from early classification models to:

- Enable variety-specific nutrient recommendations based on early phenological cues;
- Optimize fertilizer application timing, minimizing waste and environmental impact;
- Adapt recommendations to **regional variation** through hybrid modeling strategies.

Despite current limitations in pixel-level accuracy, the integration of predictive AI models like EarlyRNN into farmer-facing tools reflects an important **step toward sustainable, data-driven agriculture**, aligning with **CropID's mission** to bridge advanced analytics and real-world agronomic needs.

Limitations and Future Directions

This study faced several constraints, both computational and methodological. Training was conducted entirely on **CPU**, as **GPU** resources *(e.g., Google Colab Pro)* proved unreliable or inaccessible in practice. This limited the scale of hyperparameter tuning and experimentation with larger or more complex architectures.

Moreover, no predefined dataset was provided at the outset. A complete benchmark analysis **(see Appendix E)** was conducted to identify and justify the selection of BavarianCrops and BreizhCrops datasets as appropriate testbeds for the CropID proof of concept.

Moving forward, several directions can be pursued to enhance model performance and operational scalability:

- **Cross-Region Evaluation**: Extend experiments to diverse agro-ecological zones beyond Europe.
- **Domain Adaptation**: Fine-tune models on local datasets to improve transferability under domain shifts.
- **Multimodal Fusion**: Integrate additional data sources—soil, weather, terrain—to improve robustness and contextual relevance.
- **Architectural Exploration**: Investigate 3D CNNs, ConvLSTMs, and attention-based models for more expressive spatio-temporal learning.
- **Data Balancing**: Apply augmentation or synthetic sampling to address class imbalance and improve minority class recall.
- **Uncertainty-Aware Prediction**: Incorporate confidence thresholds to better control early prediction decisions.
- **Continual Learning**: Enable dynamic model adaptation to unseen crop types or evolving phenological patterns.

These directions align with the broader CropID vision of delivering a robust, adaptive, and farmer-centric system capable of precision crop classification and optimized fertilizer recommendations at scale.

Conclusion

In summary, the EarlyRNN model demonstrates strong potential for early-season crop classification within familiar data domains, particularly on dominant crop types with stable phenological patterns. However, its limitations in handling rare crops, adapting across regions, and maintaining reliability under domain shift reveal critical areas for refinement. These findings underscore the importance of incorporating adaptation mechanisms, balanced datasets, and multimodal information to build models capable of robust and scalable deployment. The next chapter will consolidate the contributions of this work and suggest concrete paths for future research and operational implementation.

General Conclusion

The **CropID** project represents a promising and forward-thinking initiative at the crossroads of **AI, remote sensing, and sustainable agriculture**. Its ambition to deliver precise, crop-specific fertilizer recommendations through multimodal data fusion addresses a pressing need in modern farming: increasing yield while minimizing environmental impact. The integration of satellite time-series, soil health metrics, and phenological insights into one unified decision-support system positions CropID as a high-potential tool for both farmers and agronomists.

Throughout this internship, significant progress was made in evaluating and implementing early-season crop classification strategies. Despite constraints such as limited computing resources and the need to benchmark datasets independently, the work laid a solid foundation for the system's modeling pipeline. The **EarlyRNN** architecture was assessed on two distinct European datasets, revealing important insights into the challenges of **cross-regional generalization**, class imbalance, and prediction reliability. These findings will inform future iterations of the platform, especially regarding domain adaptation and multimodal model integration.

Beyond the technical outcomes, the project was carried out within a **highly research-focused environment**, combining methodological rigor with real-world relevance. This alignment between innovation and application underscores CropID's potential impact—not only in precision agriculture but also as part of **OCP Nutricrops' broader strategy** to support sustainable, localized fertilizer use.

In conclusion, the internship contributed to both the **advancement of CropID's scientific foundation** and the broader goal of creating AI tools that are **trustworthy, adaptive, and grounded in agronomic realities**. The challenges encountered provided valuable direction for future research, and the work accomplished during this internship will serve as a stepping stone toward building more robust, context-aware crop intelligence systems.

On a personal level, this experience has deepened my conviction to pursue a research-oriented career in the fields of **sustainable agriculture and artificial intelligence**. Working on CropID has not only strengthened my technical and analytical skills but has also given me a clearer vision of the societal impact I hope to contribute to through my future studies and doctoral research. It reaffirmed my passion for interdisciplinary innovation and my desire to further explore the intersection of **technology, sustainability, and real-world agricultural challenges**.

References

1. Barriere, V., Claverie, M., Schneider, M., Lemoine, G., & d'Andrimont, R. (2024). Boosting crop classification by hierarchically fusing satellite, rotational, and contextual data. *Remote Sensing of Environment*, 305. https://doi.org/10.1016/j.rse.2024.114110
2. Ashraf, M., Chen, L., Innab, N., Umer, M., Baili, J., Kim, T.-H., & Ashraf, I. (2024). Novel 3-D Deep Neural Network Architecture for Crop Classification Using Remote Sensing-Based Hyperspectral Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17, 12649–12665. https://doi.org/10.1109/JSTARS.2024.3422078
3. Hu, Y., Hu, Q., & Li, J. (2025). CMINet: A Unified Cross-Modal Integration Framework for Crop Classification From Satellite Image Time Series. *IEEE Transactions on Geoscience and Remote Sensing*, 63. https://doi.org/10.1109/TGRS.2024.3522942
4. Kim, D.-W., Jang, G., & Kim, H.-J. (2025). Development of CNN-Based Semantic Segmentation Algorithm for Crop Classification of Korean Major Upland Crops Using NIA AI HUB. *IEEE Access*, 13, 8425–8438. https://doi.org/10.1109/ACCESS.2025.3527502
5. Wang, Y., Feng, L., Sun, W., Wang, L., Yang, G., & Chen, B. (2024). A lightweight CNN-Transformer network for pixel-based crop mapping using time-series Sentinel-2 imagery. *Computers and Electronics in Agriculture*, 226. https://doi.org/10.1016/j.compag.2024.109370
6. Lei, L., Wang, X., Hu, X., Zhang, L., & Zhong, Y. (2024). PhenoCropNet: A Phenology-Aware-Based SAR Crop Mapping Network for Cloudy and Rainy Areas. *IEEE Transactions on Geoscience and Remote Sensing*, 62. https://doi.org/10.1109/TGRS.2024.3483110
7. Koukos, A., Jo, H.-W., Sitokonstantinou, V., Tsoumas, I., Kontoes, C., & Lee, W.-K. (2024). Towards Global Crop Maps with Transfer Learning. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 1540–1545. https://doi.org/10.1109/IGARSS53475.2024.10641793
8. Hoppe, H., Dietrich, P., Marzahn, P., Weiß, T., Nitzsche, C., Freiherr von Lukas, U., Wengerek, T., & Borg, E. (2024). Transferability of Machine Learning Models for Crop Classification in Remote Sensing Imagery Using a New Test Methodology: A Study on Phenological, Temporal, and Spatial Influences. *Remote Sensing*, 16(9). https://doi.org/10.3390/rs16091493
9. Qin, X., Guo, H., Su, X., Zhao, Z., Wang, D., & Zhang, L. (2025). Spatiotemporal masked pre-training for advancing crop mapping on satellite image time series with limited labels. *International Journal of Applied Earth Observation and Geoinformation*, 137. https://doi.org/10.1016/j.jag.2025.104426
10. Yuan, H.-T., Huang, K.-K., Duan, J.-L., Lai, L.-Q., Yu, J.-X., Huang, C.-W., & Yang, Z. (2024). Generalized few-shot learning for crop hyperspectral image precise classification. *Computers and Electronics in Agriculture*, 227. https://doi.org/10.1016/j.compag.2024.109498
11. Račič, M., Oštir, K., Zupanc, A., & Čehovin Zajc, L. (2024). Multi-Year Time Series Transfer Learning: Application of Early Crop Classification. *Remote Sensing*, 16(2). https://doi.org/10.3390/rs16020270
12. Pham, V.-D., Tetteh, G., Thiel, F., Erasmi, S., Schwieder, M., Frantz, D., & van der Linden, S. (2024). Temporally transferable crop mapping with temporal encoding and

deep learning augmentations. *International Journal of Applied Earth Observation and Geoinformation*, 129. https://doi.org/10.1016/j.jag.2024.103867

13. Mirzaei, A., Bagheri, H., & Khosravi, I. (2023). Enhancing Crop Classification Accuracy through Synthetic SAR-Optical Data Generation Using Deep Learning. *ISPRS International Journal of Geo-Information*, 12(11). https://doi.org/10.3390/ijgi12110450
14. Ma, X., Li, L., & Wu, Y. (2025). Deep-Learning-Based Method for the Identification of Typical Crops Using Dual-Polarimetric Synthetic Aperture Radar and High-Resolution Optical Images. *Remote Sensing*, 17(1). https://doi.org/10.3390/rs17010148
15. Li, H. Y., Lawrence, J. A., Mason, P. J., & Ghail, R. C. (2023). UNSUPERVISED WINTER WHEAT MAPPING BASED ON MULTI-SPECTRAL AND SYNTHETIC APERTURE RADAR OBSERVATIONS. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 48(1/W2-2023), 1411–1416. https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-1411-2023
16. Zhai, X., Han, W., Li, X., & Huang, S. (2024). PLGCA: A Progressive Local-Global Context-Aware Semantic Segmentation Network for Crop Remote Sensing Mapping. *2024 6th International Conference on Electronics and Communication, Network and Computer Technology (ECNCT)*, 491–495. https://doi.org/10.1109/ECNCT63103.2024.10704428
17. Zhang, W., Tang, P., Meng, Y., Zhao, L., Zhao, Z., & Zhang, Z. (2024). Crop type classification of remote sensing image time series based on multi-scale spatial-temporal global attention model. *National Remote Sensing Bulletin*, 28(11), 2865–2877. https://doi.org/10.11834/jrs.20243557

Supplementary Resources:
Further references and supportive material reviewed during this research are available through the project's dedicated website: https://crop-ai-research.vercel.app/

Appendices

Appendix A — Spatial and Cartographic Visualizations of BavarianCrops

This appendix provides supplementary visualizations that illustrate the spatial distribution of crop types, model prediction layout, and cartographic variations used in this study. These figures are intended to support the interpretation of results discussed in Chapter 4.

- *Figure A.1 — Simulated Crop Classification Raster View*

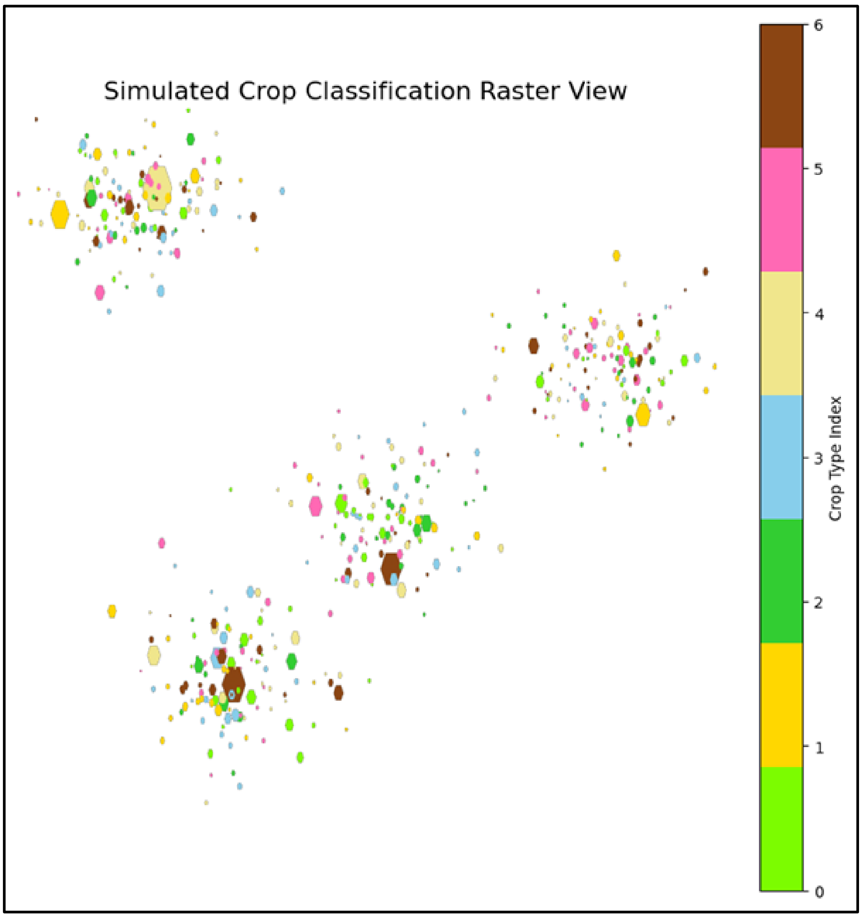


Figure 31: Simulated crop classification raster view. Each color-coded symbol represents a predicted crop type. This visualization illustrates spatial clustering and segmentation patterns learned by the model.

- *Figure A.2 — Ground Truth Crop Distribution Map*

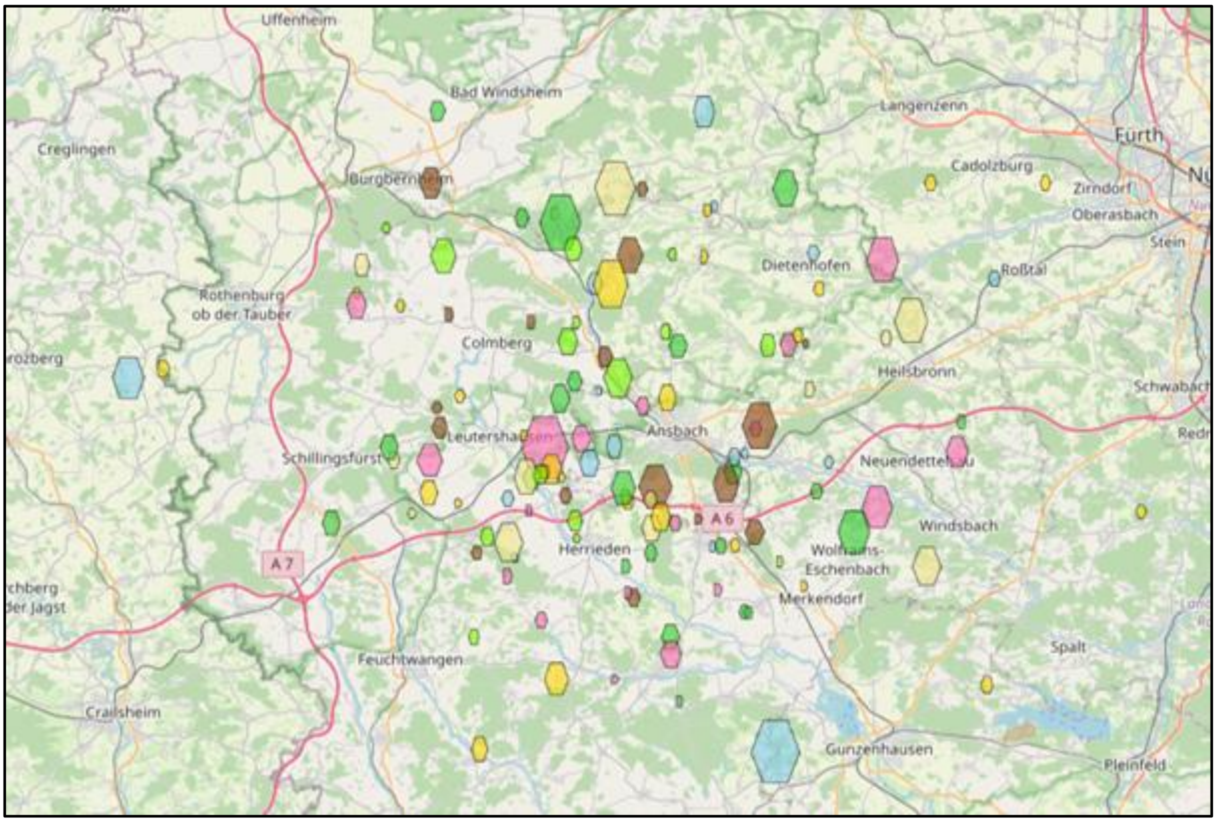


Figure 32: Geographical distribution of ground truth crop parcels over the study region. This map offers spatial context to the dataset used for training and evaluation.

◇ *Figure A.3 — Comparison of Basemaps for Ground Truth Visualization*

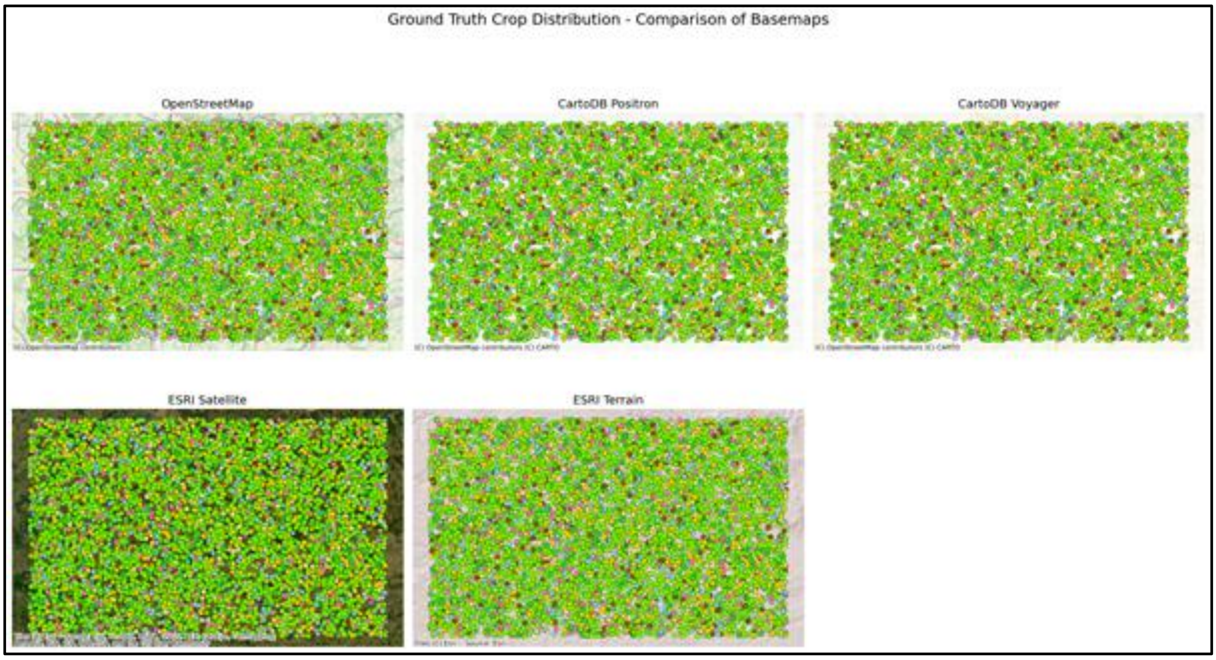


Figure 33: Ground truth crop distribution shown over five different basemap styles (OpenStreetMap, CartoDB Positron, CartoDB Voyager, ESRI Satellite, and ESRI Terrain). This comparison highlights how different geographic tiles impact map readability

Appendix B — Crop-Specific Spatial Visualizations

This appendix provides a set of interactive HTML-based spatial maps, one for each crop class in the dataset. These visualizations allow the user to explore the geographic distribution and prediction performance of each crop type individually. The maps offer a practical complement to the aggregated results discussed in Chapter 4.

B.1 All Crops Combined

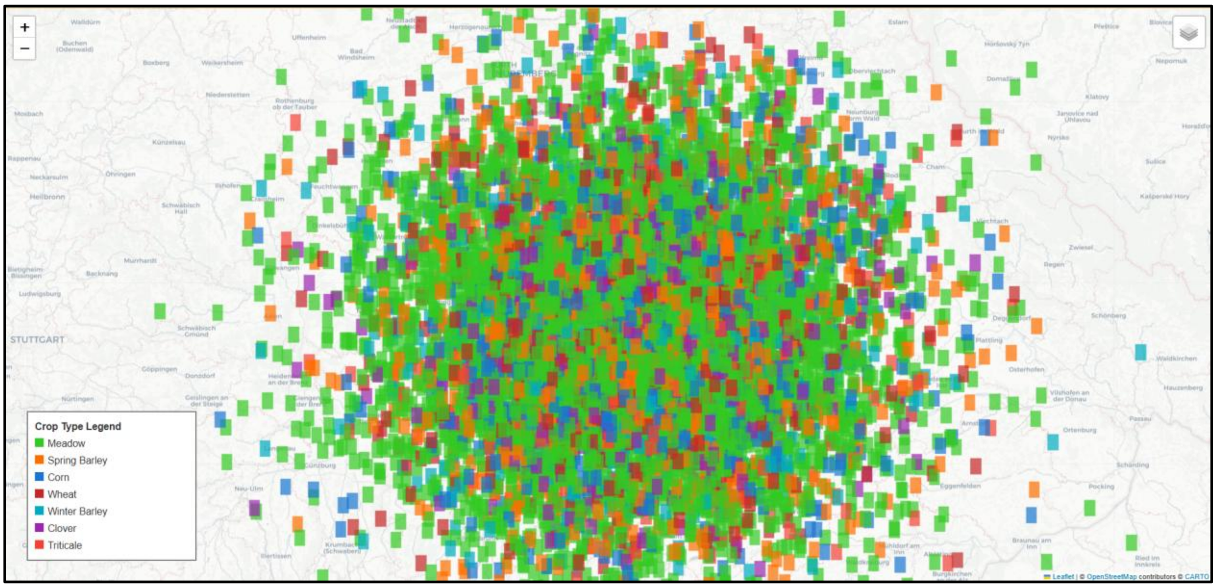


Figure 34: Interactive map showing the full spatial distribution of all crop types in the dataset.

B.2 Meadow

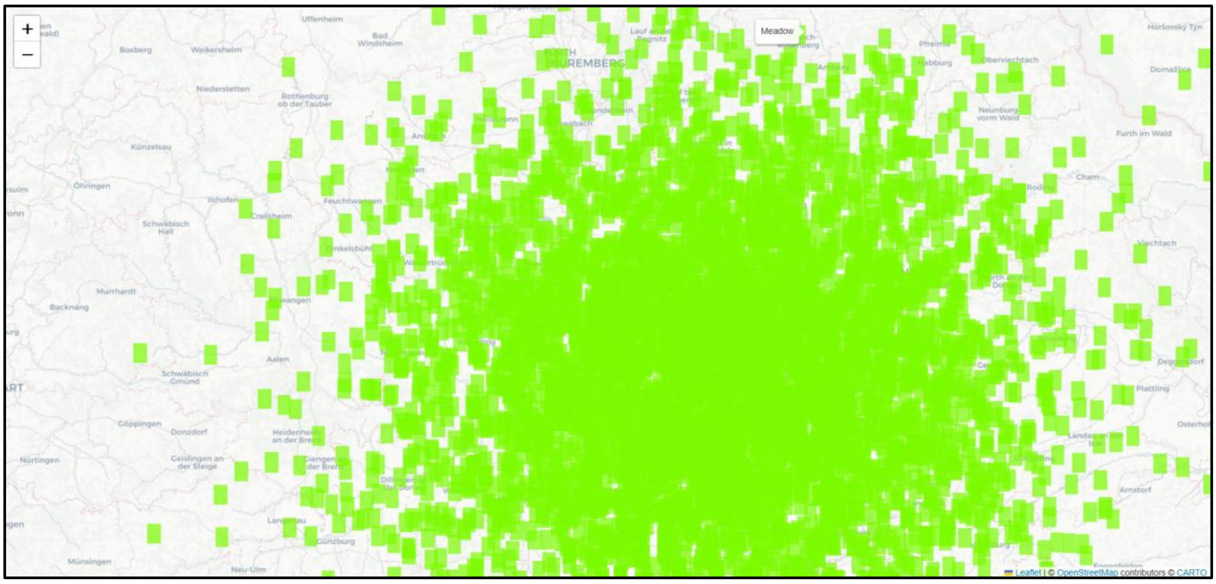


Figure 35: Spatial representation of prediction results and ground truth for the "Meadow" class.

B.3 Spring Barley

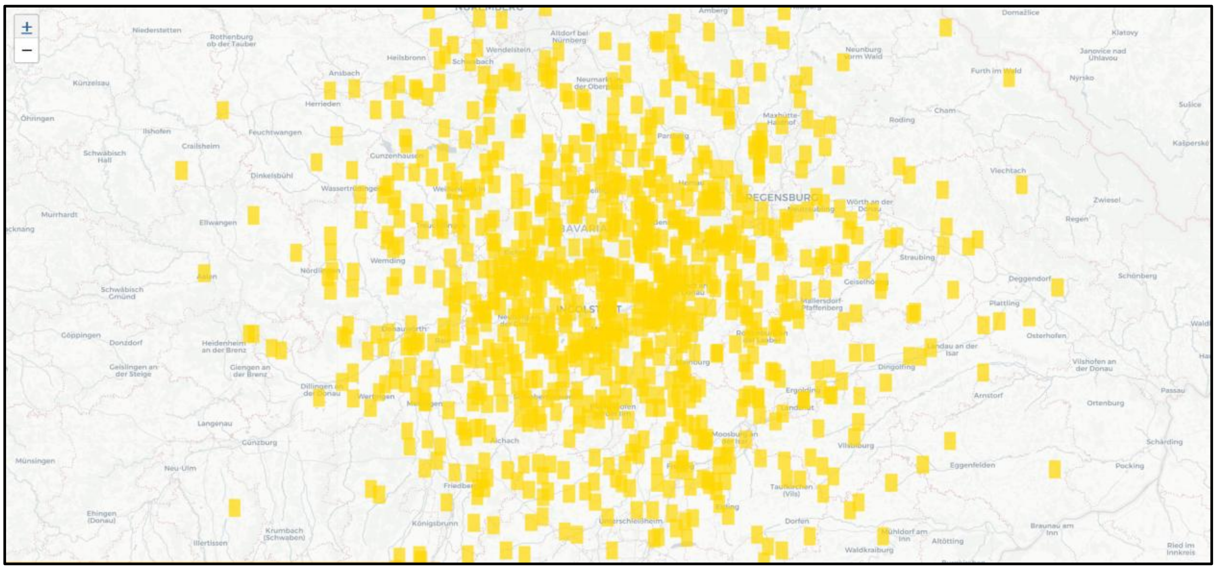


Figure 36: Map focused on the distribution and prediction of Spring Barley fields.

B.4 Winter Barley

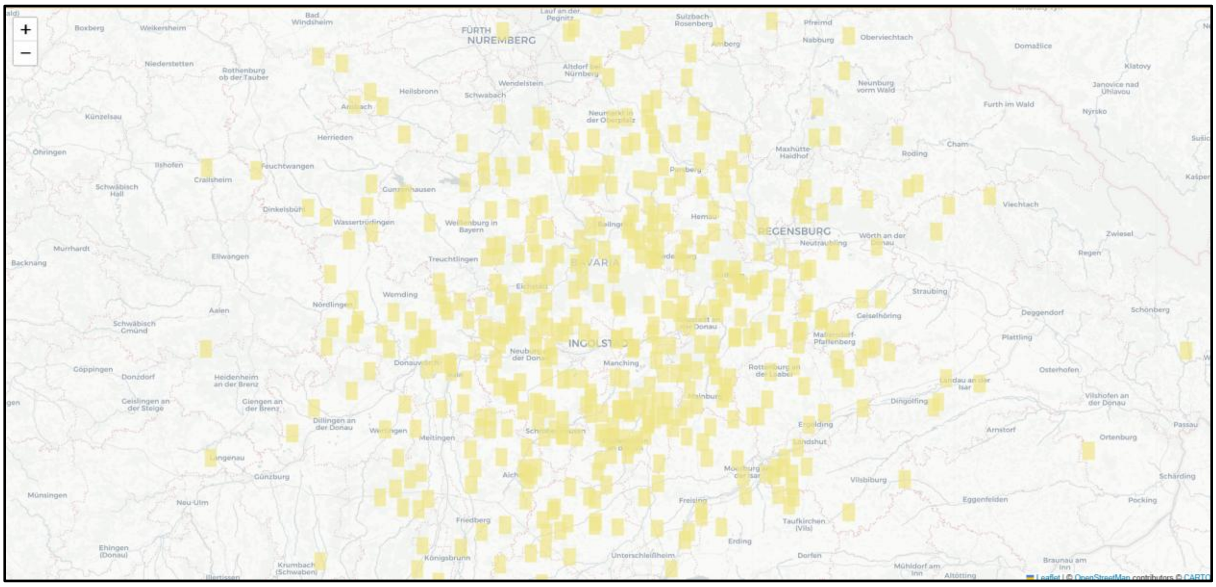


Figure 37: Map showing spatial prediction outcomes for Winter Barley

B.5 Wheat

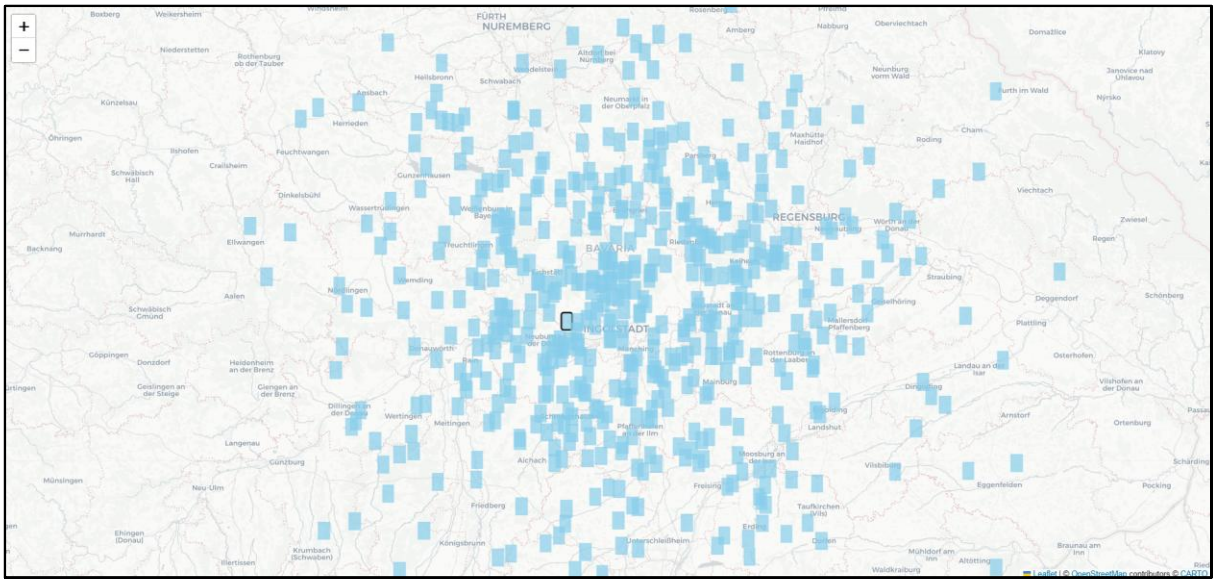


Figure 38: Interactive visualization for the Wheat class, highlighting spatial accuracy and misclassification zones.

B.6 Corn

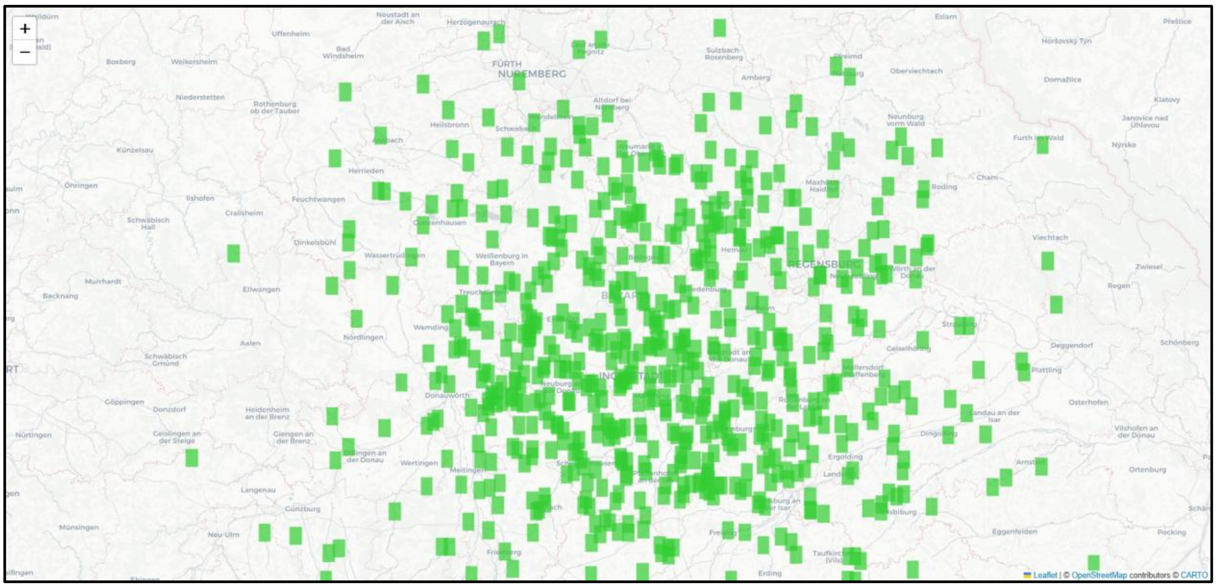


Figure 39: Map of Corn crop predictions, reflecting both correct and incorrect classifications.

B.7 Clover

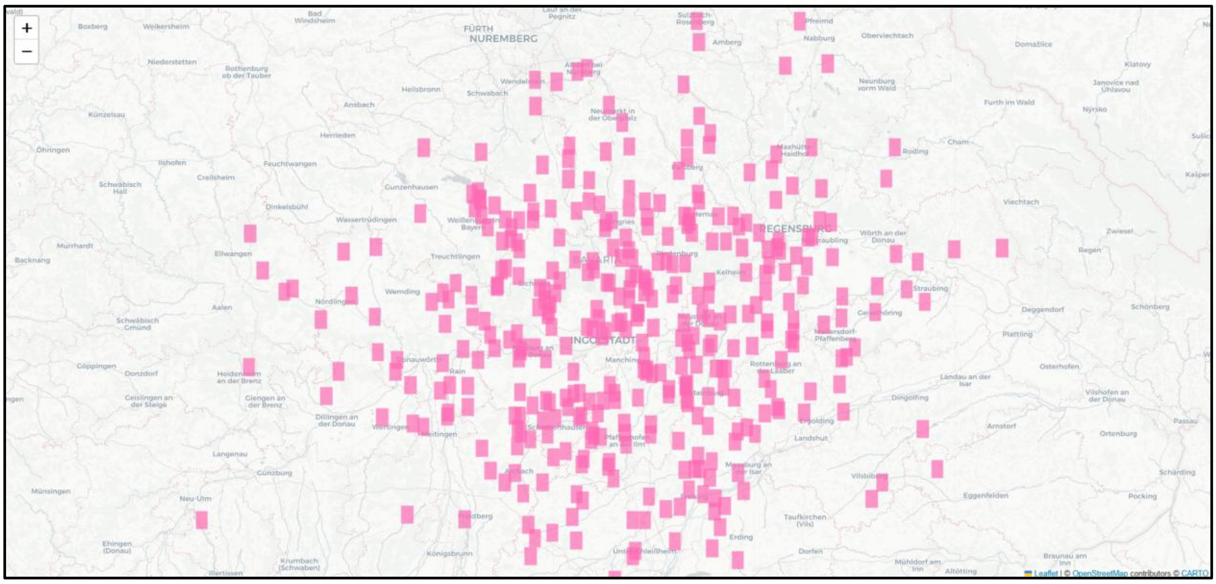


Figure 40: Spatial layout of clover predictions across the study area

- *B. 8 Triticale*

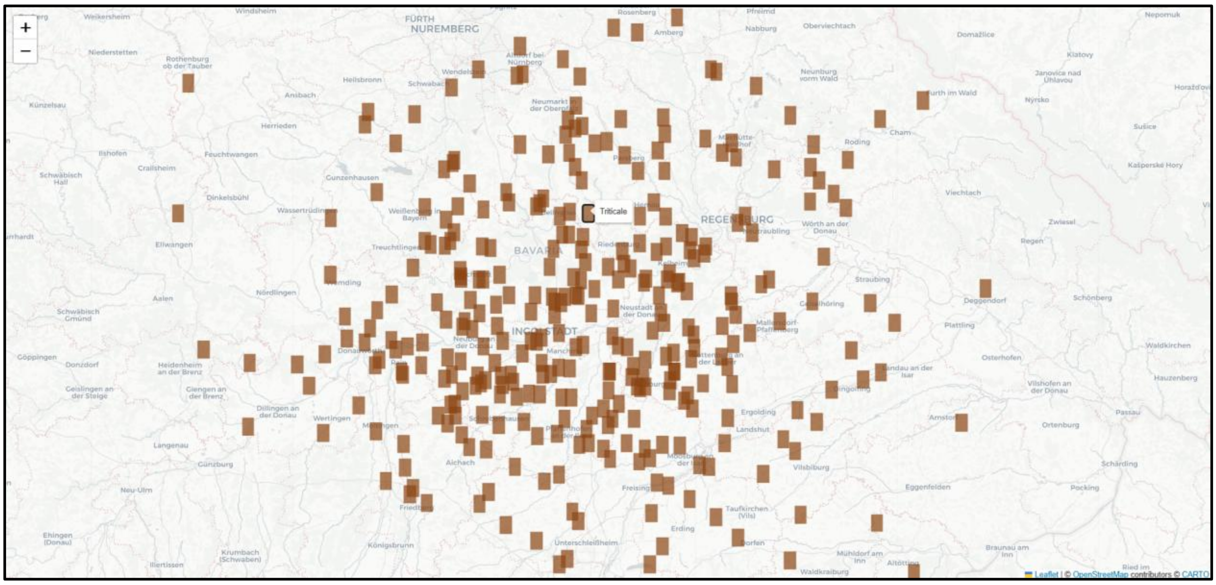


Figure 41: Spatial layout of Triticale predictions across the study area

Appendix C — Confusion Matrices for BavarianCrops and BreizhCrops

This appendix provides a visual comparison of the normalized confusion matrices obtained from the EarlyRNN model on the BavarianCrops and BreizhCrops datasets. These matrices offer insight into class-wise prediction behavior, recall distribution, and common misclassification patterns.

◇ *Figure C.1 — Normalized Confusion Matrices*

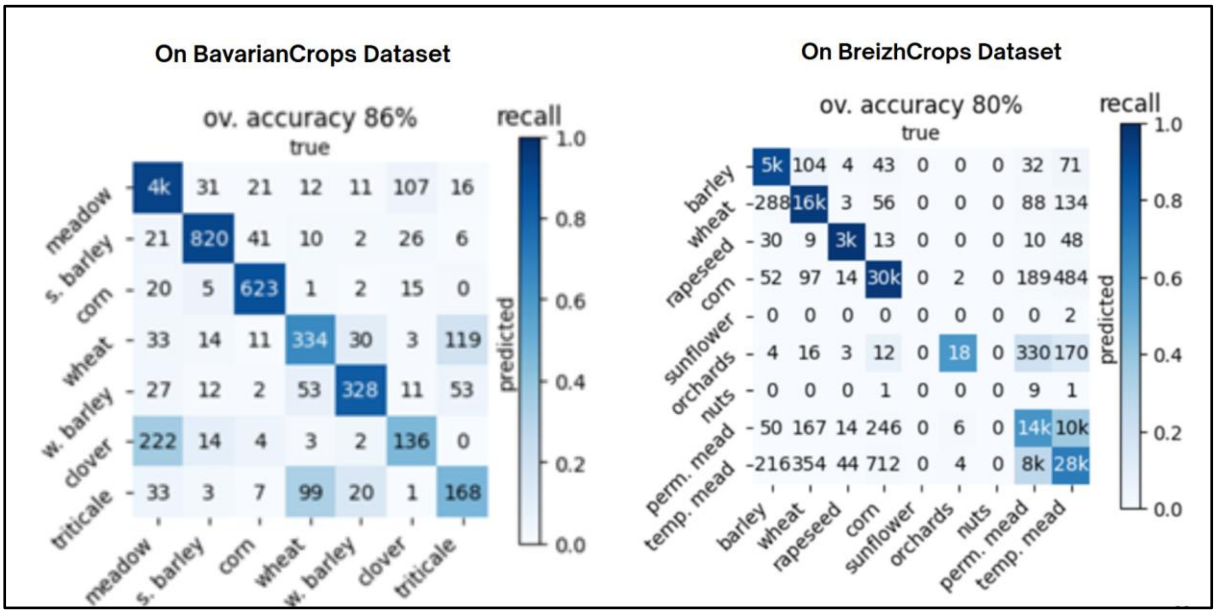


Figure 42: Normalized confusion matrices for EarlyRNN on the BavarianCrops dataset (left) and the BreizhCrops dataset (right). Each cell represents the proportion of samples from a true class (rows) that were predicted as a given class (columns).

Appendix D: Visual Evaluation of BreizhCrops Predictions

This appendix presents a series of visualizations that complement the quantitative performance analysis of the EarlyRNN model applied to the **BreizhCrops** dataset. These figures provide spatial and class-specific insights into the prediction accuracy, distribution of errors, and model behavior across different crop types.

◇ *Figure D.1 – Grid-Based Classification Results*

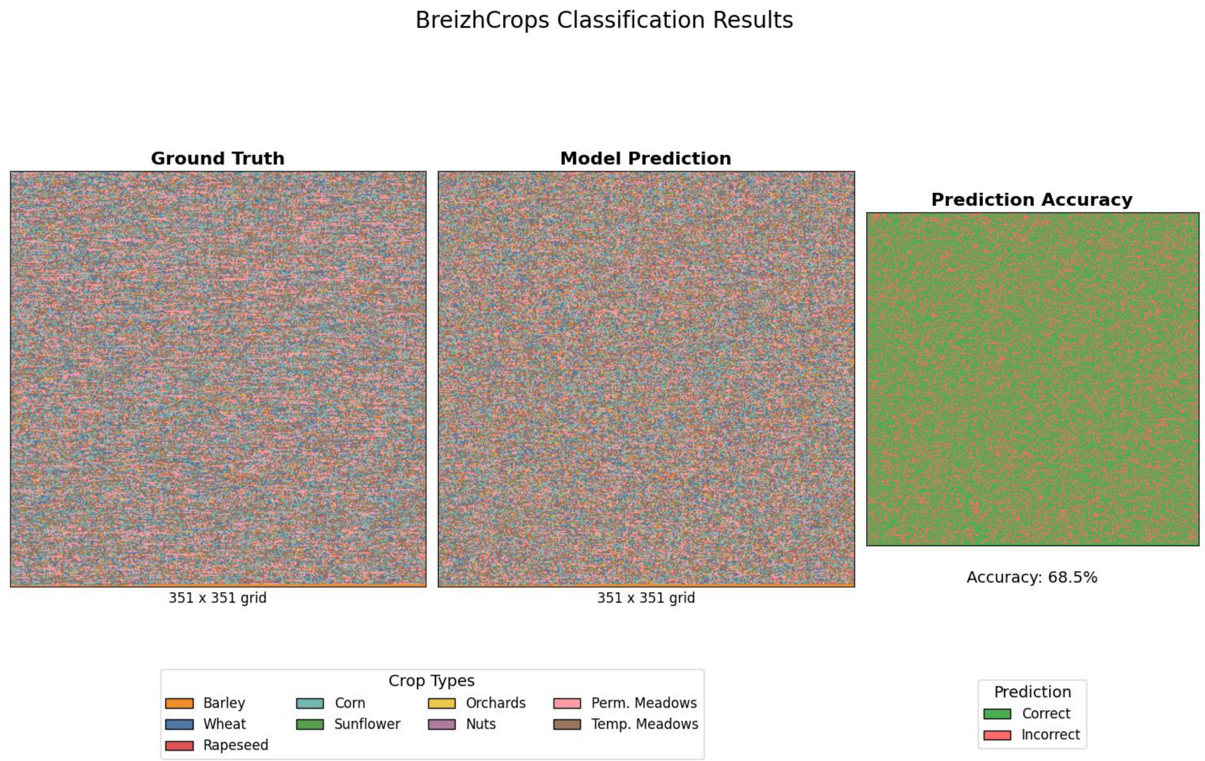


Figure 43: Comparison of ground truth labels, model predictions, and accuracy map for the BreizhCrops dataset. Green indicates correct predictions; red highlights misclassifications.

Description:
This figure shows a side-by-side comparison of the ground truth labels, model predictions, and per-pixel prediction accuracy. Each square represents a sample in a 2D layout, providing a visual abstraction of how well the model distinguishes between crop types. The rightmost panel shows correct predictions in green and incorrect ones in red, with an overall classification accuracy of **68.5%**.

- *Figure D.2 – Per-Class Prediction Accuracy*

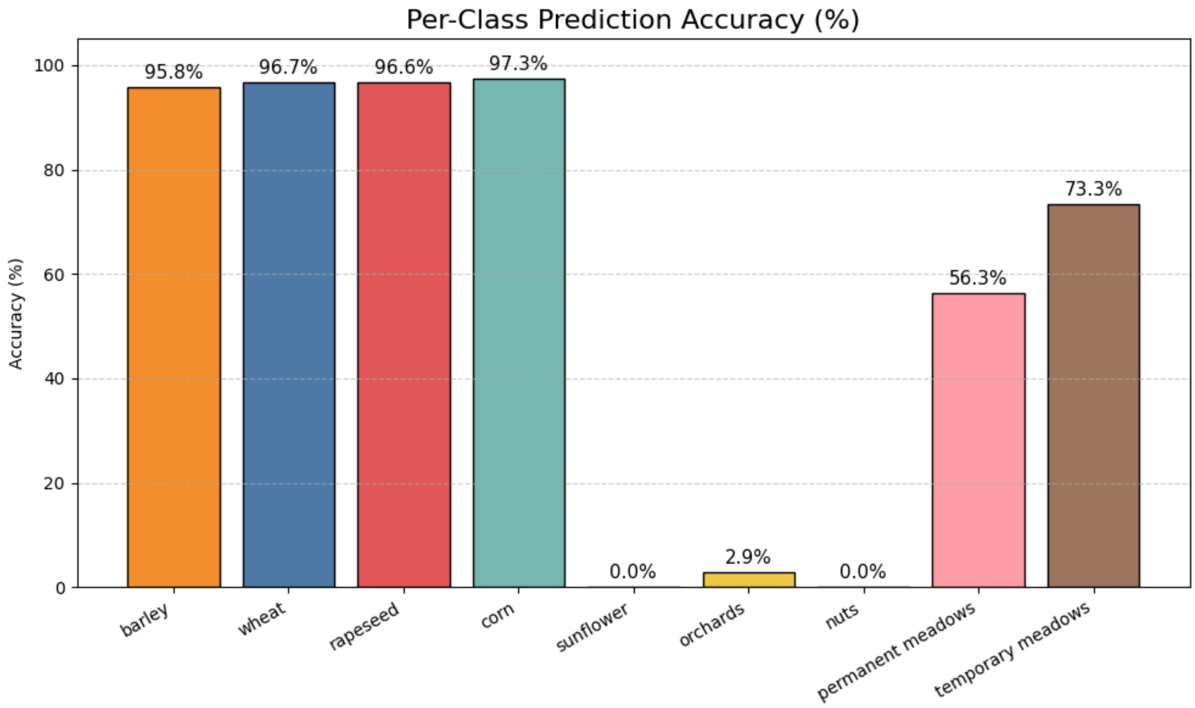


Figure 44: Bar chart showing prediction accuracy for each crop type. High performance is observed for cereals; lower accuracy for rare or spectrally similar classes.

Description:

This bar chart highlights the classification performance for each individual crop class. The model demonstrates strong performance on major crop types such as **barley**, **wheat**, **rapeseed**, and **corn**, all exceeding **95% accuracy**. However, performance degrades significantly for underrepresented classes like **orchards**, **nuts**, and **sunflower**, suggesting a need for class balancing or improved temporal signal modeling.

- *Figure D.3 – Spatial Accuracy Maps per Crop Class*

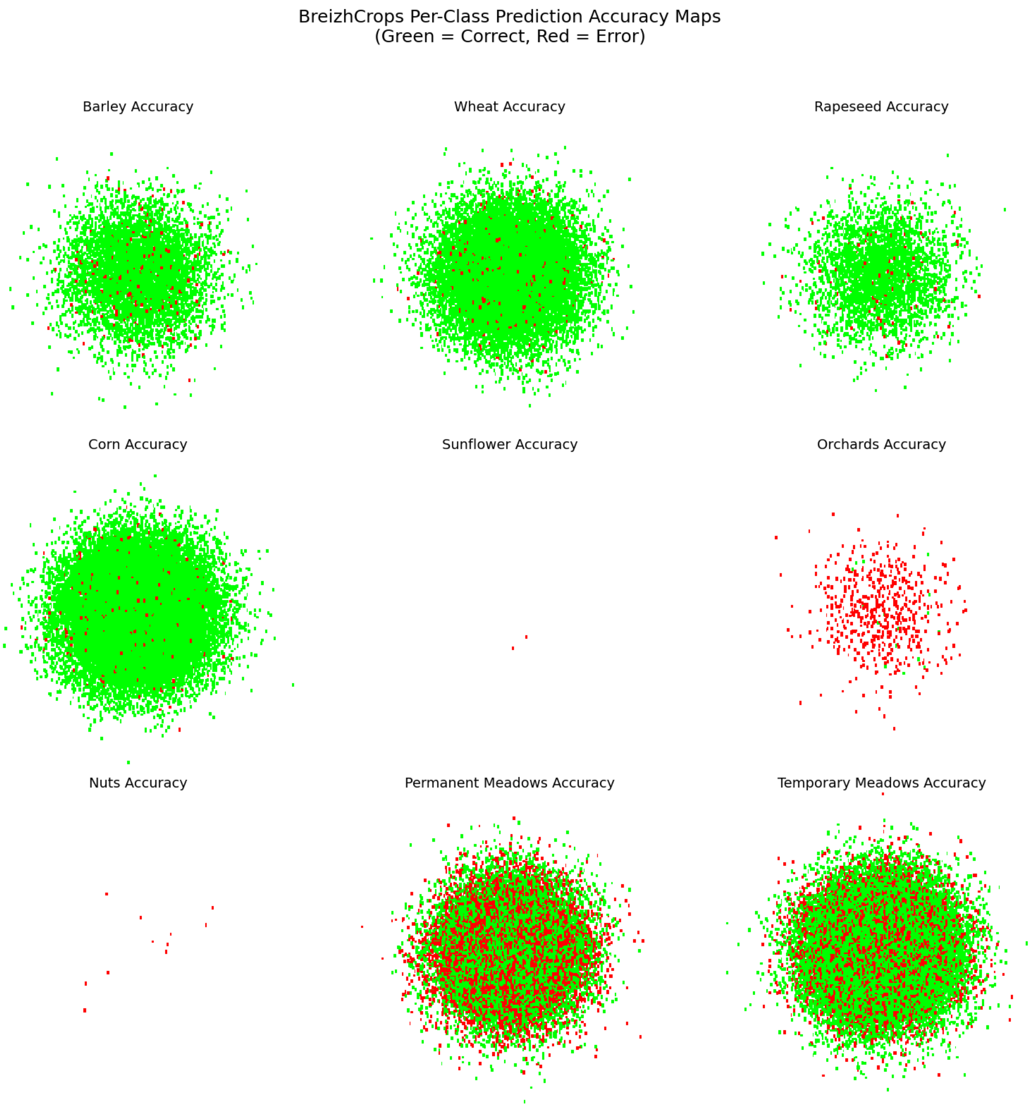


Figure 45: Per-class spatial distribution of prediction accuracy. Correctly classified plots are shown in green; misclassified plots in red

Description:
This composite visualization displays spatial prediction accuracy for each crop type across Brittany. Each subplot corresponds to one class, with **green** patches indicating correct classifications and **red** indicating misclassifications. Crops with high performance (e.g., cereals) show dense green regions, while less frequent crops display scattered red errors. This spatial breakdown is valuable for identifying geographic zones where the model underperforms.

Appendix F: Datasets for Crop Classification in Precision Agriculture

This appendix provides an exhaustive overview of publicly available datasets relevant to precision agriculture, focusing on crop classification tasks. It includes datasets from various modalities such as satellite imagery (multispectral, hyperspectral, radar), UAV-based data, soil and fertility information, and multimodal fusion sources. Each entry details the dataset's category, type, source, spatial/temporal resolution, and access links for easy reference.

This collection serves as a valuable resource for researchers and practitioners seeking to identify, select, and integrate diverse data sources to develop robust, accurate, and scalable crop classification models. The datasets compiled here reflect the state-of-the-art in data availability and foster reproducible research and innovation in agricultural monitoring.

Table 9: Satellite-Based Direct Imagery Datasets

Category	Type	Source	Resolution	Access Link
Multispectral	Satellite	Landsat Series	30m–60m	https://landsat.gsfc.nasa.gov/
	Satellite	Sentinel-2 (MSI)	10m–60m	https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2
Hyperspectral	Satellite	Hyperion	30m	https://data.nasa.gov/dataset/EO-1-Hyperion/ethf-arwz/about_data
	Satellite	EnMAP	30m	https://www.enmap.org/
SAR (Radar)	Satellite	Sentinel-1	10m–40m	https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1
	Satellite	ALOS PALSAR	10m–100m	https://www.earthdata.nasa.gov/data/catalog/...
LiDAR	Satellite	ICESat-2	Global	https://icesat-2.gsfc.nasa.gov/
	Satellite	GEDI	25m footprint	https://gedi.umd.edu/

Table 10:Multimodal Crop Classification Datasets

Dataset Name	Description	Use Case	Access Link
CropScape	National cropland data layer (USDA)	Land cover, crop monitoring	https://nassgeodata.gmu.edu/CropScape/
National-Scale Crop Maps (Germany)	Sentinel + Landsat crop maps	Classification, segmentation	https://zenodo.org/records/5153047
SICKLE	Multi-sensor for cropping parameters	Multispectral crop classification	https://arxiv.org/pdf/2312.00069
Sen4AgriNet	Sentinel-2 with farmer-declared labels	Monitoring, benchmarking	https://github.com/orion-ai-lab/s4a
Sentinel-2 Benchmark	Clean multi-year classification set	Crop type recognition	https://arxiv.org/abs/2204.00951
USDA Cropland Data Layer	Sentinel + Landsat land cover	Agricultural zone analysis	https://www.nass.usda.gov/...
Planted Dataset	Multi-sensor forest & crop ID	Temporal remote sensing	https://arxiv.org/html/2406.18554v1

Dataset Name	Description	Use Case	Access Link
BigEarthNet	Archive for remote sensing ML	Semantic segmentation, classification	https://paperswithcode.com/dataset/bigearthnet
BREIZHCROPS	UAV + Sentinel time series	Crop type mapping	https://arxiv.org/pdf/1905.11893
ClarkCGA Dataset	Multi-temporal crop classifier	Remote sensing temporal learning	https://github.com/ClarkCGA/...
CropAndWeed Dataset	UAV-based crop and weed images	Binary/multi-class labeling	https://github.com/cropandweed/cropandweed-dataset
Early Crop Classification Dataset	Early-season crop fusion set	Time-sensitive crop mapping	https://www.mdpi.com/2072-4292/15/3/799
MDAS Benchmark	Hyperspectral + LiDAR + microwave fusion	Precision Ag, benchmarking	https://essd.copernicus.org/articles/15/113/2023/
PASTIS (Temporal Attention Models)	Sentinel-2 + temporal attention models	Multimodal time series	https://github.com/VSainteuf/pastis-benchmark
CropSpectral Dataset	Hyperspectral for plant species and stress	Health, stress detection	https://cropdata.org/cropspectral
Hierarchical Crop Fusion	Crop rotation + satellite + context fusion	Context-aware classification	https://arxiv.org/abs/2305.12011
GEOBON Soil Nutrients	Earth observation + sampling data	Soil fertility & health	https://geobon.org/bons/thematic-bon/soil-bon/
Global Crop Fertilization (1961–2019)	Crop-specific NPK application rates	Fertilizer trend analysis	https://arxiv.org/abs/2406.10001
Global Plant Nitrogen Traits	Plant nitrogen trait values	Nutrient needs research	https://www.nature.com/articles/s41597-024-03357-2
LandPKS Soil	Soil data + land potential	Land suitability scoring	https://landpotential.org
OpenLandMap Soil	Soil maps via remote sensing + statistics	Global mapping, agriculture	https://openlandmap.org
Soil Macronutrient Assessment	NPK monitoring via RS techniques	Fertility management	https://www.mdpi.com/2072-4292/14/1/81
SoilGrids250m	Global soil property maps from ML	Soil health modeling	https://www.isric.org/explore/soilgrids
Variable Rate N Management	Crop zones + soil + RS for N management	Precision fertilizer application	https://www.eaps.purdue.edu/ebdl/...
WoSIS (World Soil Info)	Global soil profiles (field + RS)	Soil profiling	https://www.isric.org/explore/wosis

Dataset Name	Description	Use Case	Access Link
Climate-Aware Yield Predictions	Yield under climate stress + CropNet	Resilient yield planning	https://arxiv.org/abs/2406.06081
Maize Yield Trials	Multistate maize yield with RS	Benchmarking, model validation	https://datadryad.org/stash/...
Crop Recommendation	Soil + climate + crop labels	Crop suitability scoring	https://www.kaggle.com/datasets/...
Crop Yield Mapping	RS yield prediction datasets	Forecasting, regional yield	https://www.kaggle.com/datasets/...
CalCROP21	173 crop types @ 5.6km resolution	Global monitoring	—
EuroCropsML	Large-scale EU crop classification	Supervised ML training	https://zenodo.org/records/14161939
Fine-Scale Classification	High-resolution spatial data	Field-level mapping	https://www.frontiersin.org/...
Callisto Dataset Collection	Repository of agriculture datasets	ML in agriculture	https: //github.com/Agri-Hub/...
Machine Learning for Ag	Dataset with NPK, pH, climate for AI modeling	Fertilizer and crop advice	https://www.kaggle.com/datasets/...
NASA Harvest Dataset	Satellite + in-situ observations	Food security, crop classification	https://nasaharvest.org/data

Hassan 1st University
National School of Applied Sciences – Berrechid

End of Final Year Internship Project Report

Major: Computer Engineering

Research and Development of an AI System for Early Crop Classification in Precision Agriculture

A Sentinel-2-Based Proof of Concept for Future Multimodal Integration

Prepared by: Salma OUMOUSSA

Supervised by:

- Prof. Diego PELUFFO
- Prof. Khalid BOUIHAT
- Prof. Adil HADDI

“Research does not answer all questions; it refines how we ask them.“

Academic year: **2024/2025**