

**Temporal Patterns and Pixel Precision: Satellite-Based Crop Classification Using Deep Learning and Machine Learning**

Sairam Venkatachalam

Disha Kacha

Devarsh Sheth

Supervised By: Michael Mann, Amir Jafari

# Abstract

Accurate crop classification using satellite imagery is crucial for agricultural monitoring, yield forecasting, and promoting sustainable resource management, yet achieving high precision remains a challenge due to the complex temporal dynamics and spectral similarities among crops. Traditional methods often rely on generated features and struggle to fully exploit the rich spatio-temporal information available in multitemporal Earth observation datasets. To address these limitations, we develop and compare deep learning-based models, including a CNN-BiLSTM hybrid, a TabTransformer ensemble, and a 3D CNN Ensemble, leveraging Sentinel-2 satellite imagery with temporal observations across multiple spectral bands and vegetation indices. We also benchmark these against classical machine learning models such as XGBoost and Random Forests. Our best deep learning model, the CNN-BiLSTM ensemble, achieves a field-level classification Cohen’s Kappa score of 0.77, outperforming traditional machine learning methods by a margin of over 0.1 without the usage of generated features. Ensemble strategies and data balancing techniques enhance performance and reduce confusion among spectrally similar classes. These findings demonstrate the advantages of sequence-based deep learning for large-scale crop classification, offering valuable insights for improving remote sensing-driven agricultural monitoring systems.

Crop Classification, Satellite Imagery, Deep Learning, Machine Learning, Remote Sensing

# Table of Contents:

[**Abstract 1**](#_j48xzc3zdaht)

[**Table of Contents: 2**](#_woi5g5bkxyd6)

[**Introduction 3**](#)

[**Study Area and Description 3**](#)

[**Remote Sensing Approach 4**](#)

[**Feature Extraction vs Deep Learning 4**](#)

[**Related Work 5**](#)

[**Problem Statement 6**](#)

[**Data Description 6**](#)

[**Satellite Imagery Data 6**](#)

[**Field Boundary Data 6**](#)

[**Crop Types and Distribution 6**](#)

[**Methodology 9**](#)

**Classical Machine Learning 9**

[**Pixel-Level**](#) **Analysis** [**11**](#)

**Field-Level Analysis 14**

**Deep Learning 18**

**Pixel-Level Analysis 18**

[**Patch-Level Analysis 27**](#)

[**Results and Discussion 31**](#)

[**Classical Machine Learning 32**](#)

[**Pixel‑Level Analysis 32**](#)

[**Deep Learning 34**](#)

[**Pixel‑Level Analysis 34**](#)

[**Patch‑Level Analysis 35**](#)

[**Discussion 37**](#)

[**Conclusion 38**](#)

[**References 39**](#_je9y3mv0b6wf)

# **Introduction**

Accurate and timely crop classification forms the foundation for precision agriculture, supporting critical activities such as yield prediction, resource management, food security monitoring, and climate-resilient farming practices . The increasing availability of high-resolution Earth observation data, particularly from the Sentinel-2 satellite constellation, has transformed large-scale agricultural monitoring by providing dense multitemporal observations with rich spectral information at spatial resolutions of 10–60 meters and a revisit frequency of approximately five days .

Despite these advancements, accurately classifying crops from satellite imagery remains challenging due to the inherent spectral similarities among different crop types, variability in planting practices, and the influence of environmental factors such as cloud cover and atmospheric conditions. Historically, remote sensing-based crop classification has relied on traditional machine learning approaches leveraging handcrafted features. However, the advent of deep learning has introduced new paradigms for automatic feature extraction, sequence modeling, and representation learning, offering the potential to overcome these limitations and improve classification performance.

This study investigates the application of deep learning and classical machine learning models for crop classification using multitemporal Sentinel-2 data. By analyzing spatial and temporal patterns in the imagery and comparing different modeling strategies, we aim to identify effective methodologies for scalable and precise crop type mapping. The work is based on field-labeled Sentinel-2 imagery provided through the MLHub dataset, specifically curated for agricultural fields in South Africa during the 2017 growing season.

## **Study Area and Description**

The study focuses on agricultural regions across South Africa, a country characterized by a wide range of agroecological zones and diverse cropping systems. Major crops cultivated in the study area include Wheat, Barley, Canola, Small Grain Grazing, and Lucerne/Medics. The region experiences significant climatic variability, ranging from semi-arid to subtropical conditions, which drives distinct seasonal growth patterns across different crop types.

Fields in the study area vary considerably in size, shape, and management practices, introducing spatial and phenological diversity into the classification task. Satellite imagery covering the 2017 growing season was used, capturing observations from early emergence through maturity and harvest phases. This temporal coverage enables the extraction of distinct phenological signatures critical for differentiating between crop types.

## **Remote Sensing Approach**

This project leverages Sentinel-2 Level-2A surface reflectance products, which provide atmospherically corrected multispectral imagery across 13 bands, ranging from the visible to the shortwave infrared regions. For our analysis, six specific sources of information were selected:

Four spectral bands: B2 (Blue, 490 nm) B6 (Red Edge 1, 740 nm) B11 (Shortwave Infrared 1, 1610 nm) B12 (Shortwave Infrared 2, 2190 nm) Two vegetation indices: Enhanced Vegetation Index (EVI) Hue (colorimetric measure derived from reflectance) Each of these sources initially spanned 11 monthly observations during the growing season. However, persistent cloud cover during satellite overpasses led to missing or unreliable data for May 2017 and June 2017, prompting their exclusion from the analysis. Consequently, each pixel was characterized by 10 valid temporal observations per feature source, resulting in a total of 60 input features per pixel.

The multitemporal nature of the dataset allows for tracking crop development over time, providing crucial information for distinguishing between crops with otherwise overlapping static spectral properties.

## **Feature Extraction vs Deep Learning**

Traditional remote sensing approaches to crop classification often involve manual or semi-automated feature extraction. Features such as normalized difference vegetation indices (NDVI), spectral band ratios, textural measures, and phenology metrics are engineered based on domain knowledge and used as inputs to classifiers like Random Forests (RF), Support Vector Machines (SVM), and Gradient Boosted Trees (GBT) . While feature engineering has become faster and more automated through modern toolkits, it may still struggle to capture the complex, nonlinear relationships embedded in multitemporal satellite data.

Deep learning offers an alternative paradigm by learning hierarchical feature representations directly from raw data, without relying heavily on manual intervention. Convolutional Neural Networks (CNNs) excel at capturing local spatial patterns, while Recurrent Neural Networks (RNNs) and Transformer-based architectures are adept at modeling temporal sequences and long-range dependencies . Hybrid models, such as CNN-Transformers, have further demonstrated the ability to integrate local spatial texture with global temporal dynamics, yielding improved classification accuracy across diverse agricultural landscapes .

By comparing classical feature extraction pipelines with deep learning-based approaches, this study aims to highlight the advantages and trade-offs associated with each methodology for large-scale crop classification.

# **Related Work**

Crop classification using satellite imagery has become a cornerstone of precision agriculture, enabling large-scale monitoring, yield estimation, and resource management. Over the past decade, advances in both remote sensing technology and machine learning algorithms have significantly improved the accuracy and scalability of crop mapping efforts.

1. **Traditional Machine Learning Approaches:** Early research in crop classification predominantly utilized classical machine learning algorithms such as Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression. These models rely on engineered features derived from the spectral and temporal properties of satellite imagery. Ensemble methods such as LightGBM and XGBoost have also gained popularity due to their ability to handle high-dimensional data and complex feature interactions. However, these models often require extensive feature engineering and may struggle with spatial-temporal heterogeneity.
2. **Deep Learning and Feature Learning:** Recent years have witnessed a paradigm shift toward deep learning, particularly convolutional neural networks (CNNs) and transformer-based architectures. Deep learning models can automatically extract hierarchical features from raw multispectral time series. and both reported that deep learning models outperformed traditional methods in terms of classification accuracy. NASA and IBM’s Prithvi-100M model exemplifies the state-of-the-art.
3. **Ensemble and Hybrid Methods:** Ensemble approaches combining multiple deep learning models have been explored. showed that an ensemble of CNNs paired with advanced feature extraction techniques achieved accuracy rates exceeding 98% for cotton crop classification.
4. **Remote Sensing Data and Feature Engineering:** Sentinel-2 has become a preferred source for agricultural monitoring. Vegetation indices such as NDVI and EVI enhance class separability.
5. **Temporal Analysis and Multi-Temporal Data:** Temporal dynamics are essential for distinguishing crops with similar spectral signatures. Multi-temporal analysis using recurrent neural networks (RNNs) has been shown to enhance classification accuracy.

# **Problem Statement**

This study aims to leverage multi-spectral Sentinel-2 satellite imagery to classify crop types within a substantial agricultural area in South Africa. The study area is comprised of distinct field polygons, with each polygon representing a field where a single crop is grown. The primary objective is to develop an effective classification methodology that utilizes the spectral and temporal information provided by Sentinel-2 imagery to accurately identify the crops cultivated in these fields. This approach is expected to enhance precision in agricultural monitoring and support informed decision-making in crop management practices.

# **Data Description**

The dataset used for this study, consists of 2 primary components - Satelite imagery and Field Boundary Data sources from

## **Satellite Imagery Data**

Sentinel-2 Level-2A surface-reflectance data were acquired using a sensor sentinal-2a monthly cloud free composites over our study area in South Africa for the period between January–December 2017 growing season. Monthly median composites were generated with a 30% s2cloudless cloud-probability filter and a morphological cloud-shadow mask clipped to the field boundaries, scaled to integer reflectance (×10 000), and exported as 10 m single-band GeoTIFFs via geemap.ee\_export\_image. From each composite we extracted bands B2, B6, B11, and B12, as well as the Enhanced Vegetation Index (EVI) and Hue.

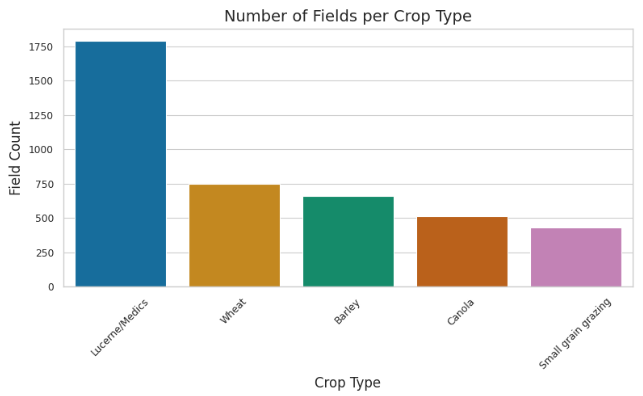
## **Field Boundary Data**

A GeoJSON file delineates the field boundaries, defining polygons that segment the imagery into homogeneous crop units. All prediction are made at the field level

## **Crop Types and Distribution**

We classify five crop types in this study: wheat, canola, lucerne/medics, barley, and small‐grain grazing.

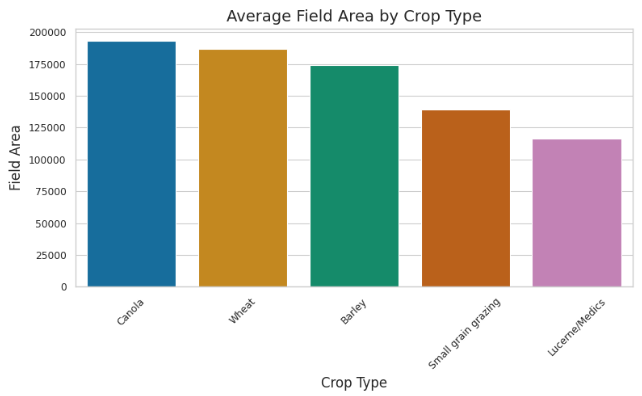
|  |
| --- |



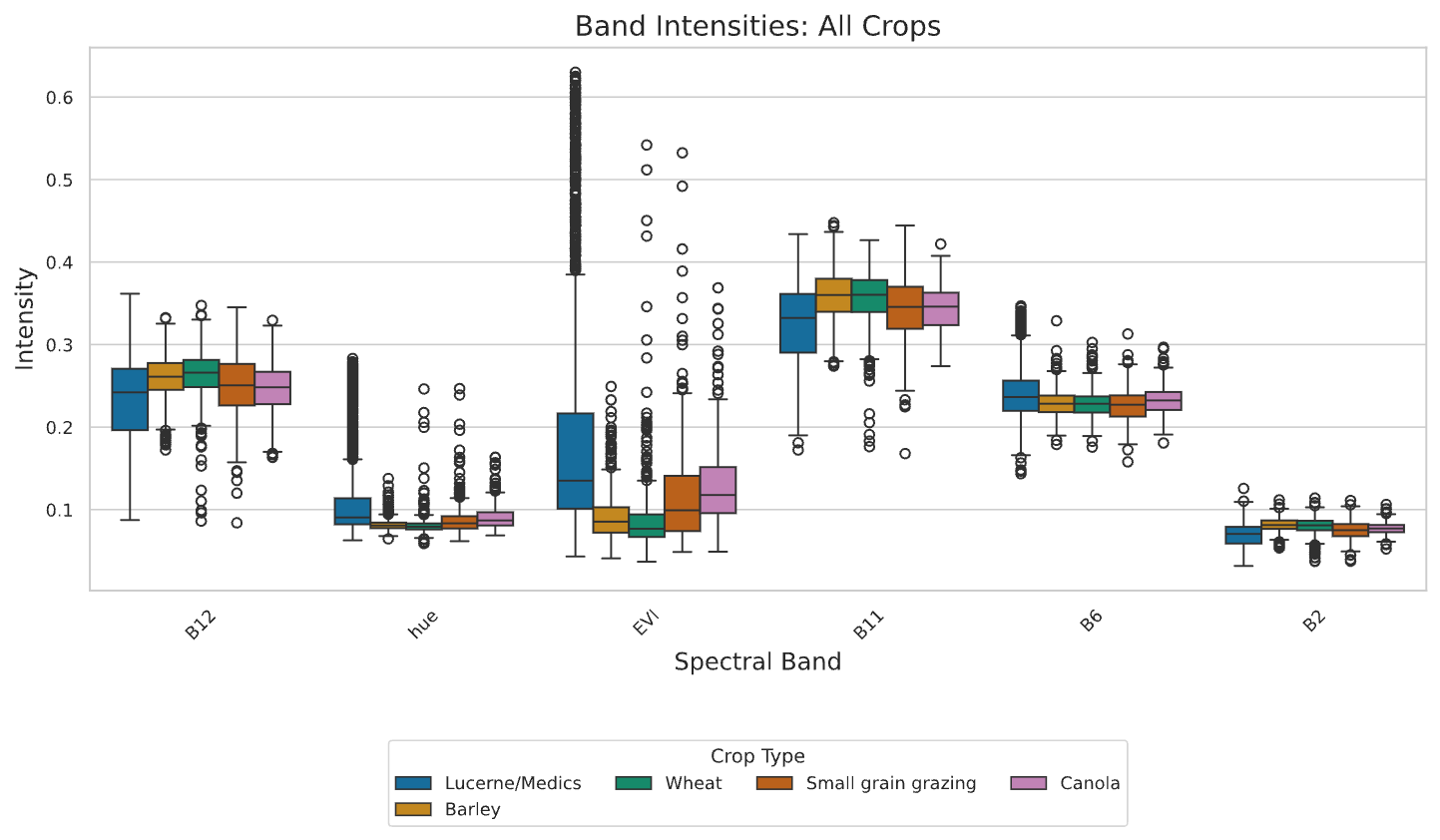
*Fig 1: Number of Fields per Crop Type*

Figure 1 illustrates the distribution of crop fields across the study area. The data is notably imbalanced, with the majority of fields dedicated to Lucerne/Medics. The imbalance in crop representation has important implications for model training and evaluation, as classification algorithms may be biased towards the majority class unless appropriate balancing techniques are applied.

Figure 2 presents the mean field area for each crop. On average, while being the majority of fields, Lucerne/Medics crop fields are smaller compared to other crop fields.

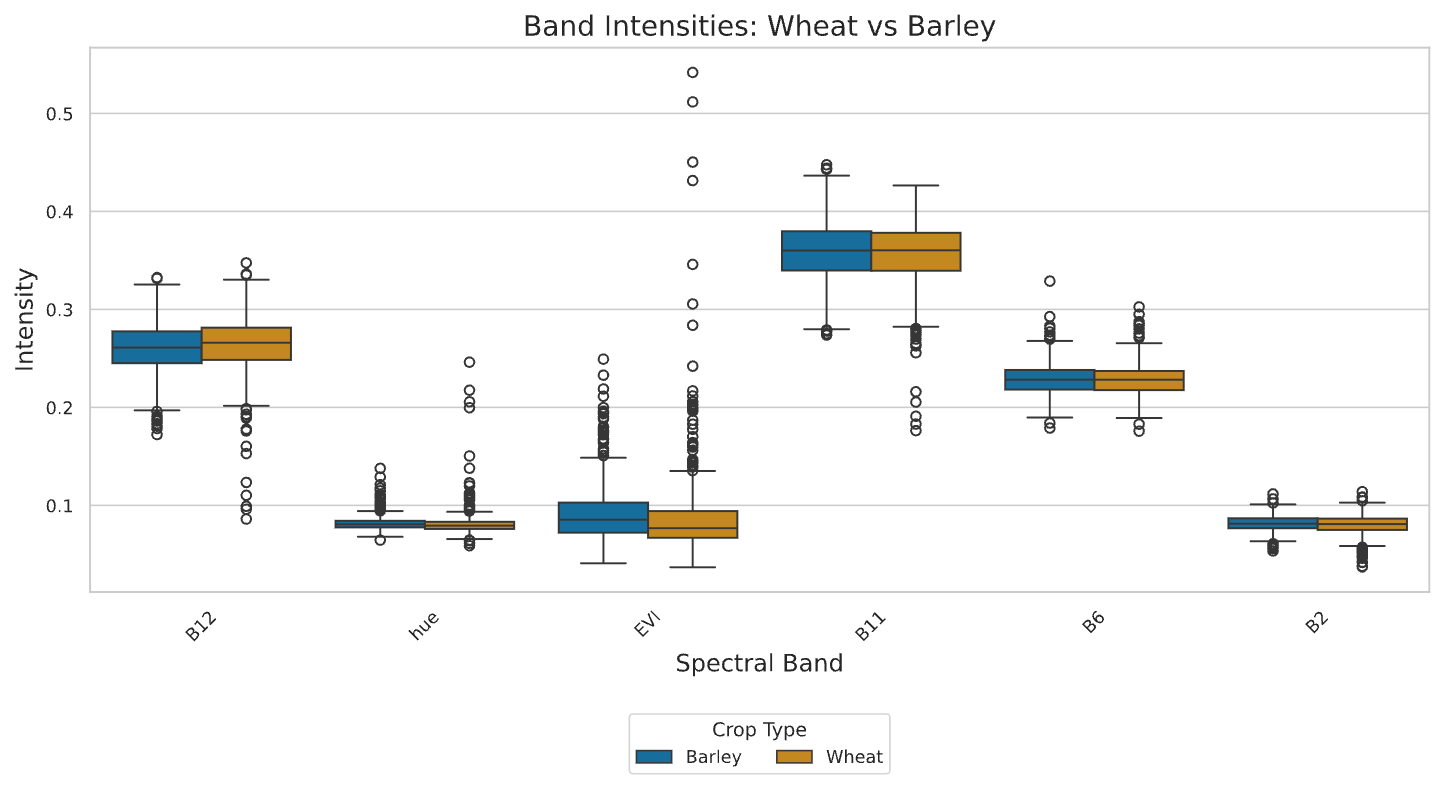


*Fig 2: Average field area by crop type.*



*Fig 3: All crops Box Plot*

Figure 3 displays boxplots of spectral band intensities for each crop type, illustrating the distribution and overlap of spectral responses. Although some differences in median and spread are evident, the substantial overlap in spectral intensities across crops demonstrates that classification based solely on a limited set of spectral bands is challenging and prone to confusion. This is particularly true for cereal crops such as wheat and barley, which exhibit highly similar spectral profiles, as shown in Figure 4. The close alignment of their spectral signatures underscores why models often struggle to distinguish between these two crops using spectral data alone .



*Fig 4. Boxplots comparing spectral band intensities of wheat and barley, highlighting their high degree of spectral similarity and potential for classification confusion.*

These findings highlight the necessity of incorporating temporal information such as phenological changes throughout the growing season alongside spectral features to improve crop classification. The unique temporal dynamics associated with crop development stages (e.g., emergence, peak greenness, senescence) provide additional discriminatory power that can help resolve ambiguities where spectral overlap exists . Integrating both spectral and temporal data is therefore critical for enhancing classification accuracy, especially in complex agricultural landscapes where crops like wheat and barley are spectrally similar.

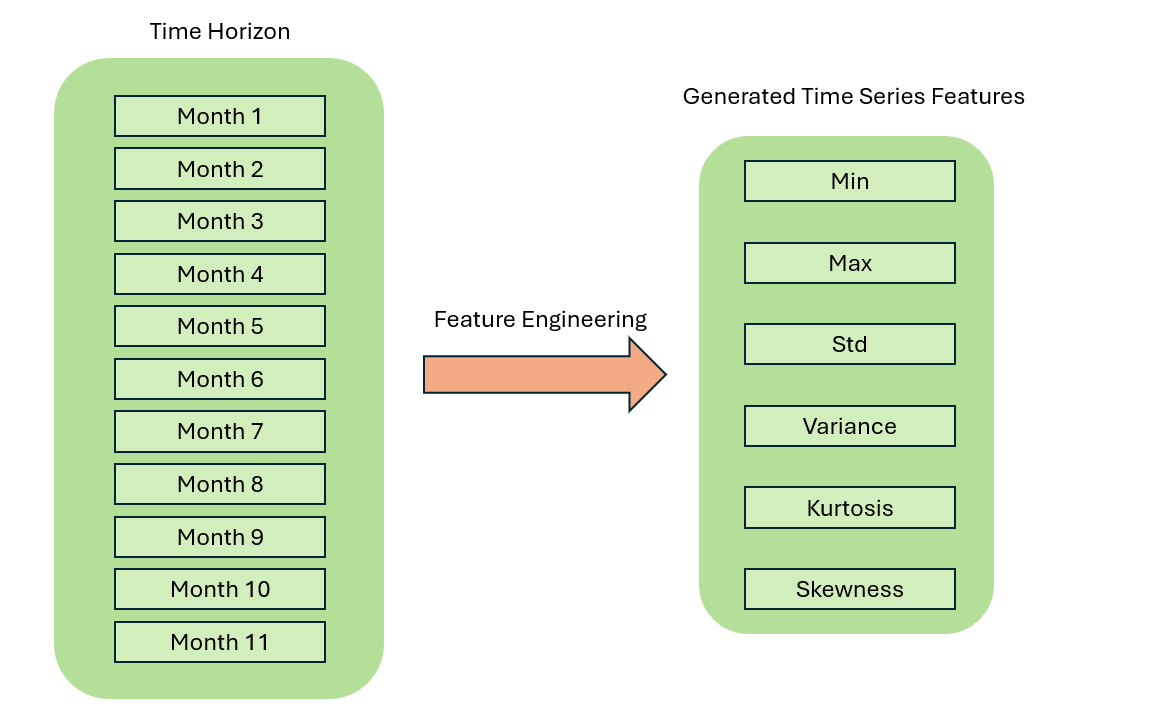
|  |
| --- |

# **Methodology**

This study details two primary modeling paradigms: Classical Machine Learning and Deep Learning. Each paradigm explores both pixel-level and field/patch-level strategies to exploit spatial and temporal information.

## **Classical Machine Learning**

In the classical approach, we automate feature extraction using the xr\_fresh toolkit to rapidly compute a comprehensive set of statistical and temporal features from each 11-month spectral time series . To address class imbalance, we apply SMOTE. Several classifiers such as Logistic Regression, Random Forest, LightGBM, and XGBoost, are then evaluated.



*Fig 5.: Extraction of statistical and temporal features from the 11-month spectral time series.*

We extract the following features from each pixel’s 11-month spectral profile:

* **Absolute Energy:** Sum of squared values, capturing overall signal magnitude.
* **Absolute Sum of Changes:** Total absolute difference between consecutive measurements, reflecting temporal variability.
* **Autocorrelation (1,2-month lag):** Correlation with its one- and two-month lagged versions, indicating series persistence.
* **Count Above Mean:** Number of points above the series mean, quantifying high-value occurrences.
* **Day of Year of Max/Min:** Julian day when the maximum and minimum occur, marking peak and low activity periods.
* **Kurtosis:** Tail heaviness of the distribution, highlighting propensity for extremes.
* **Linear Trend:** Slope of a least-squares fit, summarizing overall upward or downward trend.
* **Longest Strike Above/Below Mean:** Longest consecutive run above or below the mean, capturing sustained periods.
* **Max/Min:** Absolute extreme values in the series.
* **Mean/Median:** Measures of central tendency.
* **Mean Absolute Change/Mean Change:** Average magnitude and signed change between consecutive points, indicating volatility and direction.
* **Mean Second Derivative:** Mean of the series’ second differences, detecting acceleration or deceleration.
* **Quantiles (0.05, 0.95):** Values at the 5th and 95th percentiles, summarizing distribution extremes.
* **Ratio Beyond** **:** Proportion of points more than  standard deviations (r=1,2,3) from the mean, highlighting outliers.
* **Skewness:** Asymmetry of the distribution around the mean.
* **Standard Deviation/Variance:** Spread measures around the mean.
* **Sum of Values:** Total sum of observations (mean × length).
* **Symmetry Looking:** Similarity when the series is reversed, measuring mirror-like behavior.
* **Time Series Complexity:** Complexity-Invariant Distance metric of sequence irregularity.

Leveraging the xr\_fresh toolkit allowed us to extract all required time series features in a quick and automated fashion, ensuring each pixel’s temporal signature is fully captured for reliable change detection and environmental analysis .

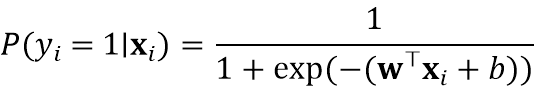
For clarity, we first present the pixel-level analysis followed by the field-level analysis, allowing a direct comparison between the two approaches.

### **Pixel-Level Analysis**

In the pixel-level analysis, each pixel is treated as an independent sample, and the extracted features for each pixel’s 11-month history are used to train classifiers. During inference, individual pixel predictions within a field polygon are aggregated through majority voting to determine the overall field classification. This method benefits from a large number of training samples and captures fine-grained temporal patterns.

1. **Baseline Classical Models: The Foundation**

* To benchmark our crop classification task, we selected four classical machine learning models: Logistic Regression, Random Forest, LightGBM, and XGBoost. These models provide a progression from linear interpretability to advanced tree-based learners capable of modeling complex, nonlinear relationships. All models were trained using time-series descriptors derived from Sentinel-2 pixel data as discussed in  [5.1](#a3d7g6n0p0ew).
* Each pixel was represented by a feature vector , constructed from the extracted time-series features across ten months for six spectral indices (B2, B6, B11, B12, EVI, Hue). This multivariate vector encodes the phenological profile of a pixel over time, capturing crop growth dynamics and seasonal patterns.
  1. **Logistic Regression (LR): Linear Interpretability**  
     In Logistic Regression, the decision boundary is a linear combination of input features. The predicted class probability is:



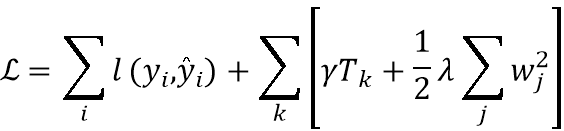
* + Each feature contributes additively to the log-odds of a class . LR provides interpretability, where the weights  reveal the importance of specific descriptors.
  1. **Random Forest (RF): Nonlinear Feature Combinations**  
     Random Forests are ensembles of decision trees trained on random subsets of features and samples . The final prediction is made via majority voting:



* + RFs capture non-linear interactions and are robust to overfitting . Feature importance can be derived from decision splits.
  1. **LightGBM: Boosted Gradient Trees with Leaf-wise Growth**  
     LightGBM builds additive decision trees to minimize residuals in a leaf-wise manner :



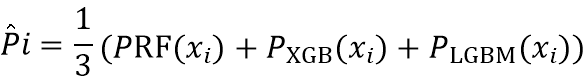
* + where  is the learning rate and  is the function learned at iteration . This approach allows efficient learning on large, high-dimensional datasets.
  1. **XGBoost: Regularized Gradient Boosting**  
     XGBoost improves on traditional boosting by incorporating regularization in its loss function :

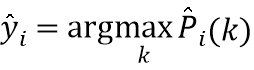


* + Regularization (via  and ) controls model complexity and prevents overfitting on rare classes .

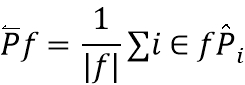
1. **Pixel-Level Ensemble with Hybrid Voting:**

* To improve pixel-level generalization and mitigate model-specific biases, we employed an ensemble of three independently trained classifiers, Random Forest, XGBoost, and LightGBM using identical feature representations . Each model in the ensemble predicted class probabilities per pixel, which were then averaged (soft voting) to compute final predictions:





* Here,  denotes the average class probability vector across  crop classes for pixel .
* While soft voting improved pixel-level metrics, it did not always yield stable field-level decisions especially for ambiguous fields. To address this, we designed a Hybrid Voting mechanism that dynamically chooses between soft voting and mode voting based on confidence margins.
  1. **Soft Voting (Field-Level)**  For each field , soft-voted probabilities were averaged across all its pixels:



* + The predicted label was the class with maximum average probability:



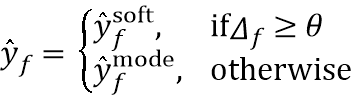
* 1. **Confidence Margin (Top-2 Gap)**  We define a top-2 class gap to quantify prediction confidence:



* + where  and  are the highest and second-highest class probabilities for field .
  1. **Mode Voting (Fallback Strategy)**  If  (empirically set to ), we revert to mode voting—i.e., the most frequent class label among pixel predictions:



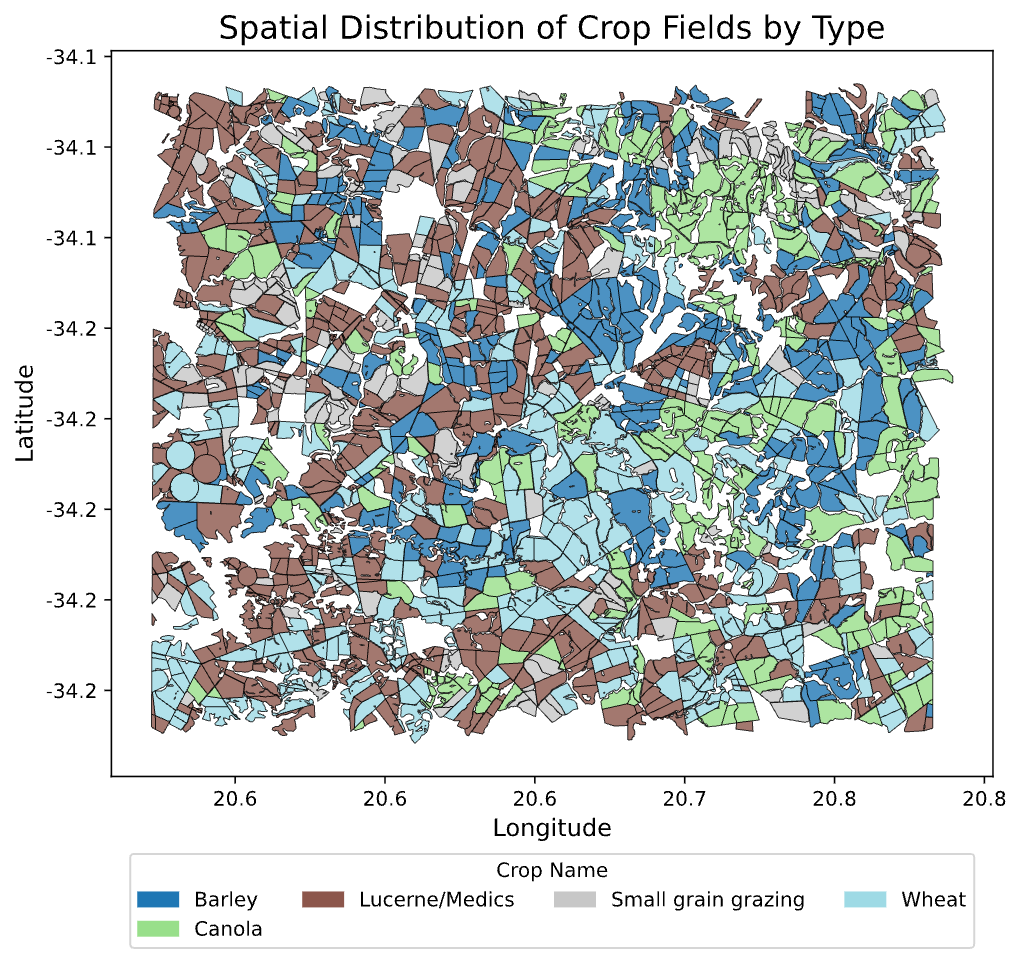
* 1. **Final Hybrid Rule**



* + This method improved field-level stability by balancing probabilistic confidence with categorical consistency.

### **Field-Level Analysis**

In field-level analysis, all pixel features within a field polygon [6](#g0pwgv7rm2kb) are aggregated by averaging to create a single feature vector per field. This aggregation reduces the influence of noisy or anomalous pixels and focuses on field-wide behavior. Classifiers are then trained on these per-field features, directly modeling the spatial coherence of each crop type at the field scale. The main disadvantage to this method is that the dataset reduces considerably; however, the advantage is that noise is eliminated, as noted by Maxwell and Warner (2019) in their discussion of object-based image analysis . Field-level analysis is particularly suited for agricultural landscapes with relatively homogenous crop cover, where aggregating pixel information provides a robust characterization of field conditions.



*Fig 6.: Illustration of field-level aggregation: pixel values are grouped by field polygon and summarized into a single feature vector.*

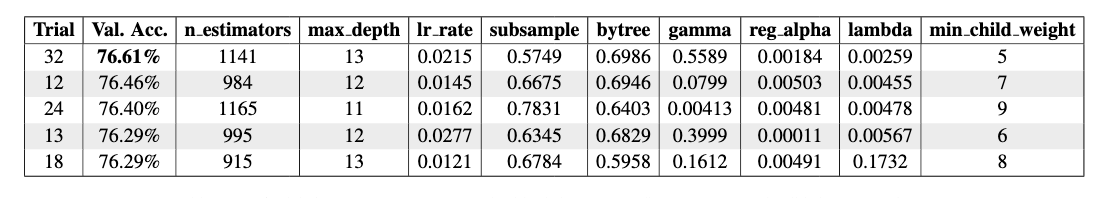
1. **SMOTETomek + Stacked Ensemble Framework**

* To address the challenge of class imbalance in field-level crop classification, we integrated a data-level resampling strategy with a model-level ensemble framework. This approach combines the **SMOTETomek** technique with a **stacked ensemble architecture**, ensuring both balanced training data and enhanced model generalization.
  1. **Data Balancing via SMOTETomek**  
     The training dataset exhibited significant imbalance, with certain crops like small grain grazing, canola being severely underrepresented as seen in Figure [1](#44v40nsh0nz4). To mitigate this, we employed the SMOTETomek resampling strategy:
     + **SMOTE (Synthetic Minority Over-sampling Technique)** synthetically generated new samples for minority classes by interpolating between existing observations in the feature space.
     + **Tomek Links** identified and removed ambiguous samples near class boundaries, enhancing class separability and reducing overlap.
  + This resampling process yielded a more balanced and clean dataset, which better represented minority classes without overwhelming the majority class distribution.
  1. **Stacked Ensemble Architecture**  
     Post-resampling, we constructed a stacked ensemble classifier composed of the following components:
     + **Base Learners:**
       - A **Random Forest (RF)** model to capture robust non-linear patterns through bagged decision trees.
       - An **XGBoost (XGB)** model to exploit gradient-boosted trees for fine-grained pattern recognition and error correction.
     + **Meta Learner:**  
       A **Logistic Regression (LR)** model trained on the outputs (probabilities or class predictions) of the base learners, as well as the original input features (enabled via passthrough=True). This meta-learner models the inter-model relationships and refines final predictions through probabilistic calibration.
  + Formally, the prediction from the stacked model is represented as:



* + where  and  are the base classifiers and  is the logistic regression model acting as the meta-learner.
  1. **Justification**  
     The rationale behind this approach was twofold:
     + SMOTETomek enables the model to fairly learn from all crop categories, especially minority crops that otherwise risk being ignored.
     + The stacked ensemble leverages the strengths of different model families: tree-based ensembles for capturing complex relationships, and linear models for interpretability and calibration.
  + This integrated methodology allows for both **resilient classification performance** and **balanced class representation** during learning, addressing both data-driven and model-driven limitations in crop type prediction.

1. **Field-Based Aggregation + XGBoost Optimization Framework**

* To achieve reliable field-level crop classification, we developed an integrated framework that first aggregates pixel-level information to field-level summaries, and then optimizes a powerful XGBoost classifier via extensive hyperparameter tuning. This two-stage methodology ensures both meaningful data representation and robust model performance.
  1. **Field-Level Data Aggregation**  
     To respect the spatial structure of the dataset and prevent data leakage between train and test sets, we employed a field-based aggregation strategy:
     + **Train-Validation-Test Split:**
       - Unique field identifiers (fid) were split into training, validation, and testing groups (80%-20% split, with training further split into train-validation).
       - This ensured that no field contributed data to multiple sets, maintaining strict independence.
     + **Aggregation Logic:**
       - **Features:** For each field, pixel-level numerical features were aggregated using their **mean values**.
       - **Labels:** The field label was determined as the **mode** (most frequent crop type) among all its pixel-level labels.
  + This aggregation process converted the pixel-level dataset into a **field-level dataset**, drastically reducing redundancy and emphasizing field-wide spectral and statistical characteristics.
  + 
  + *Table 1.: Top 5 Trials from Hyperparameter Search with Their Corresponding Parameters and Validation Accuracies*
  1. **Extensive XGBoost Hyperparameter Tuning via Optuna**  
     Following aggregation, a hyperparameter tuning phase was conducted using the **Optuna** optimization library to maximize model generalization:
     + **Search Space:**
       - Number of estimators (n\_estimators) ranging from 200 to 1500.
       - Tree depth (max\_depth) ranging from 3 to 15.
       - Learning rate (learning\_rate) from 0.005 to 0.3 (log-uniform distribution).
       - Regularization terms (reg\_alpha, reg\_lambda) optimized on log scales.
       - Tree subsampling parameters (subsample, colsample\_bytree) between 0.5 and 1.0.
       - Additional parameters like gamma and min\_child\_weight.
     + **Optimization Objective:**
       - 5-Fold Stratified Cross-Validation was performed within each trial.
       - The **weighted F1 score** across folds was used as the evaluation metric.
     + **Selection Process:**
       - 75 optimization trials were conducted to thoroughly explore the hyperparameter space.
       - The best parameter set found by Optuna was used for final model training.
  + This extensive tuning process resulted in a highly tailored XGBoost model specifically optimized for the field-level crop classification task.
  1. **Justification**  
     The rationale behind this framework was threefold:
     + **Field-Level Aggregation** ensured that the model captured consistent patterns across entire fields, rather than noisy pixel-level artifacts.
     + **Optuna-Based Tuning** allowed for discovering a near-optimal hyperparameter configuration efficiently without manual grid search, which can be seen in Table 1.
  + This integrated approach enabled the development of a highly robust, generalizable field-level crop type classifier while adhering to best practices in both data preparation and model optimization.

## **Deep Learning**

While classical machine learning offers valuable interpretability and computational efficiency, the feature engineering process, although increasingly automated and efficient, may still struggle to fully capture the complex spatial and temporal dependencies inherent in satellite imagery . To overcome these limitations and leverage recent advancements in representation learning, we turn to deep learning approaches, which have shown great promise in automatically extracting hierarchical features from raw data for various remote sensing tasks .

Deep Learning models ingest raw spectral values and learn hierarchical representations that capture both spatial and temporal dependencies. We explore several architectures: 2D Convolutional Neural Networks (CNNs) that treat spectral bands as input channels, 3D CNNs that jointly model spatial and temporal dimensions, transformer-based models for long-range temporal attention, and ensemble approaches that combine multiple deep architectures for improved robustness.

### **Pixel-Level Analysis**

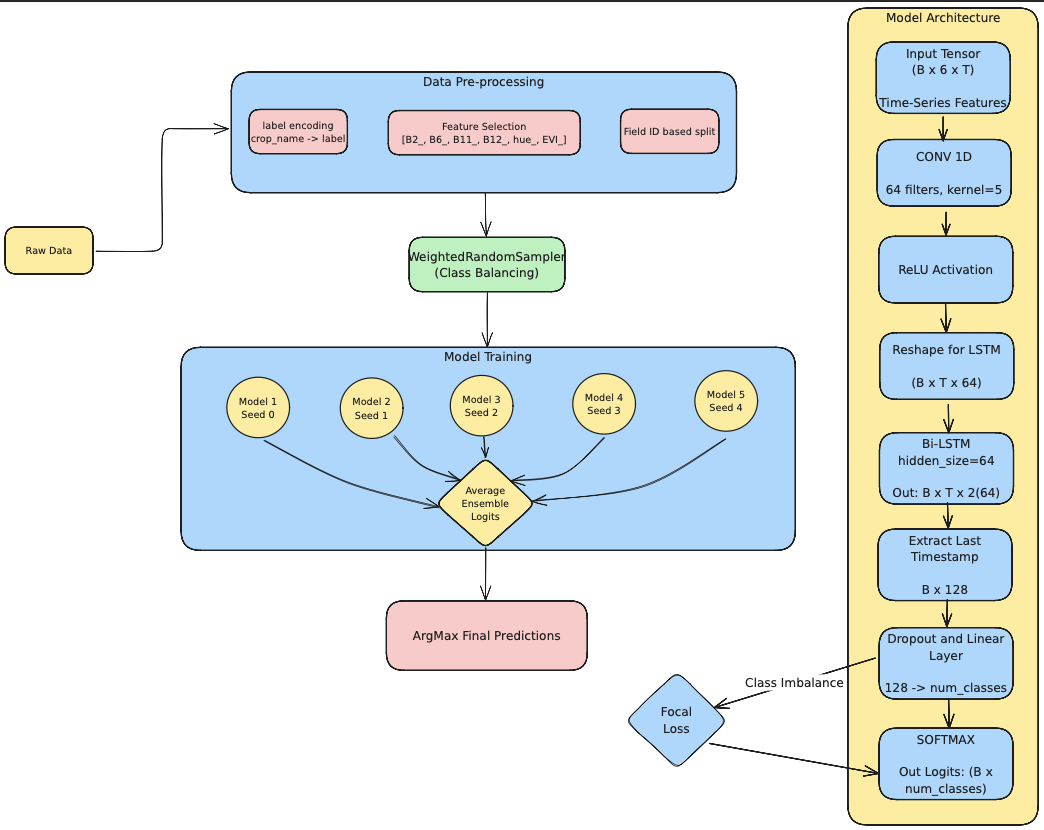
Accurately modeling agricultural land cover at the pixel level is challenging due to seasonal variability, crop phenology, and mixed-crop patterns commonly observed in satellite imagery. Building on the multitemporal features extracted from Sentinel-2 imagery, we constructed a structured dataset where each row corresponds to an individual pixel characterized by 60 temporal features spanning the 2017 growing season.

Each pixel is linked to a unique field identifier (fid) and labeled according to the ground-truth crop type, covering five major crop classes: Wheat, Barley, Canola, Small Grain Grazing, and Lucerne/Medics. To ensure unbiased model evaluation and prevent spatial information leakage, we employed a field-aware data splitting strategy. In this approach, all pixels belonging to a given field are assigned exclusively to either the training, validation, or test set, thereby avoiding scenarios where a model could inadvertently learn from spatially neighboring pixels during both training and evaluation.

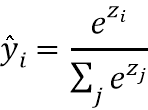
This pixel-level organization serves as the foundation for subsequent modeling stages, enabling both classical machine learning and deep learning architectures to leverage temporal crop signatures for classification.

For the TabTransformer models, we applied z-score standardization to ensure scale consistency across numeric inputs. In contrast, the CNN-based models were trained directly on the raw or scaled reflectance values without explicit normalization, allowing the convolutional layers to learn directly from the temporal magnitude patterns.

1. **CNN + BiLSTM Ensemble with Focal Loss:**  
   To capture the intricate temporal dynamics of crop growth patterns from satellite data, we developed a hybrid deep learning architecture that combines the local pattern recognition capability of 1D Convolutional Neural Networks (CNNs) with the sequence modeling power of Bidirectional Long Short-Term Memory (BiLSTM) networks [7](#uhook5qfg7s4). This model is uniquely designed to leverage the multitemporal nature of remote sensing data and mitigate class imbalance via Focal Loss, while also stabilizing performance through ensembling.

*Fig 7.: Architecture of the CNN + BiLSTM model used for pixel-level crop classification. The model combines convolutional temporal feature extraction with sequential modeling through a BiLSTM layer.*

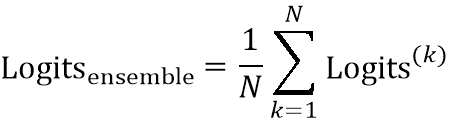
* 1. **Temporal Representation Through CNN + BiLSTM:**  
     The architecture processes an input tensor of shape , where each of the six channels corresponds to a spectral or vegetation index (B2, B6, B11, B12, hue, EVI) recorded across 10 months. The sequence begins with a 1D Convolutional Layer with 64 filters and a kernel size of 5, applied along the temporal axis. Post-convolution, a ReLU activation is applied, followed by a tensor permutation to format  for compatibility with the BiLSTM.
  + This is then processed by a Bidirectional LSTM with a hidden size of 64. The effectiveness of BiLSTMs in agricultural classification settings, particularly for Sentinel-1 and Sentinel-2 based time-series data, has been well-documented in prior research demonstrating up to 89.1% accuracy across 15 crop classes .
  + From the BiLSTM output of shape , we extract the last timestep representation, followed by dropout and a fully connected layer. The softmax layer outputs class probabilities:



* 1. **Handling Class Imbalance: Focal Loss:**  
     To address class imbalance, we used the Focal Loss, defined as:



* + where  is the weighting factor,  the focusing parameter (set to 2.0), and  the probability for the true class. This imbalance-aware loss is beneficial for underrepresented crops .
  1. **Balanced Training via Weighted Sampling:**  
     A WeightedRandomSampler ensured equal representation of all crop classes by sampling inversely proportional to class frequencies.
  2. **Ensemble Learning for Robustness:** To reduce variance, we trained five identical CNN + BiLSTM models with different seeds. Their logits were averaged:

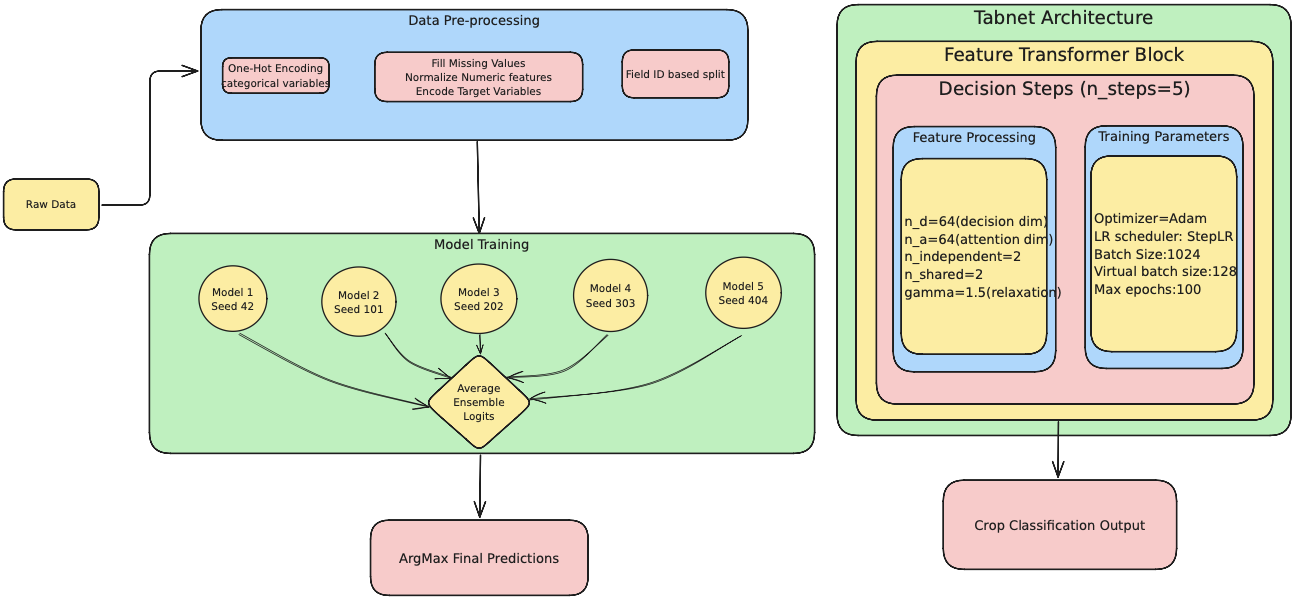


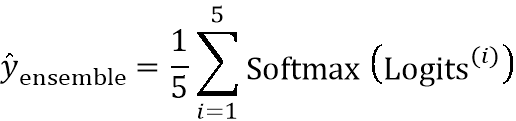
* + This improved generalization and stabilized performance .
  1. **Field-Level Aggregation: From Pixels to Parcels:** To align with field-level annotations, we applied majority voting:



* + This reduced noise and improved accuracy, a strategy validated in prior remote sensing work .

1. **Ensemble of TabTransformer Models (TabNet):**

* While sequence-based architectures such as CNNs and LSTMs are effective for capturing dynamic temporal patterns in remote sensing data, certain crop classification scenarios call for a different perspective, one that treats satellite observations as structured tabular inputs, where each pixel’s multiple measurements (e.g., spectral bands and vegetation indices across different months) are organized into rows and columns like a spreadsheet. To address this, we incorporated TabNet, a deep learning architecture specifically designed for learning from tabular data. TabNet not only processes multitemporal inputs in this flattened format, but also enables interpretable learning through its attention-driven architecture [8](#bbx57ep0fz3p).
  1. **TabNet for Structured Temporal Features:**
  + TabNet departs from traditional deep networks by integrating feature selection and attention into each layer of its architecture. It does this through two key components: the Feature Transformer and the Attentive Transformer. At every decision step, the Attentive Transformer generates a sparse mask over the input features, dynamically selecting a subset of dimensions most relevant for prediction. This mask guides the Feature Transformer, which performs nonlinear transformations on the masked inputs.
  + This iterative process repeated across a series of decision steps enables TabNet to focus on different subsets of features at different depths in the network. Such selective attention empowers the model to uncover complex patterns that arise across combinations of spectral and vegetation index features aggregated from satellite time-series. The model’s ability to navigate high-dimensional tabular input spaces without relying on temporal order has shown to be especially effective in flattened multitemporal settings, as well as when additional metadata or engineered features are included .
  1. **Training Setup and Parameterization:**
  + The network architecture consists of:
    - Decision dimension () = 64
    - Attention dimension () = 64
    - Independent block width = 2
    - Shared block width = 2
    - Relaxation parameter () = 1.5
    - Number of decision steps = 5
  + Optimization was performed using Adam with a StepLR learning rate scheduler. The model was trained with a batch size of 1024, a virtual batch size of 128 (for stabilization), and a maximum of 100 epochs. Early stopping was used to prevent overfitting by monitoring performance on a validation set constructed using field ID-based stratification.
  + *Fig 8.: Architecture of the TabTransformer (TabNet) ensemble used for pixel-level crop classification. Each TabNet model applies iterative attention-based feature selection. Final predictions are obtained by averaging across five models, followed by field-level aggregation using majority voting.*
  1. **Ensembling for Generalization and Robustness**
  + Given the stochastic nature of training and the sensitivity of deep tabular models to initialization, we constructed an ensemble of five TabNet models, each initialized with a different random seed (42, 101, 202, 303, 404). This diversity in training paths encourages each model to learn slightly different representations, making the ensemble more robust to noise and variance.
  + After all five models were trained independently, we averaged their softmax outputs for each input instance to compute the final class probabilities:



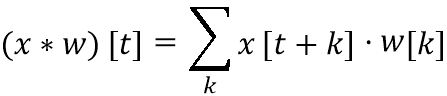
* + The final predicted label was obtained via argmax over these averaged probabilities.
  1. **Field-Level Aggregation for Agricultural Relevance**
  + Although TabNet operates on pixel-level features, practical applications in crop monitoring require field-level decisions. To translate pixel-level predictions into agriculturally actionable outputs, we applied a majority voting strategy over all pixels associated with the same fid (field ID):

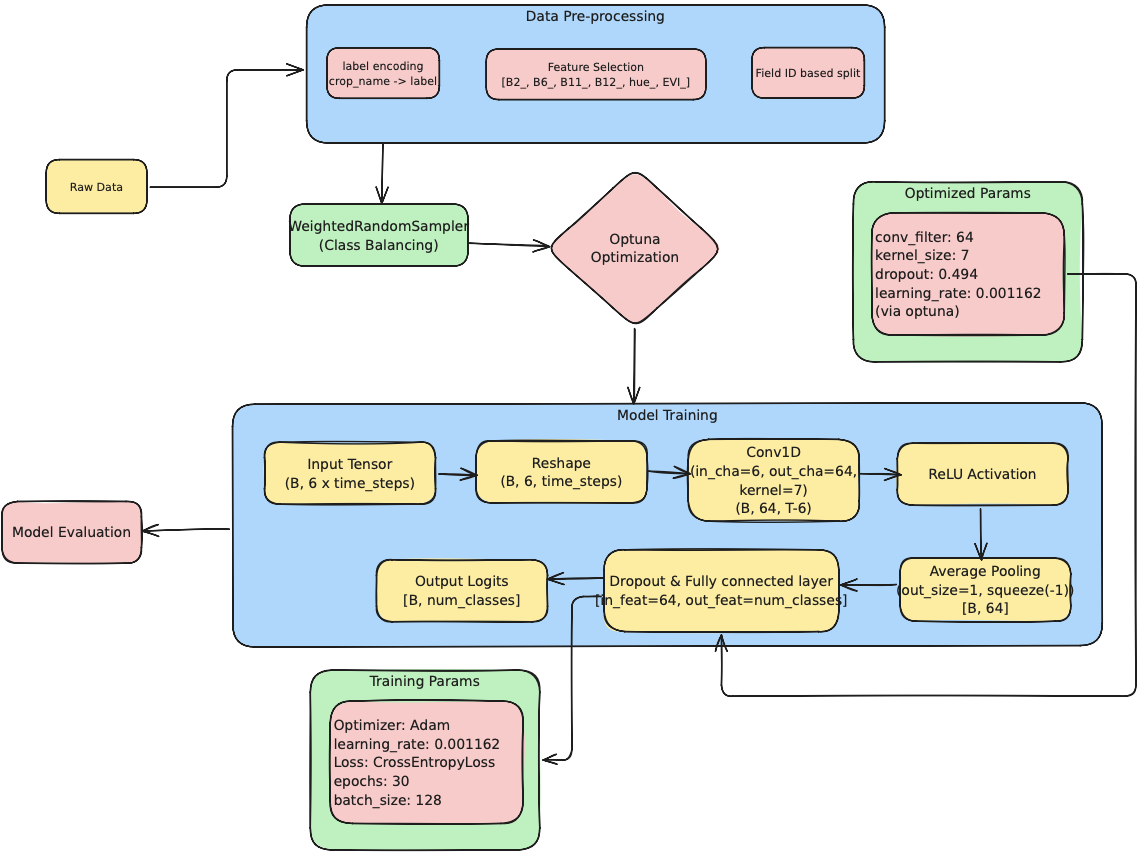


* + This ensures that the final output reflects the dominant class label per agricultural parcel, aligning predictions with the spatial resolution of ground truth annotations and real-world decision-making units.
  1. **Interpretable Deep Learning for Crop Monitoring:**
  + A standout advantage of TabNet lies in its interpretability. The sparse masks generated at each decision step can be analyzed to identify which features were most influential in making a prediction. This feature-level transparency is highly desirable in agricultural contexts where decision-makers need insight into why a particular crop label was assigned whether due to reflectance patterns in specific bands or seasonal vegetation indices.
  + Recent studies have validated the effectiveness of TabNet in remote sensing classification tasks, showing competitive accuracy and enhanced interpretability when compared to MLPs and ResNets in tabular crop classification .

1. **1D CNN with Optuna Hyperparameter Tuning:**

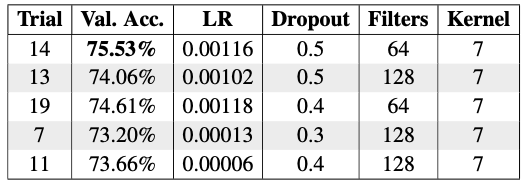
* To build a streamlined and computationally efficient baseline for crop classification, we implemented a 1D Convolutional Neural Network (1D CNN) trained on temporal satellite features. This model focuses on identifying subtle phenological patterns within multitemporal reflectance signals by applying convolutional filters across time. Despite being lightweight, the architecture leverages automated feature extraction and hyperparameter tuning to compete with deeper temporal models such as LSTMs and BiLSTMs .
* The pipeline begins with reshaping each input tensor to represent multichannel time-series sequences—six features (B2, B6, B11, B12, EVI, and hue) across monthly intervals. This representation allows the model to interpret the evolution of spectral characteristics throughout the crop lifecycle [9](#hxikosc2ufi6).
* As depicted in Figure 10, the model architecture consists of:
  + A 1D convolutional layer (nn.Conv1d) with a configurable kernel size and filter count
  + ReLU activation
  + Global average pooling to condense temporal features
  + Dropout to reduce overfitting
  + A fully connected classification head with softmax output
* The key operation in this model is the 1D convolution across time, formally defined as:



* Where  is the temporal input signal for a feature,  is the convolutional filter, and  is the time index. This allows the model to learn temporal dynamics such as crop emergence, peak greenness, and senescence from raw sequences without manual feature crafting.
*   
  *Fig 9.: Architecture of the 1D CNN model optimized via Optuna. The model uses a single convolutional layer followed by ReLU activation, adaptive pooling, and a fully connected classification head.*
  1. **Hyperparameter Optimization with Optuna**
  + Rather than relying on fixed architectural parameters, we used Optuna, an efficient, automated hyperparameter tuning framework. The goal was to identify the best combination of the following parameters:
    - Learning rate (log-uniform:  to )
    - Dropout rate (uniform: 0.2 to 0.5)
    - Kernel size 
    - Number of convolutional filters 
  + Each configuration (trial) was evaluated over 25 training epochs using a class-weighted sampler to address severe imbalance in crop classes. The validation accuracy guided Optuna’s pruning and selection mechanism. We formally define the hyperparameter optimization problem as:



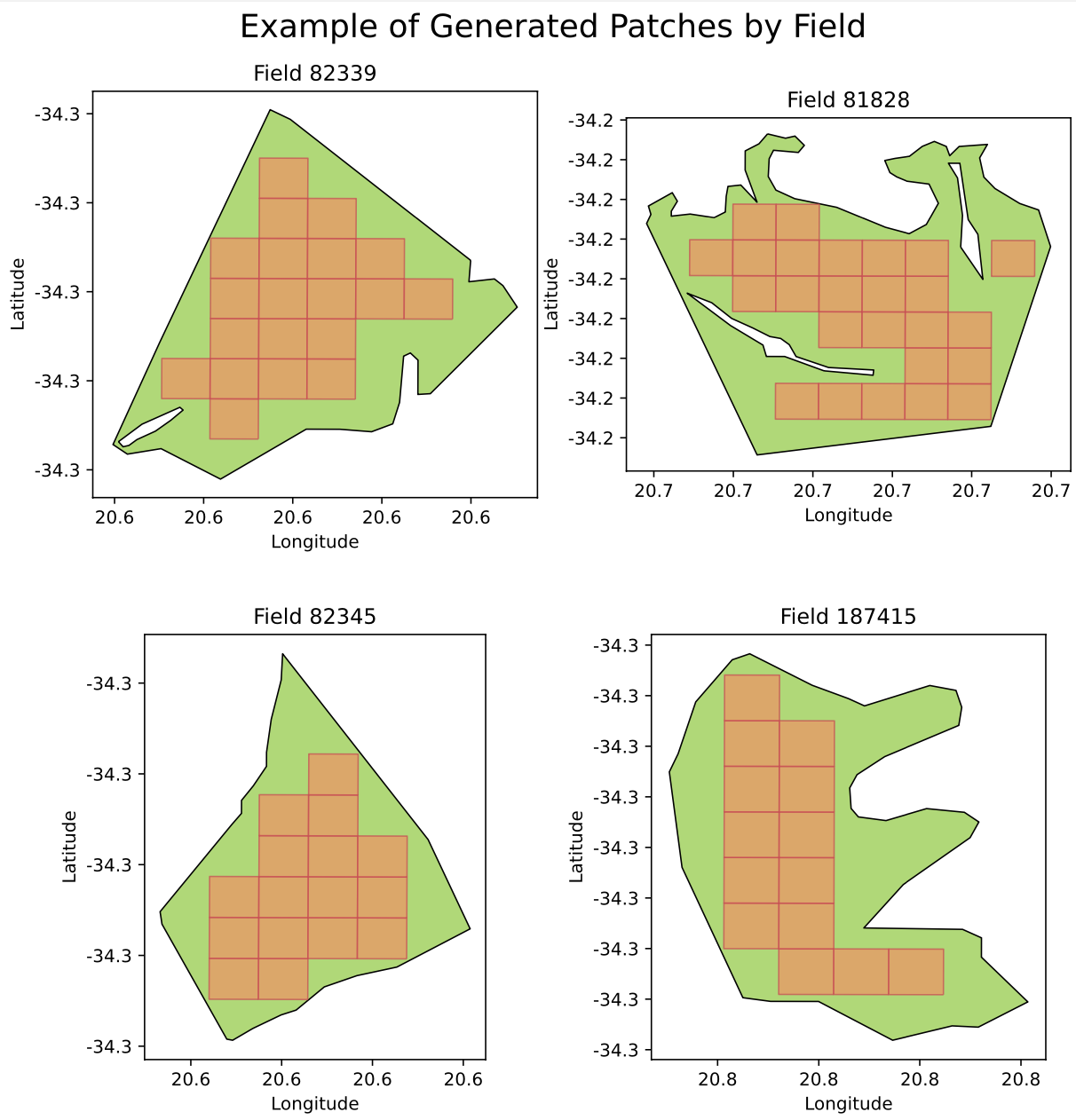
* + Where:
    -  is a set of hyperparameters: 
    -  is the search space defined above
    -  is the CNN model trained using configuration 
    -  is the validation accuracy after 25 epochs
  + This optimization problem was solved using the Tree-structured Parzen Estimator (TPE) algorithm implemented within the Optuna framework. This setup yielded the best validation accuracy of 75.5%, with robust generalization confirmed through field-level aggregation.
  1. **Why This Configuration Was Chosen:**
  + The selected trial balanced accuracy and convergence speed. The use of a larger kernel (7) helped capture wider temporal patterns, such as seasonal trends in vegetation indices. A moderate filter count (64) offered enough representational power without overfitting, and dropout  helped maintain generalization. Optuna dynamically explored a wide search space and converged on this optimal setting based on cross-validation performance.
  1. **Field-Level Aggregation and Evaluation:**
  + To align predictions with agricultural decision-making, we aggregated pixel-level outputs to the field-level using majority voting on field IDs. Evaluation metrics (accuracy, Cohen’s kappa, F1-score) were computed at both levels. A field-level confusion matrix was also generated for interpretability.



*Table 2.: Top Hyperparameter Configurations Based on Validation Accuracy*

### **Patch-Level Analysis**

To better capture spatial patterns and enable the use of matrix-based deep learning architectures, we transitioned to a patch-level analysis. We divide each field into a grid of fixed-size square patches (100x100 pixels), with each patch retaining the full 10-month spectral history for every pixel as shown in Figure [10](#60ltw1n9vusu).



*Fig 10.: Illustration showing how patches are generated for each field*

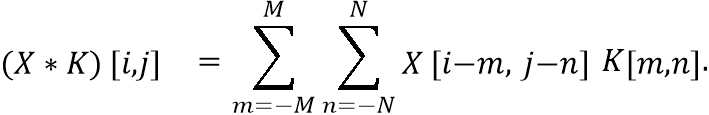
This approach offers several advantages:

* **Spatial Context:** Unlike pixel-level models, patches explicitly encode spatial relationships among neighboring pixels, capturing local patterns and textures characteristic of different crop types.
* **Data Augmentation:** By extracting multiple patches from a single field, we effectively augment the training dataset, reducing overfitting and improving generalization performance, particularly in data-scarce scenarios.
* **Compatibility with CNNs:** Patch-based inputs are naturally compatible with convolutional neural networks (CNNs), which are designed to learn hierarchical spatial features through convolution operations.
* **Multi-Scale Feature Extraction:** CNNs applied to patches can capture features at multiple scales, ranging from fine-grained textures within each patch to broader field-level patterns.

The trade-off involves increased computational complexity due to processing a larger spatial matrices and omission of boundary pixels.

1. **Convolution for feature generation:**

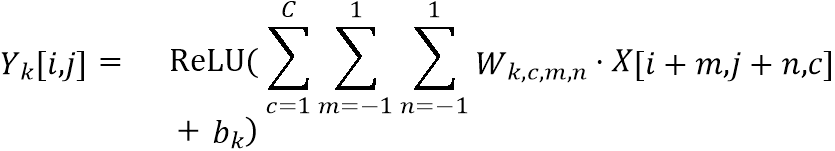
* Central to our patch-level architectures is the 2D convolution operation, which applies a small learnable kernel across an input image. Formally, for input feature map  and kernel  of size , we write:



* Convolutions exploit local connectivity, weight sharing, and translation equivariance, making them ideal for learning hierarchical spatial features in satellite patches.

1. **Patch-Based Architectures:**

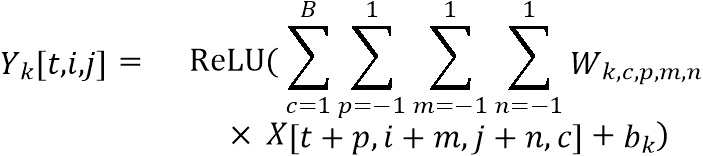
* We investigated four distinct patch-based architectures for crop classification, each leveraging different strategies to exploit the spectral, spatial, and temporal information present in remote sensing data.  
  1. **Multi-Channel 2D CNN:**  
     This approach treats each spectral band and monthly observation as a separate channel, constructing a unified spectral-temporal profile for each patch. Let  be the input patch (, ). Each Conv2D block with  filters computes:



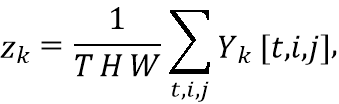
* + where , . After two such blocks and flattening to , class scores are:

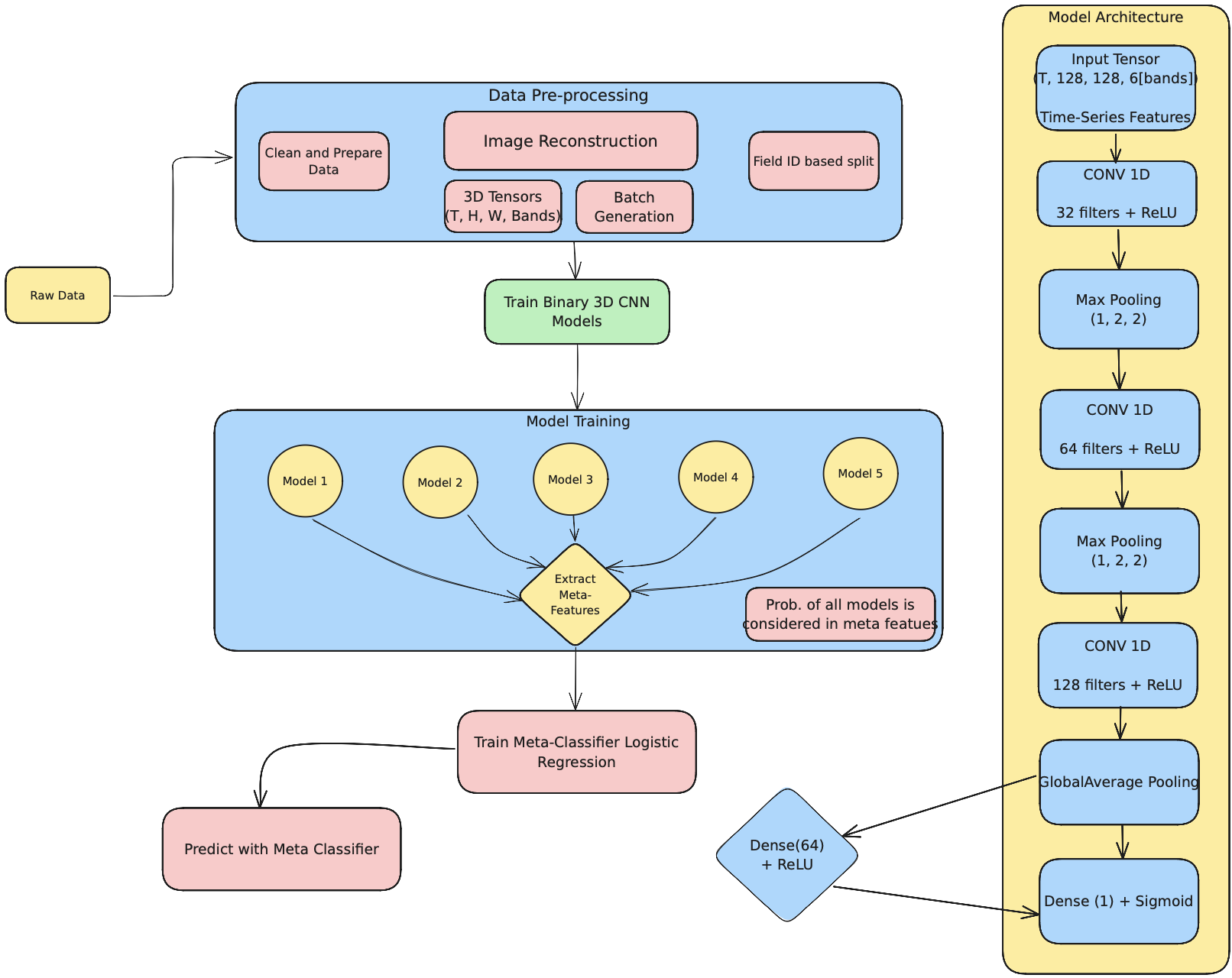


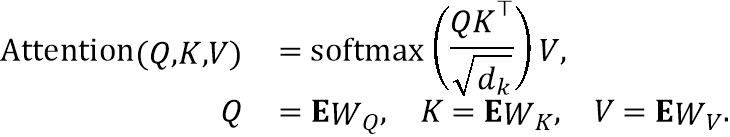
* + This simple yet effective architecture is widely used as a benchmark in remote sensing, as 2D CNNs have demonstrated robust performance in extracting spatial-spectral features from satellite imagery .
  1. **3D CNN:**  
     The 3D CNN architecture processes the spectral and temporal dimensions jointly, enabling the model to capture temporal changes and phenological patterns critical for accurate crop classification. Reshape into  with , . A Conv3D block does:



* + with . After three Conv3D–Pool3D layers, global average pooling yields:



* + and the final softmax is as above.
  + 
  + *Fig 11.: This architecture* [*11*](#s7wbh9shurcj) *provided a clear improvement in performance, as 3D CNNs are particularly effective at capturing both spatial and temporal dependencies in remote sensing data* .
  1. **Transformer-Based Model:**  
     In this approach, each patch is flattened into  tokens  and embedded to . Each multi-head attention layer computes:



* + A final ‘[CLS]‘ embedding  is classified by:



* + Transformer-based architectures have shown promise in remote sensing by capturing long-range dependencies and global context .
  1. **Ensemble of Class-Specific 3D CNNs:**  
     This ensemble approach combines the outputs of five class-specific 3D CNN experts , each outputting a class probability . A logistic regression meta-classifier then combines them:



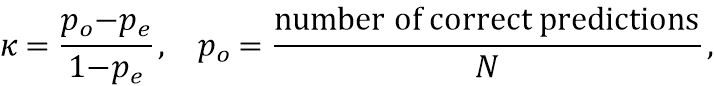
* + This strategy leverages the strengths of specialist models while retaining the robustness of ensemble learning, resulting in the highest performance among patch-level approaches .

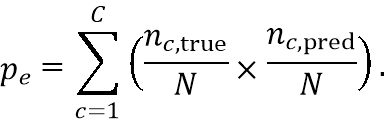
# **Results and Discussion**

This section presents the main findings of our study, highlighting key performance metrics for each modeling approach. All models were evaluated on an independent, held-out test set to ensure their ability to generalize to unseen data.

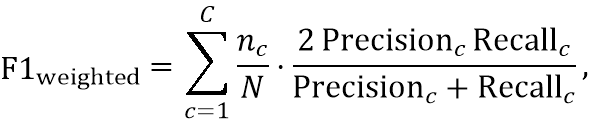
To account for the highly imbalanced class distribution in our data, we report three metrics that are more informative than overall accuracy:

* **Cohen’s Kappa (Kappa)** Measures agreement between predicted and true labels, corrected for chance:

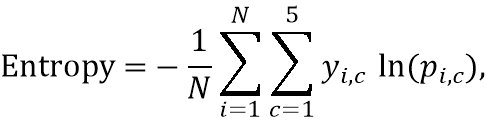




* **Weighted F1 Score (F1)** The harmonic mean of precision and recall, weighted by the support of each class:



* where  is the number of true instances of class .
* **Entropy (Log Loss)** Quantifies the average surprise of the predicted probabilities (adapted from the standard cross‐entropy formula):



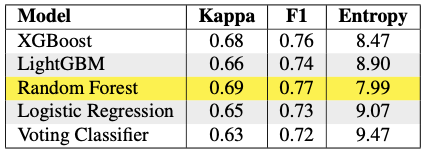
* where  is the one‐hot true label for sample  in class ,  is the model’s predicted probability for that class, and there are 5 crop categories.

We avoid reporting plain accuracy here because an imbalanced test set can yield deceptively high accuracy simply by predicting the majority class. In contrast, Cohen’s Kappa and the weighted F1 score reward correct classification across all classes and higher values indicate better agreement and balanced precision/recall, while entropy (log loss) measures prediction uncertainty, so lower values indicate more confident, accurate probability estimates.

## **Classical Machine Learning**

### **Pixel‑Level Analysis**

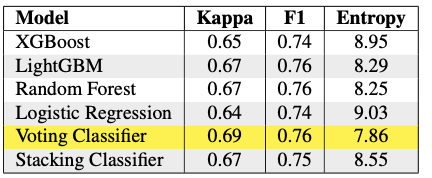
Table 2 summarizes the out-of-sample performance of classical machine learning models at the pixel level. Random Forest achieves the highest performance, with a Cohen’s Kappa of 0.69 and an F1 score of 0.77, indicating substantial agreement beyond chance. Figure 12 shows the confusion matrix of the pixel-level random forest model.

**

*Table 3.: Out-of-sample performance for Classical ML Pixel-Level Analysis*

**Field‑Level Analysis**

Table 3 presents the field-level results for the classical models and ensemble approaches. The Voting Classifier achieves the best performance, with a Cohen’s Kappa of 0.69 and an F1 score of 0.76, highlighting the advantage of ensemble methods over individual classifiers.

**

*Table 4.: Out-of-sample performance across Classical ML models at Pixel and Field levels*

Combining model predictions at the field level leads to improvements in classification reliability.

Figures 13 and 14 present the confusion matrices for the Stacking and Voting ensemble classifiers evaluated at the field level. Both models demonstrate strong diagonal dominance, indicating high agreement between predicted and true crop classes.

|  |
| --- |

*Fig 12.: Random Forest(Pixel Level) Confusion Matrix*

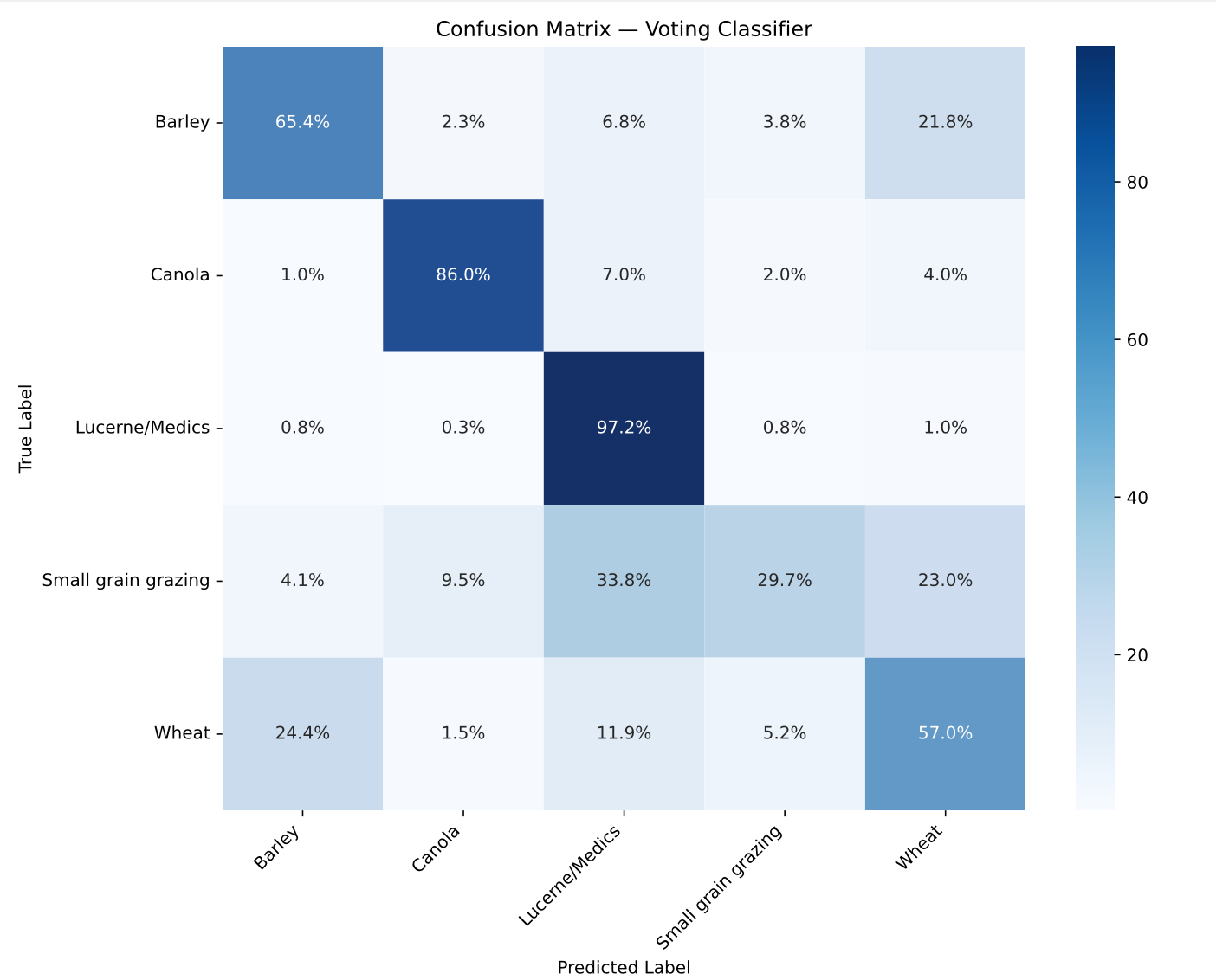
For the Stacking Ensemble (Figure 13), Lucerne/Medics achieves the highest classification accuracy, with over 94% of fields correctly identified. Wheat and Canola, however, exhibit some misclassification, particularly as Lucerne/Medics and Barley, which suggests that these classes share spectral-temporal similarities during certain months. The model performs poorly on Small Grain Grazing, with notable confusion spread across Barley and Wheat, indicating challenges in distinguishing these crops based solely on their seasonal satellite signatures.

|  |
| --- |

*Fig 13.: Confusion matrix for Stacking Ensemble at field level.*

In the case of the Voting Ensemble (Figure 14), the Cohen’s Kappa and F1 score improves, with Lucerne/Medics again achieving the best performance at over 97%. Wheat is classified more accurately than in the stacking approach, indicating enhanced stability resulting from ensemble consensus. Although minor crops such as Barley and Small Grain Grazing continue to show moderate confusion, their misclassification rates are notably lower compared to the stacking strategy, highlighting the robustness of the voting-based aggregation.

These confusion matrices highlight that both ensemble methods effectively consolidate pixel-level predictions into robust field-level outputs. The Voting Ensemble offers better generalization, particularly for dominant crops, while the Stacking Ensemble provides competitive performance across all classes.



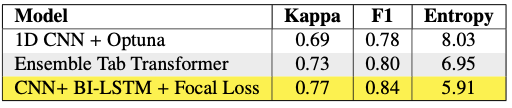
*Fig 14.: Confusion matrix for Voting Ensemble at field level.*

## **Deep Learning**

### **Pixel‑Level Analysis**

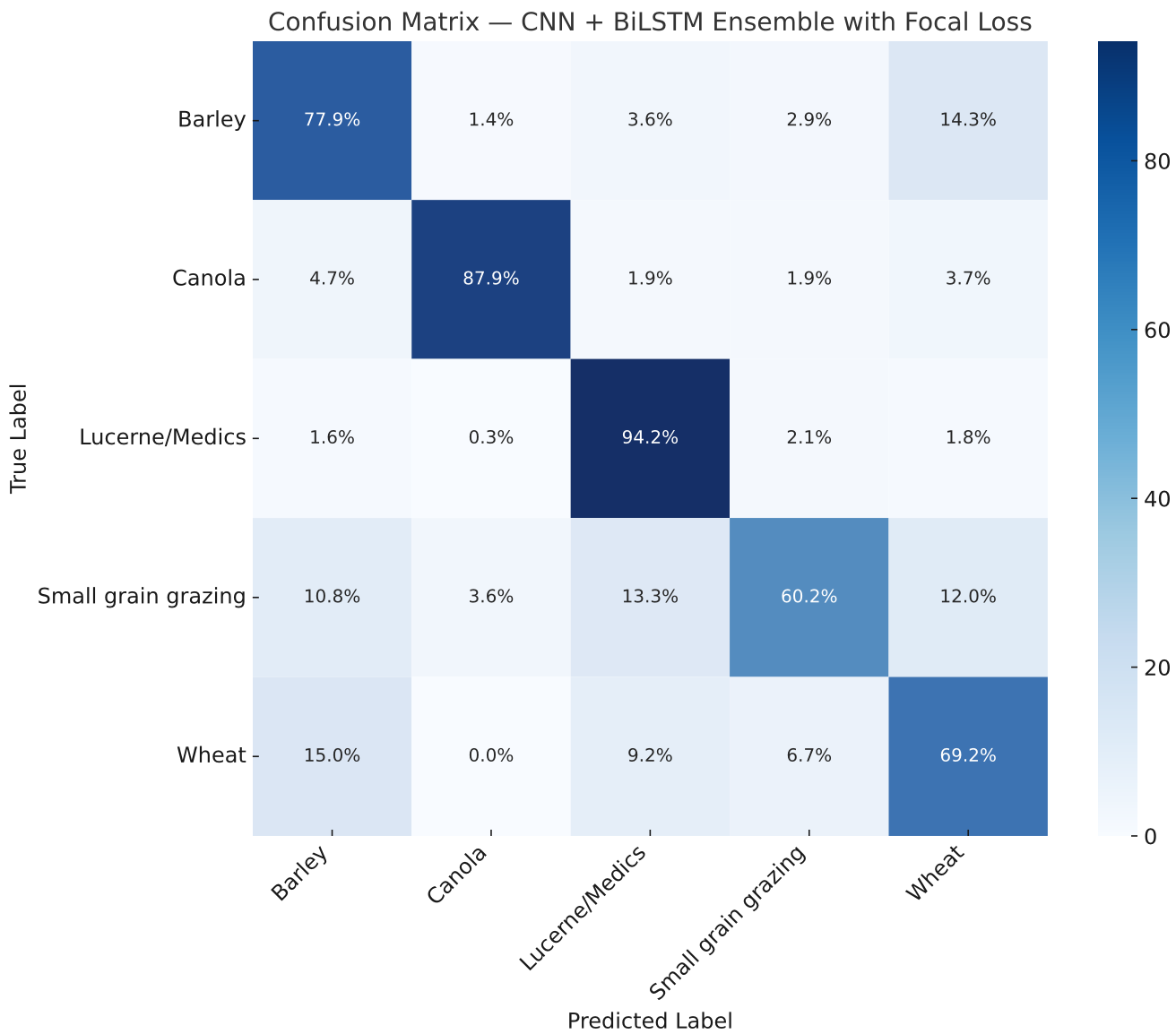
We evaluated all three models—CNN + BiLSTM + Focal Loss, Ensemble TabTransformer, and 1D CNN with Optuna—on a held-out test set.

* **Performance Comparison:**
* The results highlight the benefits of temporal modeling and imbalance-aware training for crop classification. The CNN+BiLSTM+Focal Loss model[[15](#6n75eu94b9wu)] achieved the highest field-level performance, with a Cohen’s Kappa of 0.77 and an F1 score of 0.84, demonstrating superior ability to capture vegetation dynamics and handle class imbalance.



*Table 5.: Comparison of test‐set performance across deep learning models at both pixel and field level.*

The Ensemble TabTransformer achieved a field-level Cohen’s Kappa of 0.73 and an F1 score of 0.80, highlighting the strength of attentive feature selection in high-dimensional crop data . Despite lacking explicit temporal modeling, its sequential decision process effectively captured latent growth patterns. The 1D CNN, while simpler, still delivered competitive results with a Cohen’s Kappa of 0.69 and an F1 score of 0.78, benefiting from Optuna-based hyperparameter optimization[[1](#eduy24j2taqu)]. This supports prior observations  that shallow Conv1D architectures, when carefully tuned, can perform strongly for vegetation classification.



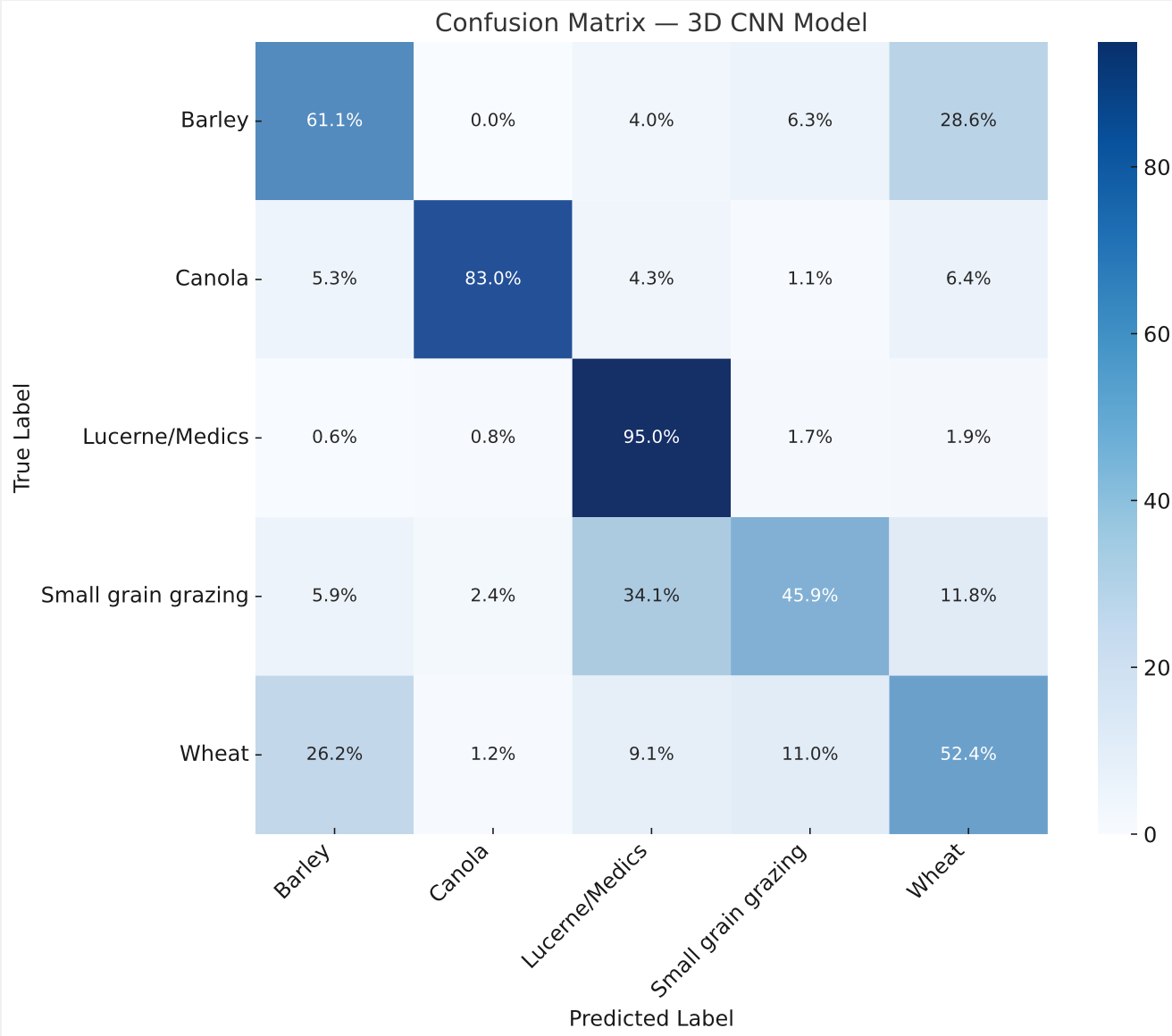
*Fig 15.: Field-level confusion matrix for CNN + BiLSTM ensemble model showing predicted vs. actual crop classes.*

### **Patch‑Level Analysis**

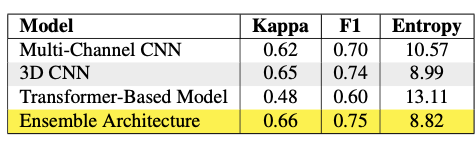
We evaluated four patch-based deep learning models on spatially coherent image regions extracted from Sentinel-2 tiles. The classification was conducted using aggregated patch-level information rather than individual pixels, preserving spatial context and allowing models to exploit neighboring structures.

* **Performance Comparison:**

The results suggest that modeling local spatial coherence through patch-level inputs can improve crop classification performance, particularly when ensemble learning strategies are employed. As shown in Table [5](#18f1zzf0au47), the Ensemble Architecture achieved the highest patch-level performance, with a Cohen’s Kappa of 0.66 and an F1 score of 0.75. This ensemble combined outputs from multiple base learners, including CNNs and transformer components, enabling a more generalized understanding of crop field characteristics Figure [16](#id2k75azkjpg).  
Among the individual models, the 3D CNN performed strongly with a Cohen’s Kappa of 0.65 and an F1 score of 0.74. By capturing volumetric relationships across spectral bands and spatial dimensions, the 3D CNN demonstrated an advantage over traditional 2D methods. The Multi-Channel CNN also achieved competitive results, reaching a Kappa of 0.62 and an F1 score of 0.70. These results highlight the effectiveness of band-specific spatial convolution for capturing vegetation structure and field texture.



* *Fig 16.: Confusion matrix for 3D CNN Model*



*Table 6.:* Out‐of‐sample performance for Deep Learning Patch‐Level Analysis

In contrast, the Transformer-Based Model underperformed, achieving a Cohen’s Kappa of 0.48 and an F1 score of 0.60. This suggests that pure attention-based architectures, without strong inductive biases like spatial locality, may struggle with small patch classification unless extensively pretrained or domain-adapted to remote sensing imagery.

# **Discussion**

Our experiments demonstrate that both classical and deep learning approaches can effectively leverage Sentinel-2 imagery for field-level crop classification, yet each paradigm has distinct strengths and limitations. Deep models, particularly the CNN + BiLSTM ensemble trained with Focal Loss, excel at capturing seasonal dynamics and mitigating class imbalance, achieving an 0.77 Cohen’s Kappa score at the field-level and an F1 score of 0.84. However, they require extensive hyperparameter tuning, and they remain sensitive to mixed pixels along field edges due to Sentinel-2’s 10–20 m resolution.

Classical machine learning approaches remain highly competitive: a pixel‐level Random Forest trained on xr\_fresh-generated time‐series features generalizes well to unseen data, and a field‐level voting classifier that aggregates those pixel predictions outperforms several deep learning architectures on the held-out test set. By combining SMOTETomek resampling with ensemble methods (e.g., Random Forest, LightGBM, XGBoost) and automated temporal feature engineering, these models achieve comparable or better accuracy with substantially lower computational cost and greater interpretability. However, because they rely primarily on temporal summaries, they may underutilize fine‐scale spatial textures unless supplemented with explicit textural indices.

Our patch-level analysis incorporates spatial context while matching the performance of classical models based on time-series features. Dividing each field into fixed-size patches enables 3D CNNs and patch-based ensembles to learn local texture and neighborhood patterns, boosting patch-level Cohen’s Kappa to 0.66. This strategy also provides implicit data augmentation by generating multiple overlapping patches per field, which increases sample diversity and reduces overfitting. Because patches are placed on a fixed grid, however, some pixels along field boundaries are left out. Additionally, processing dozens of patches per field significantly increases computational cost and memory requirements.

Looking ahead, several avenues could further improve performance across paradigms:

* **Higher resolution or increased revisit frequency.** PlanetScope or UAV imagery and biweekly acquisitions would reduce edge effects and better capture rapid phenological events. Using all Sentinel 2 observations rather than monthly composites could also achieve the same.
* **Advanced augmentation.** Spectral band mixing, elastic deformations, or GAN-based patch generation could bolster deep model robustness, particularly for underrepresented crops.
* **Dynamic field delineation and ancillary data.** Change-detection algorithms, soil or elevation maps, and in-situ surveys would strengthen robustness in heterogeneous landscapes.
* **Hybrid workflows.** Combining time-series features with patch-based representation learning may yield a best-of-both-worlds solution—balancing interpretability, spatial detail, and temporal dynamics.

# **Conclusion**

This study demonstrates the complementary strengths of deep learning, classical machine learning, and patch-based approaches for multitemporal crop classification using Sentinel-2 data. Deep learning models, particularly the CNN + BiLSTM ensemble trained with Focal Loss—lead the pack, achieving an F1 score of 0.84 and a Cohen’s Kappa of 0.77 at the field level. Patch-based methods further enrich spatial context: our 3D CNN patch ensemble reached a patch-level Cohen’s Kappa of 0.66 through implicit data augmentation and local texture modeling, at the expense of higher compute demands and exclusion of boundary pixels.

Meanwhile, classical machine learning models such as LightGBM and XGBoost combined with SMOTETomek resampling and ensemble voting on xr\_fresh-generated temporal features, deliver competitive field-level performance with far less tuning and far greater interpretability. These approaches remain appealing for operational deployments where transparency and resource efficiency are paramount.

To summarize, deep and patch-based paradigms excel at automatically extracting rich spatiotemporal and local features, while classical methods offer reliable, interpretable alternatives. The choice of method should weigh deployment constraints, accuracy targets, spatial resolution, interpretability, and computational budget. Hybrid workflows that fuse temporal feature engineering, spatial representation learning, and higher-resolution imagery hold the greatest promise for scalable, accurate, and equitable agricultural monitoring.

# References

Yunxiang Zhang, Yuxin Zhang, Jun Zhang, Fulin Wang, and Xinyu Wang. (2023). A Spatio-Temporal Feature Fusion Network for Crop Classification Using Multi-Temporal Sentinel-2 Data. *ISPRS International Journal of Geo-Information*, 12(11), 450.[[Article]](https://www.mdpi.com/2220-9964/12/11/450)

S. Ali, M. Usama, M. Usman, M. Rizwan, and A. Rehman, “A hybrid deep learning model for crop classification using Sentinel-2 time series data,” *Comput. Electron. Agric.*, vol. 206, p. 107673, 2023. [Online]. Available: [[Article]](https://www.sciencedirect.com/science/article/pii/S0168169923000303)

A. Ienco, R. Gaetano, C. Dupaquier, and D. H. T. Minh, “Land cover classification via multitemporal spatial data by deep recurrent neural networks,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2142–2154, 2017. [Online]. Available: [[Article]](https://ieeexplore.ieee.org/document/7891032)

Y. Li, Y. Zhang, and Z. Wei, “Multi-Temporal Crop Classification via Transformer-Based Deep Learning Model,” arXiv preprint arXiv:2402.02121, 2024. [Online]. Available: [[Article]](https://arxiv.org/abs/2402.02121)

R. Wang, Y. Wang, Z. Shao, and Y. Zhang, “Deep learning for crop classification in remote sensing images: A review,” *ISPRS J. Photogramm. Remote Sens.*, vol. 186, pp. 63–77, 2021. [Online]. Available: [[Article]](https://www.sciencedirect.com/science/article/pii/S0924271623000679)

Michael L. Mann, Lisa Colson, Rory Nealon, Ryan Engstrom, and Stellamaris Nakacwa. (2023). Lite Learning: Efficient Crop Classification in Tanzania Using Traditional Machine Learning & Crowd Sourcing. SSRN Preprint. [[Article]](https://ssrn.com/abstract=5090897)

A. Chauhan, “Crop Classification via Satellite Image Time Series and PSETAE Deep Learning Model,” Medium, 2023. [Online]. Available: [[Article]](https://medium.com/geoai/crop-classification-via-satellite-image-time-series-and-psetae-deep-learning-model-c685bfb52ce)

A. K. Rangarajan, R. Purushothaman, M. Prabhakar, and C. Szczepański, “Crop identification and disease classification using traditional machine learning and deep learning approaches,” *J. Eng. Res.*, vol. 11, no. 1B, pp. 1–12, 2023. [Online]. Available: [[Article]](https://kuwaitjournals.org/jer/index.php/JER/article/view/11941)

K. K. Gadiraju, B. Ramachandra, Z. Chen, and R. R. Vatsavai, “Weakly supervised deep learning for rapid crop cover mapping with multi-temporal satellite imagery,” in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2020. [Online]. Available: [[Article]](https://dl.acm.org/doi/abs/10.1145/3394486.3403375)

X. Liu et al., “Deep learning for crop classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-14, 2022.

Y. Zhang et al., “Transformer-based crop classification,” *ISPRS Journal of Photogrammetry*, vol. 195, pp. 200-215, 2023.

NASA/IBM, “Prithvi-100M technical report,” Tech. Rep., 2023.

Z. Khan et al., “Ensemble methods for precision agriculture,” *Computers and Electronics in Agriculture*, vol. 216, p. 108495, 2024.

L. Wang et al., “Multi-temporal crop classification,” *Agricultural Systems*, vol. 203, p. 103517, 2022.

M. Rußwurm and M. Körner, “Temporal CNN for crop analysis,” *Remote Sensing*, vol. 10, no. 12, p. 1930, 2018.

Radiant Earth Spot The Crop Challenge [[Github]](https://github.com/radiantearth/spot-the-crop-challenge)

James B. Campbell and Randolph H. Wynne. (2011). *Introduction to Remote Sensing*. The Guilford Press.

Michael L. Mann. (2024). *mmann1123/xr\_fresh: SpeedySeries (0.2.0)*. Zenodo. [[Software]](https://doi.org/10.5281/zenodo.12701466)

B. D. Wardlow, S. N. Melendez, and C. M. Justice, “Towards an operational system for cropland monitoring using MODIS time series data,” *Remote Sens. Environ.*, vol. 110, no. 3, pp. 349–361, 2007.

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. (2013). *An Introduction to Statistical Learning*. Springer.

Leo Breiman. (2001). Random forests. *Machine Learning*, 45(1), 5-32.

Gérard Biau. (2008). Consistency of random forests. *The Annals of Statistics*, 36(4), 2032-2068.

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30.

Tianqi Chen and Carlos Guestrin. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794).

Jerome H. Friedman. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29(5), 1189-1232.

Thomas G. Dietterich. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning*, 40(2), 139-157.

Aaron E. Maxwell, Timothy A. Warner, and Fang Fang. (2019). Implementation of machine-learning classification in remote sensing: an overview. *Remote Sensing*, 11(14), 1633.

Xiao Xiang Zhu, Devis Tuia, Lorenzo Mou, Gui-Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8-36.

Rui Li, Shuang Zheng, Chao Zhang, and Chenguang Huang. (2019). Object detection in remote sensing images based on improved Faster R-CNN. *IEEE Access*, 7, 35813-35822.

I. Ijabs, “Artificial Intelligence and Democracy: A Political-Theoretical Perspective,” *Baltic Journal of Modern Computing*, vol. 8, no. 4, pp. 607–616, 2020. [Online]. Available: [[Article]](https://www.bjmc.lu.lv/fileadmin/user_upload/lu_portal/projekti/bjmc/Contents/10_4_02_Ijabs.pdf)

J. A. et al., “Deep Learning Models for the Classification of Crops in Aerial Imagery: A Review,” *Remote Sensing Reviews*, vol. 46, no. 1, pp. 1–24, 2023. [Online]. Available: [[Article]](https://www.sciencedirect.com/science/article/pii/S0168169923000303)

N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, “Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data,” *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 778–782, May 2017. doi: 10.1109/LGRS.2017.2681128. [Online]. Available: [[Article]](https://ieeexplore.ieee.org/document/7890411)

A. K. et al., “A Deep Learning Approach for Dealing with Tabular Data in Crop Classification,” *IEEE Access*, 2024. [Online]. Available: [[Article]](https://ieeexplore.ieee.org/abstract/document/10826760)

“Deep learning based multi-temporal crop classification,” *Remote Sensing of Environment*, vol. 221, pp. 175–190, 2019. [Online]. Available: [[Article]](https://www.sciencedirect.com/science/article/abs/pii/S0034425718305418)

Ryan M. Rustowicz, A. Singh, M., and A. Davis. (2019). Semantic segmentation of crop type in African smallholder agriculture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 75-82).

Shaowu Ji, Chenglong Shen, Yuxiang Zhu, Shuang Zhang, and Yanfeng Lu. (2018). 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sensing*, 10(1), 75.

Y. Zhao, W. Li, M. Li, and J. Wang, “Cropformer: A new generalized deep learning classification framework for multi-scenario crop classification,” *Front. Plant Sci.*, vol. 14, p. 1130659, 2023. [Online]. Available: [[Article]](https://www.frontiersin.org/articles/10.3389/fpls.2023.1130659/full)

M. O. Turkoglu, S. D’Aronco, G. Perich, F. Liebisch, C. Streit, K. Schindler, and J. D. Wegner, “Crop Type Classification with Satellite Imagery Using Deep Learning,” *Remote Sens. Environ.*, vol. 256, p. 112331, 2021. [Online]. Available: [[Article]](https://www.sciencedirect.com/science/article/pii/S0034425721003230)

“Crop Classification via Satellite Image Time Series and PSETAE Deep Learning Model,” Medium. [Online]. Available: [[Article]](https://medium.com/geoai/crop-classification-via-satellite-image-time-series-and-psetae-deep-learning-model-c685bfb52ce)

“Crop Classification with Satellite Data,” *MDPI Proceedings*, vol. 82, no. 1, p. 95, 2024. [Online]. Available: [[Article]](https://www.mdpi.com/2673-4591/82/1/95)

“Crop Monitoring with Deep Learning on Satellite Data,” in *ACM KDD*, 2020. [Online]. Available: [[Article]](https://dl.acm.org/doi/abs/10.1145/3394486.3403375)

“Crop Type Classification with Satellite Imagery Using Deep Learning,” *Remote Sensing of Environment*, 2021. [Online]. Available: [[Article]](https://www.sciencedirect.com/science/article/pii/S0034425721003230)

“Crop Classification using Machine Learning and Remote Sensing,” *Journal of Engineering Research*, 2022. [Online]. Available: [[Article]](https://kuwaitjournals.org/jer/index.php/JER/article/view/11941)