Spatial Generalization of Crop Yield Prediction Models Using UAV Imagery and Machine Learning

by

Wilfried Sankara

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements for the
Master's degree in Computer Science
with a concentration in Applied Artificial Intelligence

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

Abstract

Global population is expected reach 9.7 Billions by 2050 leading to an increase in food demand significantly, placing then a pressure on agricultural systems to improve productivity in a sustainable manner. Precision agriculture, supported by remote sensing and machine learning (ML), has emerged as a promising tool for optimizing resource use and improving crop yield prediction. Among most popular remote sensing platforms, Unmanned Aerial Vehicles (UAVs) has been used widely due to its ability to capture high-resolution images. Despite encouraging results reported in prior studies, the ability of UAV-based ML models to generalize across different spatial contexts remains understudied.

This thesis investigates the spatial generalization performance of ML models for crop yield prediction using UAV multispectral orthomosaics and yield monitor data collected from three agricultural fields—two canola fields and one corn field—located in Manitoba and Alberta, Canada. Linear Regression and Random Forest models are evaluated using a tile-based representation (5 m $\times$ 5 m) derived from five spectral bands and two vegetation indices, namely the Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Red Edge (NDRE).

Model performance were assessed under multiple training–testing configurations, including intra-field and inter-field baselines, Leave-One-Field-Out (LOFO) evaluation, and spatially aware sampling approaches such as Leave-One-Block-Out (LOBO) and Leave-One-Cluster-Out (LOCO). Results indicate that models trained on combined raw spectral bands consistently outperform those relying only on vegetation indices,

suggesting that richer spectral information improves predictive performance. However, cross-field generalization remains limited, highlighting the challenges of transferring models across heterogeneous spatial environments. Performance variability between fields further emphasizes the importance of consistent data collection and evaluation protocols.

Acknowledgments

I would like to express my sincere gratitude to all those who have supported me throughout my thesis journey. I am deeply thankful to my supervisors, Professors Iluju Kiringa and Tet Yeap, for their guidance and precious feedback. I would also like to extend my special thanks to Professor Azzedine Boukerche for his financial support during this journey.

I also extend my gratitude to all my fellow graduate students from the University of Ottawa Smart IoT Lab, particularly Dr Patrick Killeen, whose generous technical help was essential to the happy outcome of my journey. Special thanks to the University of Ottawa for providing the necessary resources and infrastructure.

I would especially like to thank my family — my father, mother, and all my siblings — for their trust, sacrifices, and unwavering support. I am also grateful to my close friends and my church who stood by me during this process.

Finally, I would like to thank God, who inspired and guided me through the Holy Spirit throughout this thesis process.

Contents

List of Figures

List of Tables

List of Acronyms

Acronym	Definition
ML	Machine Learning
UAV	Unmanned Aerial Vehicle
NDVI	Normalized Difference Vegetation Index
NDRE	Normalized Difference Red Edge
EVI	Enhanced Vegetation Index
SAVI	Soil-Adjusted Vegetation Index
VI	Vegetation Index
PA	Precision Agriculture
RGB	Red-Green-Blue
MS	Multispectral
NIR	Near-Infrared
GIS	Geographic Information System
GPS	Global Positioning System
CNN	Convolutional Neural Network
LR	Linear Regression
RF	Random Forest
OLS	Ordinary Least Squares

SVM	Support Vector Machine
SVC	Support Vector Classification
SVR	Support Vector Regression
CV	Cross-Validation
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
R^2	Coefficient of Determination
r	Pearson's Correlation Coefficient
LOFO	Leave-One-Field-Out
LOBO	Leave-One-Block-Out
LOCO	Leave-One-Cluster-Out
Man Eds	Manitoba Ed's Field
Man Home	Manitoba Home Field
Alb PP	Alberta Precision Planter Field

Chapter 1

Introduction

The world's population is projected to exceed 9.7 billion by 2050 [103], raising a critical challenge to global food security. Global food systems must produce more food with fewer resources, while facing land degradation and climate change. There is an urgent need to rethink agricultural production systems in order to address this issue.

In response, precision agriculture (PA) has emerged as a promising approach to optimize resource use and improve crop productivity. Precision agriculture (PA) contributes significantly to sustainable food production by helping farmers monitor crops and make informed decisions. The goal is not only to maximize yield but also to optimize the use of resources such as water and fertilizer while minimizing environmental impacts [74, 90].

Recent technological advances—particularly the use of UAVs—have accelerated this transformation. UAVs capture high-resolution imagery that bridges the gap between satellite observations and ground-based measurements, offering a flexible and detailed view of field conditions. When associated with machine learning (ML) models [81], UAV imagery can reveal spatial and spectral patterns associated with crop health and yield. Among its various applications, crop yield prediction has become especially valuable, supporting proactive planning for harvesting, logistics, and market

strategies [90].

1.1 Motivation and Problem Statement

Despite these advantages, the generalization ability of UAV-based yield prediction models, especially their ability to maintain accuracy when applied to new spatial contexts, remains underexplored. Yield monitor data collected by combine harvesters offer precise ground truth but are inherently noisy and spatially heterogeneous [81]. Even within a single field, distinct zones of low and high yield are common. Moreover, environmental and management conditions vary considerably between fields. This raises two fundamental questions:

1. Can a model that performs well on one field transfer to another, with either similar or different crop types?
2. Can a model trained on one part of a field accurately predict yield in other, unseen parts of the same field?

While many studies have demonstrated promising results in UAV-based crop yield prediction using machine learning (ML) [73, 62, 9], most have not incorporated spatially aware validation strategies or explicitly examined the effect of spatial sampling on model performance. Several existing studies rely on random train–test splits, which ignore spatial autocorrelation. Recent work has begun to address this gap: [55] introduced spatial constraints in their evaluation and found that conventional non-spatial methodologies often produce overly optimistic performance estimates. However, systematic investigation across multiple spatial contexts and evaluation strategies remains limited.

This thesis aims to clarify how spatial heterogeneity, image spectral information, and evaluation strategies affect the performance of machine learning (ML) models for UAV-based yield prediction.

1.2 Research Objectives and Approach

The objectives of this study are to address the challenges outlined earlier by conducting an evaluation of widely used machine learning (ML) models under various spatial contexts. Specifically, the study investigates how linear regression (LR) and random forest (RF) models perform when trained and tested on data drawn from different fields and from distinct spatial regions within a single field.

This work aims to assess the impact of spatial autocorrelation on model performance and transferability, as well as the role that spectral information—including raw spectral bands and VIs—plays in spatially aware experiments.

The analysis uses UAV multispectral orthomosaics and yield-monitor data collected from three agricultural fields in Manitoba and Alberta, Canada (two canola fields and one corn field). Each field is divided into 5 m $\times$ 5 m tiles, from which five spectral bands (Blue, Green, Red, Red Edge, and NIR) and two VIs (NDVI, NDRE) are extracted as predictive features.

The models are evaluated under four main training–testing configurations:

- **Intra-field baseline:** standard train–test split within a single field;
- **Inter-field baseline:** standard train–test split applied to combined datasets from multiple fields;
- **Leave-One-Field-Out (LOFO):** training on all fields except one and testing on the remaining field;
- **Spatially aware strategies:** LOBO and LOCO, which explicitly account for the spatial structure within fields.

1.3 Thesis Outline

- Chapter 2: Background and related work, covering smart farming concepts.

- Chapter 3: Methodology for yield prediction, including data collection, preprocessing, and model development.
- Chapter 4: Experimental results, analysis, and evaluation of the proposed models.
- Chapter 5: Conclusion, discussion, and directions for future work.

Chapter 2

Background and Related Work

2.1 Yield prediction

2.1.1 Precision Agriculture

Precision agriculture (PA) refers to the use of advanced technologies with the goal of improving farmer productivity.[74] describes the PA cycle in the following steps: data collection, data interpretation, and application. This process is illustrated in Figure 2.1. Sensor technology plays a crucial role in collecting and analyzing relevant data, which is one of the key components of precision agriculture [90]. The data, as detailed by [108], contains various aspects, such as crop yield, soil composition, moisture levels and so on.

To implement precision agriculture, several types of data are necessary. Among the most crucial is imagery data, which can be obtained through satellite or drone-based methods. These imagery datasets can be combined with yield data, such as yield quantity and moisture content, as well as field-specific data like field boundaries and GPS coordinates. This combination of data can be helpful for tasks like yield prediction, crop monitoring, crop detection and so on. Data from imagery, sensors, and yield reports must be cleaned and processed before use. Raw data frequently

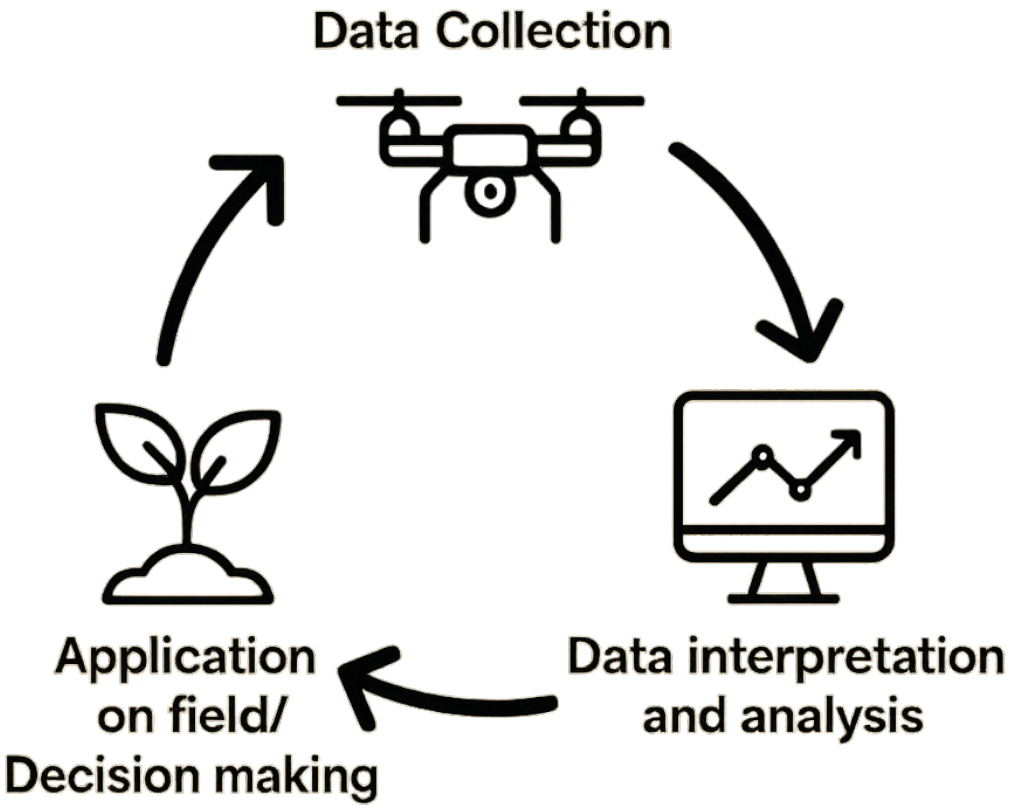

Figure 2.1: Precision agriculture cycle

contains abnormalities, outliers, and inconsistencies, which, if ignored, could lead to erroneous analysis. Cleaning techniques described in [58] are used to remove these abnormalities. The cleaning procedure ensures that the data is reliable and accurate, which is critical for further analysis.

After the data has been cleaned, statistical and machine learning techniques are used to extract insights and inform decision-making. For instance, [13] illustrates how machine learning and statistical approaches can be used to predict crops. [55] used methods including linear regression, random forests, and cross-validation. One takeaway from this study is the optimal timing for image capture, which can help reduce the frequency and cost of image acquisition across the growing season. By understanding when and how to collect imagery, farmers can make data-driven decisions that optimize input use and increase productivity.

2.1.2 Remote sensing

Remote sensing is the technique of collecting information about an object, location, or phenomenon without physically interacting with them. This is typically done using sensors. Yield data imagery can be collected using sensors mounted on satellites and drones. [74] highlights the strengths and weaknesses of each. While satellites are great for capturing images of large areas, their limitations in spatial and temporal resolution have resulted in a rising preference for UAVs in certain applications.[72, 50, 60] examine some use cases of UAVs in precision agriculture, including crop monitoring, weed mapping, crop spraying tasks, fertilizer management. UAVs provide high spatial resolution and allow data collection at flexible, user-defined intervals, making them a valuable tool for precision agriculture.

The integration of Red-Green-Blue (RGB) and multi-spectral cameras increases their capacity for detailed analysis, since they can collect information beyond human vision. These capabilities contribute to more effective farm management, reduced input costs, and increased crop productivity.

2.1.2.1 Red-Green-Blue (RGB) and multi-spectral imagery using UAVs

Precision agriculture UAVs can be equiped with RGB and/or multispectral sensors, enabling the collection of full-field imagery. RGB (Red, Green, and Blue) is the abbreviation for the three primary colours of light used in colour imaging systems to capture the visible spectrum. These RGB images are excellent for detecting changes in the landscape, such as deforestation, crop health, and climate change. Furthermore, the many colour channels of RGB photography can be employed as features in advanced artificial intelligence models such as Convolutional Neural Networks (CNNs). These AI models can evaluate RGB data and perform tasks such as classification, pattern recognition, and prediction. For instance, [10] successfully integrated RGB imagery with CNNs for crop yield prediction, demonstrating the model potential to extract insights from low-cost RGB camera images. [9] also used RGB imagery data on 3D CNNs models and got promising results.

Multi-Spectral (MS) camera collects data from multiple wavelengths of light, typically beyond the visible spectrum. This type of photography is captured using sensors that can detect light in a variety of wavelengths or bands, including visible light and the Near-Infrared (NIR) area of the electromagnetic spectrum. Multi-spectral imagery collects data in a variety of bands, offering more detailed information than RGB alone. According to [55], crop yields may be forecasted using RGB and MS images but MS imagery gives the most promising results.

Both RGB and MS bands can be used to compute Vegetation Indices (VIs), which are valuable features for precision agriculture models.

2.1.2.2 Vegetation indices (VIs)

Vegetation Indices (VIs) are mathematical formulas employed to measure the density, health, and condition of vegetation. These indices originate from the reflectance measurements of various wavelengths of light captured by remote sensing technologies such as satellites, drones, or other sensors. Observations have shown that healthy

vegetation reflects strongly in the NIR and absorbs most of the visible light, while stressed or unhealthy vegetation exhibits different reflection characteristics. By analyzing the light reflected by vegetation, VIs can indicate plant health, moisture levels, and other critical agricultural factors.[55] used and compared and ranked the corn grain prediction power of 55 vegetation indices. [66] used VIs alongside convolutional neural network for yield crop prediction. Some of the most common VI used for yield prediction are :

1. **Normalized Difference Vegetation Index (NDVI)** — NDVI quantifies the difference between the red light absorbed by vegetation and the near-infrared light reflected by it. It helps moniroting vegetation density and health :

$$\text{NDVI} = \frac{\rho_{\text{NIR}} - \rho_{\text{Red}}}{\rho_{\text{NIR}} + \rho_{\text{Red}}}$$

2. **Normalized Difference Red Edge Index (NDRE)** — NDRE replaces the red band in NDVI with the red-edge band, improving sensitivity to chlorophyll concentration changes, particularly during later growth stages:

$$\text{NDRE} = \frac{\rho_{\text{NIR}} - \rho_{\text{RedEdge}}}{\rho_{\text{NIR}} + \rho_{\text{RedEdge}}}$$

3. **Enhanced Vegetation Index (EVI)** — EVI adjusts for atmospheric effects and canopy background noise, making it particularly useful in areas with dense vegetation:

$$\text{EVI} = G \times \frac{\rho_{\text{NIR}} - \rho_{\text{Red}}}{\rho_{\text{NIR}} + C_1 \times \rho_{\text{Red}} - C_2 \times \rho_{\text{Blue}} + L}$$

   where $G = 2.5$, $L = 1$ (canopy background adjustment), and $C_1 = 6$, $C_2 = 7.5$ (aerosol resistance coefficients).

4. **Soil-Adjusted Vegetation Index (SAVI)** — SAVI incorporates a soil-brightness correction factor L to minimize background interference from ex-

posed soil, which is especially important in areas with sparse vegetation cover:

$$\text{SAVI} = \frac{(1+L)(\rho_{\text{NIR}} - \rho_{\text{Red}})}{\rho_{\text{NIR}} + \rho_{\text{Red}} + L}$$

where L typically equals 0.5 for intermediate vegetation cover.

For this study, we chose to work with two (02) vegetation indices, NDVI and NDRE, because they have been widely used in previous research and have demonstrated a good fit for crop yield prediction.

2.1.2.3 Geospatial data

Geospatial data refers to information describing objects by combining location characteristics such as longitude and latitude with attribute information (characteristics of the object) and temporal information. They can be collected through various methods, including satellite, aerial photography, GPS devices and field surveys. Geospatial data are often processed using Geographic Information Systems (GIS), which are technology designed for gathering, managing , analyzing and visualizing geospatial data. GIS allows the association of maps with data resulting in a comprehensive visual representation, such as hurricane maps, urban development maps, and yield precision maps, which are useful for analysis and decision making. The GIS software used for this study are QGIS and PIX4D. There are two main types of geospatial data :

1. Raster data : This type of data is represented in a grid format, where each pixel holds a specific value. Examples : satellite imagery, temperature maps...
2. Vector data : Information is stored using points, lines and polygons. Points represent discrete locations (e.g., weather stations), lines represent features (e.g., rivers, roads) and polygons represent areas (e.g., fields).

The distinction between raster and vector representations is illustrated in Figure 2.2.

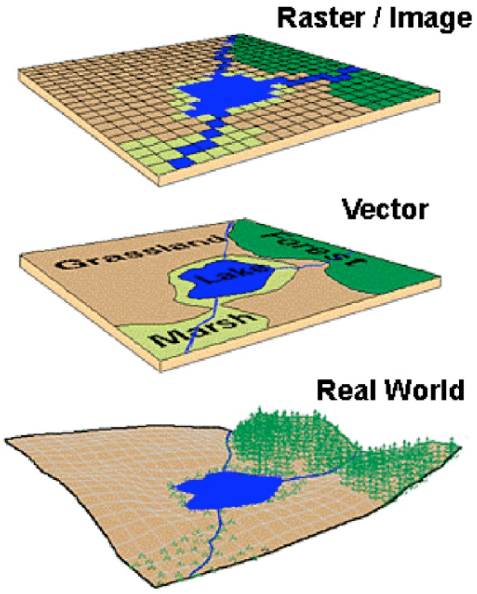

Figure 2.2: [25] Comparison between raster, vector and real-world data

2.1.2.4 Orthomosaics

An orthomosaic is a high-resolution image generated by stitching together several smaller images, known as orthophotos[21]. Orthophos are corrected for lens distortion, camera tilt, perspective and topographic relief to guarantee that all features are aligned to their true geographic positions. The picture is similar to a picture got from Google Earth but offers a higher resolution and informations. The precision and accuracy of orthophotos makes the orthomosaics equivalent to maps in terms of spatial fidelity. This allows orthomosaics to be used for tasks such as measuring true distances, analyzing spatial relationships, conducting georeferencing with high accuracy, extracting features and performing spatial and temporal analysis to track changes in the landscape over time.

Orthomosaics are widely used in fields such as remote sensing, precision agriculture, urban planning, and environmental monitoring. For example, [73] uses orthomosaic for yield prediction on canola. [52] uses orthomosaic for identifying blueberry bushes

and locating them.

2.2 Machine Learning

2.2.1 Models

With the growing interest on precision agriculture and the development of the technology, predicting crop yield has become an important task for decision-making in agriculture. It can help farmers to determine what to grow and when to grow. In the past, farmers used to rely on their experience and personal understanding to estimate the crop yield but this method tend to likely be not accurate. Population growth, climate change and soil loss are many factors that can impact the yield. Corn is one the most predicted product [55, 10, 53]. But other crop are also involved in yield prediction. [83] did a study on soybean yield prediction. [91] worked on cotton yield prediction. Many machine learning techniques can be use for crop yield prediction.[56] did what they call a Systematic Literature Review to extract and synthesize the algorithms and features that have been used in crop yield prediction studies. They found that the most applied algorithm is Artificial Neural Networks in these models. And among these deep learning papers, Convolutional Neural Network (CNN) is found to be the most widely used. But others machine learning techniques such as Linear Regression and Random Forest has been also used.

2.2.1.1 Random Forest

Random forest in one of the most popular machine learning techniques. It works by creating a number of decision trees during the training phase. Each tree is constructed by including some randomness which helps to prevent overfitting. The random forest works by in several steps. It create random subsets of the whole dataset with replacement to train each tree. At each node of the tree, a random subset of features is used. Each tree operate independently and give an output. At

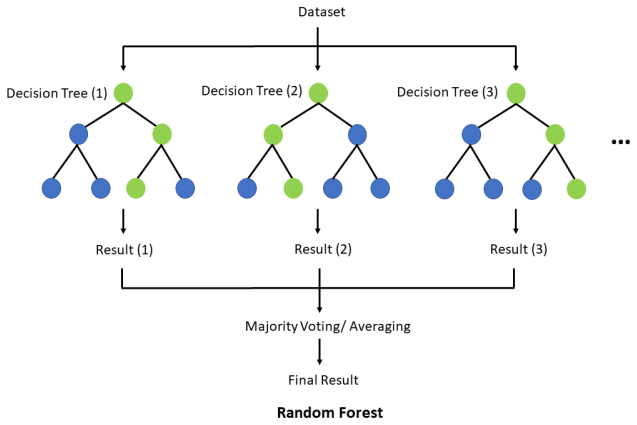

Figure 2.3: [26] Random Forest Explained

the end the result of each tree is either aggregated by either average of all results for regression task or by majority voting for classification tasks. [55, 89, 48] use random forest for their crop prediction tasks.

2.2.1.2 Linear Regression

Linear regression is a machine learning technique used to model a relationship between at least two variables by fitting a linear equation to the observed data. The output is commonly referred to as the dependent variable, while the involved variables in the process of prediction are called independent variables or predictors. Linear regression is characterized by its simplicity, interpretability, and low computational cost. [55] included linear regression in his study for corn yield prediction. The general form of a multiple linear regression model is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon$$

where:

- y: Dependent variable (e.g., yield)

- $x_1, x_2, \ldots, x_n$: Independent variables (e.g., vegetation indices, spectral band values)
- β_0: Intercept term
- $\beta_1, \beta_2, \ldots, \beta_n$: Coefficients representing the effect of each predictor
- ϵ: Error term accounting for unexplained variability

The model parameters β_i are typically estimated using the Ordinary Least Squares (OLS) method, which minimizes the sum of squared residuals between observed and predicted values.

2.2.1.3 Support Vector Machine (SVM)

Support Vector Machines (SVM) are supervised learning algorithms that can be used for both classification and regression tasks. As a classifier, sometimes called Support Vector Classification (SVC), the model seeks to find the optimal decision boundary—called a *hyperplane*—that best separates data points from different classes by maximizing the distance between the hyperplane and the nearest data points on each side [97]. Larger distance means better model. Many models are not linearly separable in their original feature space. To address this, the *kernel trick*, a technique that maps the input data into a higher-dimensional feature space where a linear separation can be used. Commonly used kernel functions include the linear kernel, polynomial kernel, radial basis function (RBF), and sigmoid kernel. In regression settings, a variant known as Support Vector Regression (SVR) is used. Unlike traditional regression techniques that aim to minimize error between predicted and actual values, SVR attempts to fit a function within a specified margin of tolerance (*epsilon-insensitive tube*) around the actual targets, while also minimizing model complexity.

In crop yield prediction, SVMs have been applied to model the complex [32, 37].

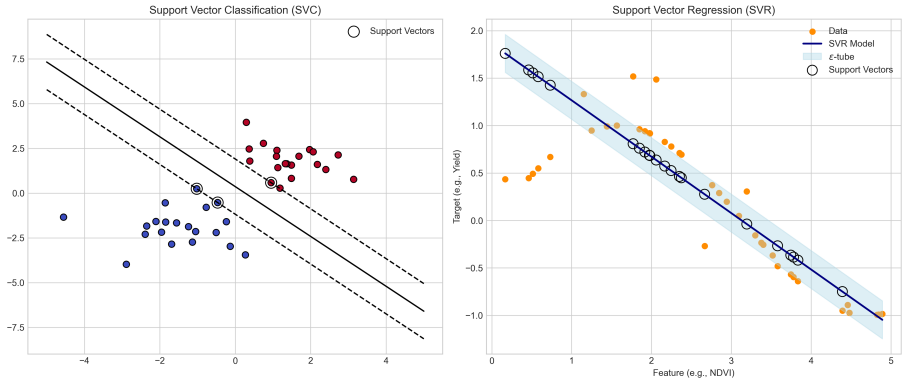

Figure 2.4: SVC and SVR visual representation

2.2.1.4 Neural Network

Neural networks are inspired by the human brain where we try to connect multiple neurons to process information and learn tasks for pattern recognition and decision-making. They are composed of node layers, containing an input layer, one or more hidden layers and an output layer. Information is processed through these layers. Weights are assigned to each node in the input layer, representing it the contribution to the output. Larger weight implies larger influence on the final output. All the weighted inputs are summed and passed through an activation function introducing non-linearity and enables to learn complex patterns. Some popular activation function are ReLU and the Sigmoid function. The output of the previous layer becomes then the input of the next layer and this process continues until the output layer is reached.

2.2.1.4.1 Convolutional Neural Networks Convolutional neural networks (CNN) are a subset of the bigger fields of deep learning and neural networks. They are distinguished from the other neural networks for its good performance in processing images and audio signals. They have mainly three types hidden of layers : convolutional layers, pooling layers and fully connected layers. The input layer receives the input data, which can be images or audio files. Data are passed to the convolutional

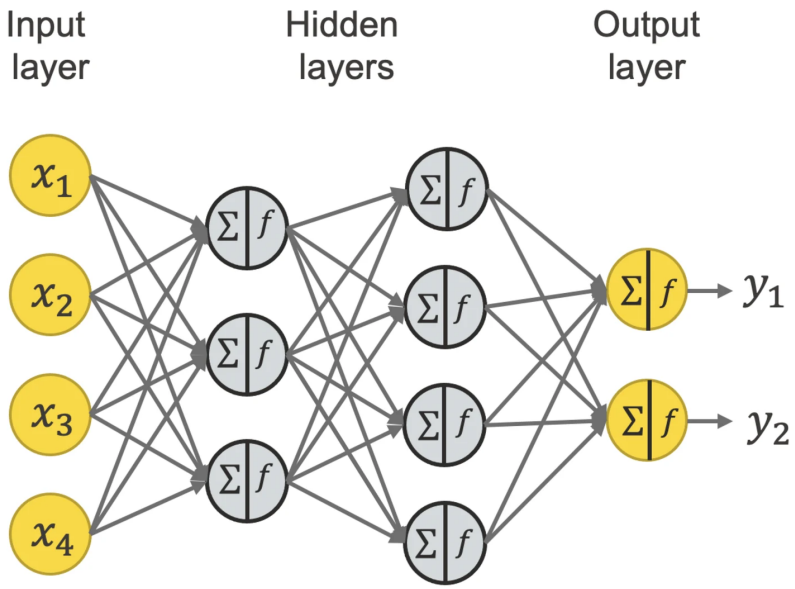

Figure 2.5: [2] Simple Neural Network

layer, whose role is to detect the features of the input data, such as edges, texture and patterns. This layer uses a set of filters, which are applied to small regions of the input data, to extract features from the image. The set of features is passed to the pooling layer, which helps downsampling the data to reduce its dimension by reducing the number of parameters allowing the network to focus on the most important features. While some information can be lost during this phase, it also helps to reduce the complexity, improve efficiency and reduce the risk of overfitting. After that the features extracted from the convolutional and pooling layers are passed to the fully connected layers, where they are used to classify the input image into one of several possible classes. The final output of the CNN is a probability distribution over the classes, indicating the likelihood that the input image belongs to each class.

CNNs are a very useful tool when it come to crop yield prediction using imagery of the field. As said earlier, [56] found in their comparative study of yield prediction using machine learning that CNN-based models are the most widely used for yield

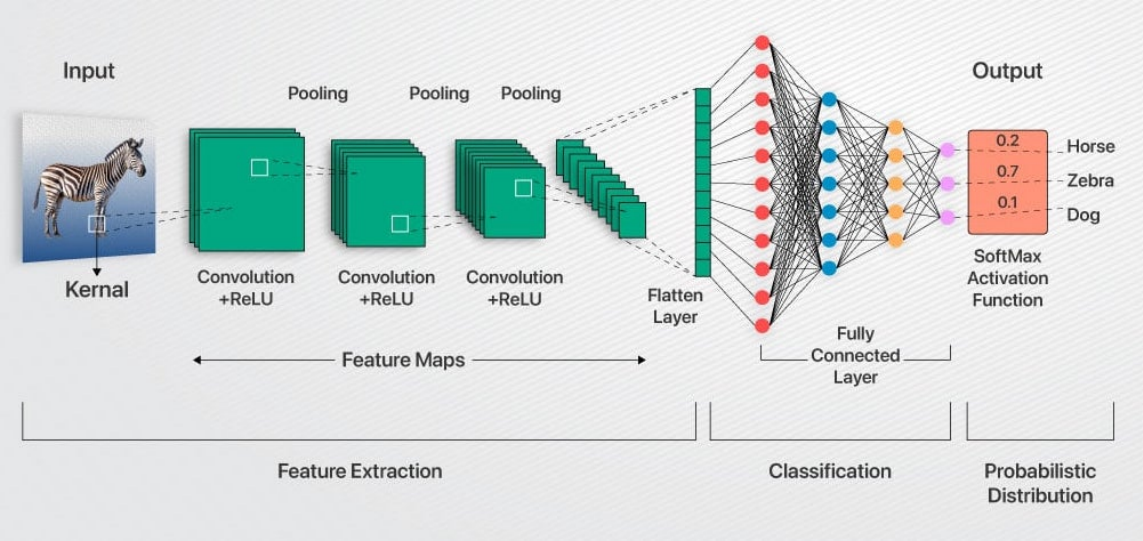

Figure 2.6: [80] Convolutional Neural Networks Explained

prediction. [62] found that CNNs are able to reduce the crop yield prediction uncertainty considerably. [91] used colour RGB images from simple mobile devices to predict the cotton yield from commercial field. [10] used two types, 2D-CNN and 3D-CNN, to predict the corn crop yield. [15] used a novel clustered 3D-CNN model to predict future prices of crops including wheat, oats, rice and soybeans.

We chose to use two machine learning models that are well-suited to our input data while also offering computational efficiency. Although convolutional neural networks (CNNs) are widely used, they typically involve higher computational costs and require larger amounts of input data.

2.2.2 Clustering

Clustering is basically grouping similar data points into distinct categories or clusters, based on a defined similarity metric [104]. The goal is to ensure that data points within the same cluster are more similar to each other than to those in other clusters. Clustering algorithms typically partition a dataset into a fixed number of clusters,

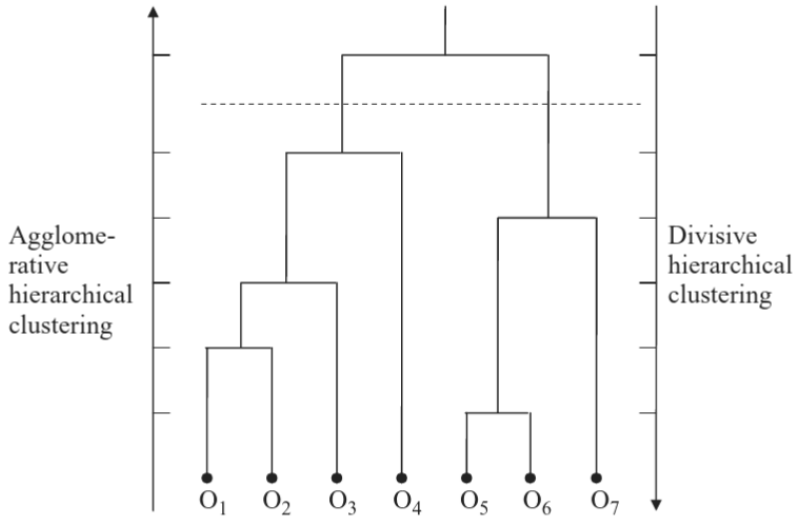

Figure 2.7: [104] Example of a dendrogram resulting from hierarchical clustering

each representing a substructure or pattern within the data.

There are two main categories of clustering techniques: *hierarchical* and *partitional* clustering.

2.2.2.1 Hierarchical Clustering

Hierarchical clustering creates a structure of nested clusters based on the pairwise proximity between data points. It does not require a predefined number of clusters. Instead, it builds a hierarchy, which can be visualized as a dendrogram. There are two main approaches:

- **Agglomerative (bottom-up):** Starts with each data point as a single cluster and recursively merges the closest pairs of clusters.
- **Divisive (top-down):** Starts with all data points in one cluster and recursively splits them into smaller clusters.

Figure 2.8: K-means clustering with K=3

2.2.2.2 Partitional Clustering

Partitional clustering assigns data points directly into K clusters, typically without generating a nested structure. Unlike hierarchical methods, partitional approaches require the number of clusters (K) to be specified in advance.

2.2.2.2.1 K-Means Clustering K-Means is one of the most widely used partitional clustering algorithms. It partitions the data into K clusters by minimizing the intra-cluster variance. The algorithm iteratively assigns each data point to the nearest cluster centroid and updates the centroids until convergence.

K-Means has been applied in agricultural context. For instance, [55] employed K-Means clustering to implement a spatial cross-validation strategy in a corn yield prediction study.

2.2.3 Evaluation metrics

Regression performance can be assessed using complementary metrics to capture absolute and relative errors, variance. [56] gives a list of the most common metrics in crop yield prediction. For this work we used the following regression evaluation metrics :

2.2.3.1 Root Mean Squared Error (RMSE).

RMSE is one of the most common metric used in regression tasks and particularly in yield prediction studies. According to [56], its the most used metric in crop yield prediction. It measures the average magnitude of errors, penalizing large deviations:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}.$$

Lower is better. Unlike the Mean Squared Error (MSE), which omits the square root, RMSE is expressed in the same units as the target variable, making it easier to interpret. In this study, RMSE is reported in bushels per acre (bu/ac), corresponding to the original scale of the yield data.

2.2.3.2 Mean Absolute Error (MAE).

MAE is very similar to MSE with the square replaced by absolute value. It's robuster than RMSE to outliers since MSE and RMSE tends to penalize large deviations.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|.$$

Lower is better. MAE is reported in the same unit as the target variable.

2.2.3.3 Mean Absolute Percentage Error (MAPE).

MAPE is one of the most interpretable metrics since it is expressed in percentage. It's scale-free and measure proportion. MAPE is not a common metric in crop yield prediction [56] compared to RMSE or R^2, but can be useful in result interpretation.[10, 62, 6]

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{\max\left(|y_i|, \varepsilon\right)} \right| \ \%.$$

We use a small floor ε (e.g., ε=0.01 t/ha) to stabilize the denominator in low-yield areas. *Lower is better.* It is interpretable as average % deviation; but can overweight errors where y_i is small.

2.2.3.4 Coefficient of Determination (R^2).

According to [56], R^2 is the second most frequently used evaluation metric in crop yield prediction studies, after RMSE.It measures the proportion of variance in the observed data that is explained by the model.Values of R^2 typically lie within the range $[0, 1]$, where higher values indicate better explanatory power. However, negative values are possible and indicate that the model performs worse than simply predicting the mean, making R^2 unbounded below. This metric has been widely adopted in yield prediction studies, including [73, 9], due to its interpretability and ability to quantify the goodness-of-fit.

$$R^2 \ = \ 1 \ - \ \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}.$$

2.2.3.5 Pearson Correlation Coefficient (r).

The Pearson correlation coefficient measures the strength and direction of the linear relationship between the observed and predicted values:

$$r = \frac{\sum_{i=1}^{n}(y_i - \bar{y})\ (\hat{y}_i - \overline{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\ \sqrt{\sum_{i=1}^{n}(\hat{y}_i - \overline{\hat{y}})^2}}.$$

Its value ranges from -1 to 1, where values close to ± 1 indicate a strong linear relationship, and values near 0 suggest little to no linear association. While the sign denotes the direction of the relationship, the absolute magnitude reflects its strength. Pearson's r is widely used in yield prediction studies to evaluate the agreement between predicted and observed spatial patterns.

2.3 Spatial Considerations

Agricultural systems carry spatial structure as a result of environmental factors, management practices, and biological processes. Understanding these spatial patterns is essential for developing realistic and generalizable predictive models in precision agriculture.

2.3.1 Spatial autocorrelation

Many yield prediction studies assume that yield observations are spatially independent. However , according to Tobler's First Law of Geography, "everything is related to everything else, but near things are more related than distant things" [93]. This principle highlights the presence of spatial autocorrelation in agricultural datasets. Validation methods commonly include random data splitting and cross-validation (CV). Cross validation is a resampling methods that partitions data into multiple training and test set in order to minimize overfitting. However, in the context of

spatially structured data, random splitting tends to underestimate model prediction error because it ignores spatial dependence [59]. This limitation has also been reported specifically in yield prediction studies [55].Despite this random split strategy is still widely applied. [77] exhibits the importance of implementing k-fold CV in yield prediction model, while other studies included the spatial dependence in their studies for crop yield prediction [33].

2.3.2 Spatial Cross validation methods

To address the challenge of random resampling, several spatial CV strategies have been proposed. These methods partition data into a training set D_{in} and a held-out set D_{out}, but instead of assigning observations randomly, folds are defined based on spatial dependence structures [59].

2.3.2.0.1 Leave-One-Block-Out (LOBO) In LOBO, the study region is divided into spatial blocks (often regular polygons such as grids). All observations within one block are assigned to D_{out} while the remaining blocks form D_{in}. This approach enforces spatial independence between training and testing sets. A challenge with spatial blocking, however, is that standard grids can group unrelated areas together, potentially producing unrealistic folds.

2.3.2.0.2 Leave-One-Cluster-Out (LOCO) LOCO is an alternative in which folds are created from spatial clusters identified through clustering algorithms. Unlike rigid grid-based blocking, clusters tend to form more natural and contiguous groupings of observations, producing circular or irregularly shaped partitions that may better reflect field structure.

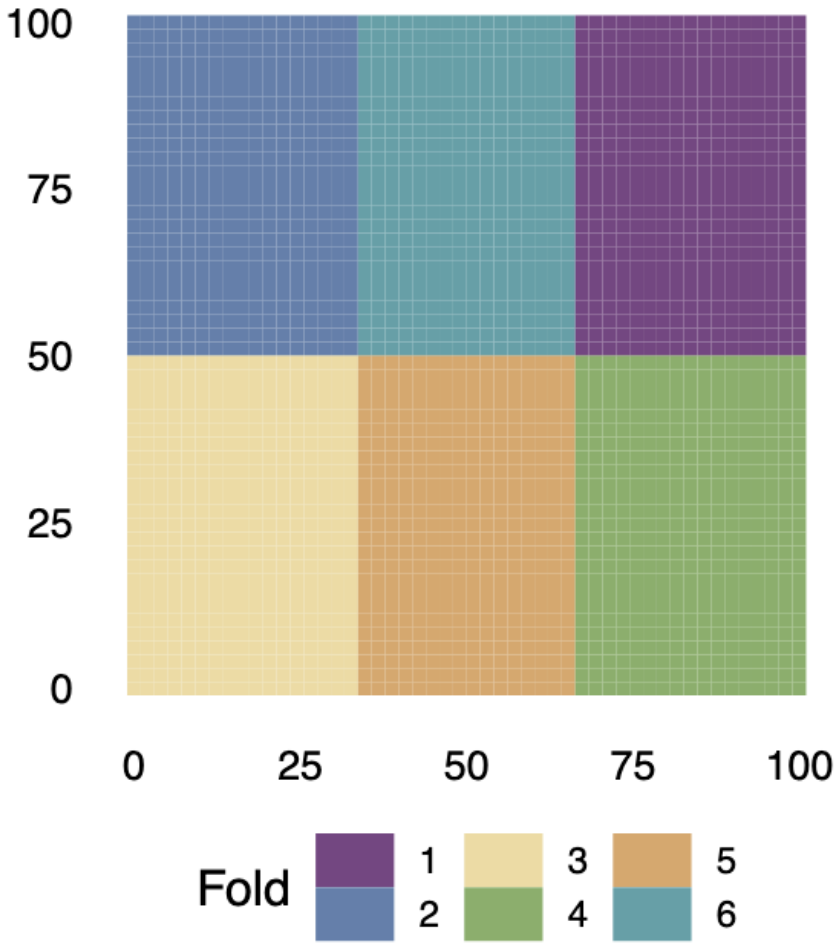

Figure 2.9: Spatially blocked CV fold assignments produced by [59] to group observations spatially

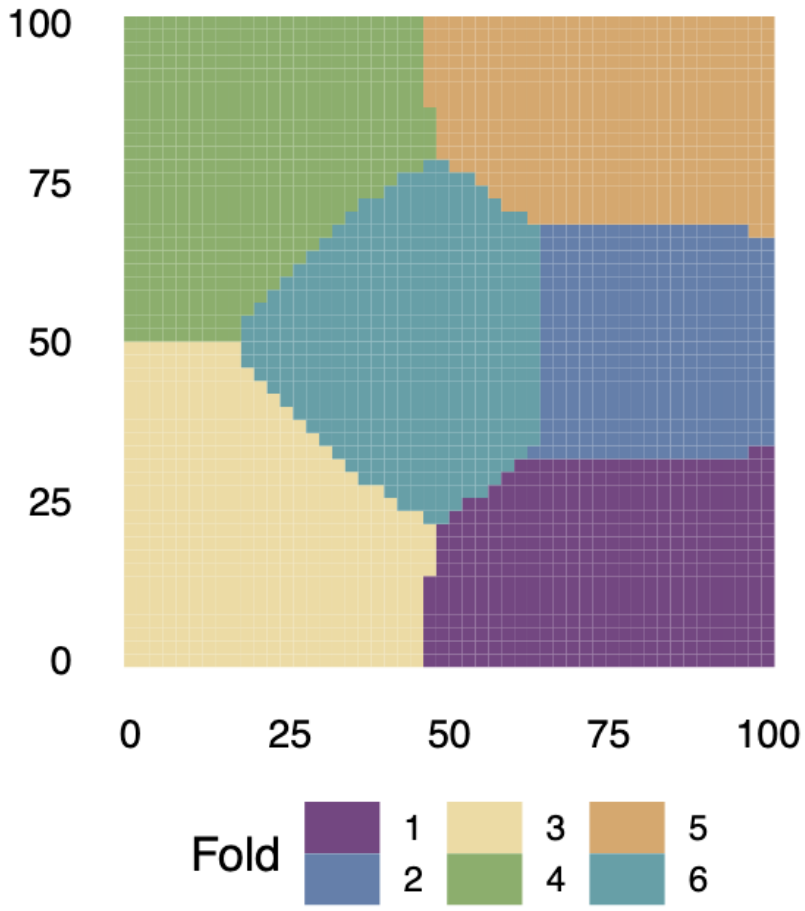

Figure 2.10: Spatial clustering CV fold assignments produced by [59] using k-means clustering to group observations spatially

2.3.2.1 Field-to-Field Transferability

While a large proportion of yield prediction studies are restricted to individual fields, an equally important dimension of spatial generalization is the ability of models to transfer across fields. This concept of field-to-field transferability refers to the extent to which a predictive model trained on data from one field remains accurate when applied to another field with different, yet potentially related, characteristics.

A number of studies have attempted to evaluate such transferability across nearby fields located within the same agricultural site. For instance, [73] investigated canola yield prediction using UAV orthomosaics from a single field, focusing on within-field variability rather than transfer across fields. In contrast, [88] and others explored models trained and validated across multiple fields, while [55, 33] explicitly assessed the challenges of spatial dependence and transferability in cross-field scenarios.

Chapter 3

Methodology

This section outlines the methodology employed to develop the proposed machine learning models for yield prediction using UAV multispectral imagery. It details the process from data acquisition and pre-processing, through dataset generation, to model design and evaluation.

3.1 Study Site

To address the yield prediction task, datasets were collected from three (03) agricultural fields located in different regions of Canada. Two fields were situated in Manitoba—Ed's (Man Eds) and Home (Man Home)—while one field, Precision Planter (Alb PP), was located in Alberta. The Man Home field was cultivated with corn, whereas Man Eds and Alb PP were cultivated with canola. Table 3.1 summarizes the key characteristics of these fields, including their size, crop type, and management dates (planting and harvest).

The heterogeneity across these fields—in terms of crop type, geography, and management practices—directly supports two objectives of this study: (i) evaluating the ability of machine learning models to generalize under real-world conditions where

input data are diverse and non-uniform, and (ii) assessing the field-specific predictive performance.

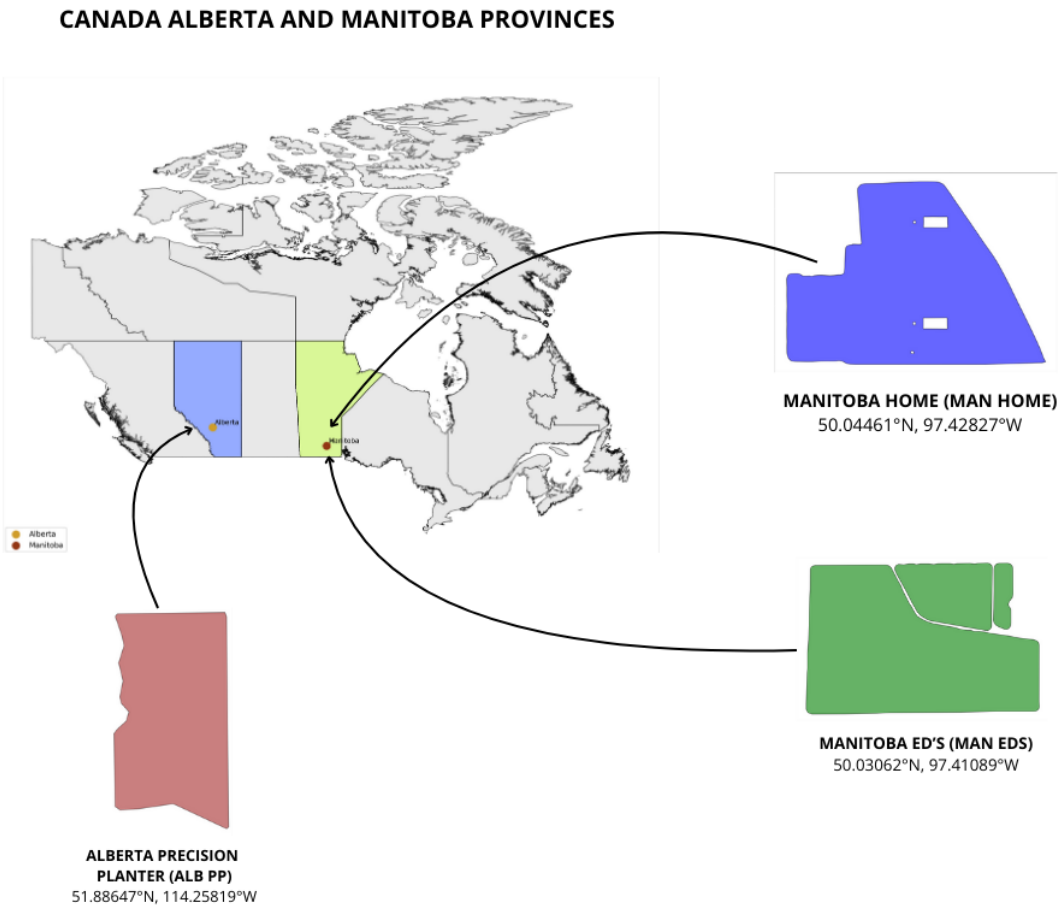

Figure 3.1: Map of field locations and boundaries in Alberta and Manitoba.

Two main sources of data were collected from these sites: (1) yield monitor data acquired during harvest and (2) UAV-based multispectral orthomosaic imagery acquired during the growing season. Together, these datasets form the basis for feature engineering, dataset construction, and machine learning model development. The timeline from planting to harvesting for each field is represented in Figure 3.11.

Field Name	Location	Size (ha)	Crop Type	Planting Date	Harvest Date
Precision Planter	Alberta, Olds College	26.29	Canola	May 30, 2024	October 16, 2024
Ed's	Manitoba, Winnipeg	93.30	Canola	June 1, 2024	September 13, 2024
Home	Manitoba, Winnipeg	66.22	Corn	May 10, 2024	October 6, 2024

Table 3.1: Field information and planting details for the three study sites.

	geometry	date	swath_width	yield	moisture	speed	duration
0	POINT (-114.26117 51.88365)	2024-09-26	37.5	17.4152	3.3	1.6001	0.917
1	POINT (-114.26117 51.88365)	2024-09-26	37.5	19.5229	3.3	1.6171	1.092
2	POINT (-114.26117 51.88366)	2024-09-26	37.5	35.8282	3.3	1.5988	1.213
3	POINT (-114.26117 51.88367)	2024-09-26	37.5	41.2722	3.3	1.5599	1.020
4	POINT (-114.26117 51.88368)	2024-09-26	37.5	34.1076	3.3	1.6832	1.211

Figure 3.2: Geopandas dataframe containing data used for yield data preprocessing

3.2 Yield Data

Yield data were collected in every field at the end of the season, consisting of continuous measurement taken at sequential points of the field. Since collected from various field and different harvester combine, data collected can be different from one field to another. But essentials data for this study-such as location coordinates, moisture levels, harvester speed, swath width, duration and time-were kept and uniformized. The data were stored in multiple shapefiles, with the primary access point being the **.shp** file extension. Tools such as Geopandas [31], a python library for geospatial data manipulation were used to read and manipulate the data. [58] suggested that many of erroneous data should be removed and not corrected. That's what have been done in this study. A sample of the yield dataset, showing the relevant columns used in this study, is presented in Figure 3.2.

3.2.1 Harvest Segment and short segment

For more precise analysis, yield data points collected from the harvester combine were grouped into segments called "harvest pass or harvest segment". These segments consist of sequential harvest points were created based on the distance between each consecutive point. They can be drawn from the data. [11] constructs its harvest segment using the time between readings. They computed the mode duration reading between two consecutive time and set a threshold of three times (3x) this mode value to be the beginning of a new segment. In our context, the process of generating harvest segments involved calculating the average distance between consecutive points harvested the same day. If the distance between two consecutive points is larger than three (03) standard deviation from the mean, the later point were designated as the starting point of a new harvest run or harvest pass. This approach assumes that during such gap, the continuous yield measurement was interrupted and possibly the swath was raised, as illustrated in Figure 3.3.

[5] demonstrated that the accuracy of the yield measurement decreases with the segment length. The longer the segment is, less are the percentage of yield data errors. This suggests removing short segments, which is exemplified in Figure 3.4. [58] suggested removing segment less than 108 m (355 feets). This is what we have done in this study.

3.2.2 Harvester Fill Mode and Finish Mode Error

The first type of error cleaning applied to our processing pipeline is what we call the harvester fill mode and finish error. The lowering and the raising of swath at the start and end of harvest pass is sometimes characterized by low yield measurement due to fill and empty mechanism. Various method were proposed to deal with this type of error. [11] suggests removing some points recorded at the beginning of the segment. [58] suggests according to their observations that the first 24s and the last 10s should be removed from each harvest pass. We applied this suggestion to our

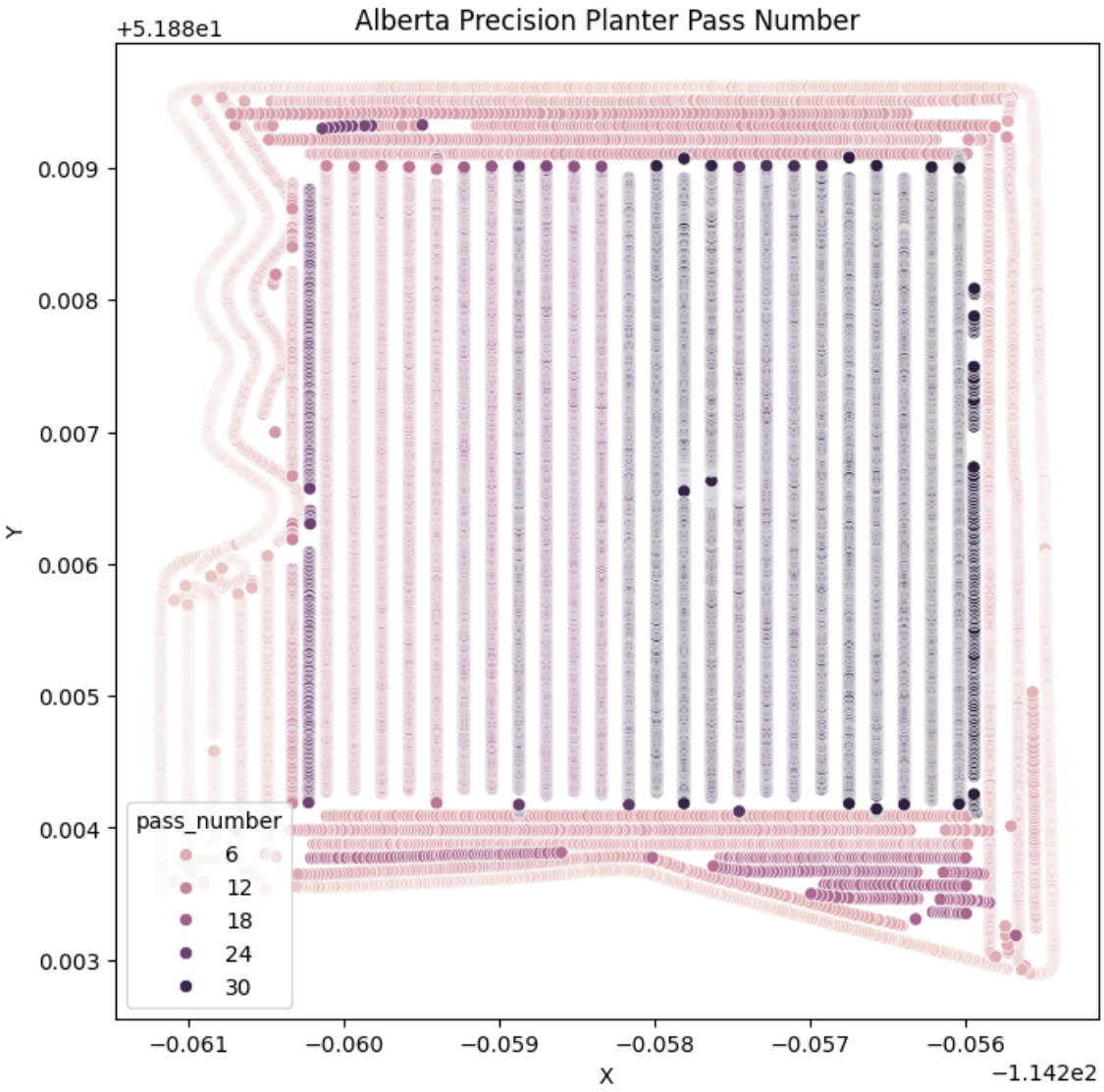

Figure 3.3: Precision Planter Harvest Pass Number

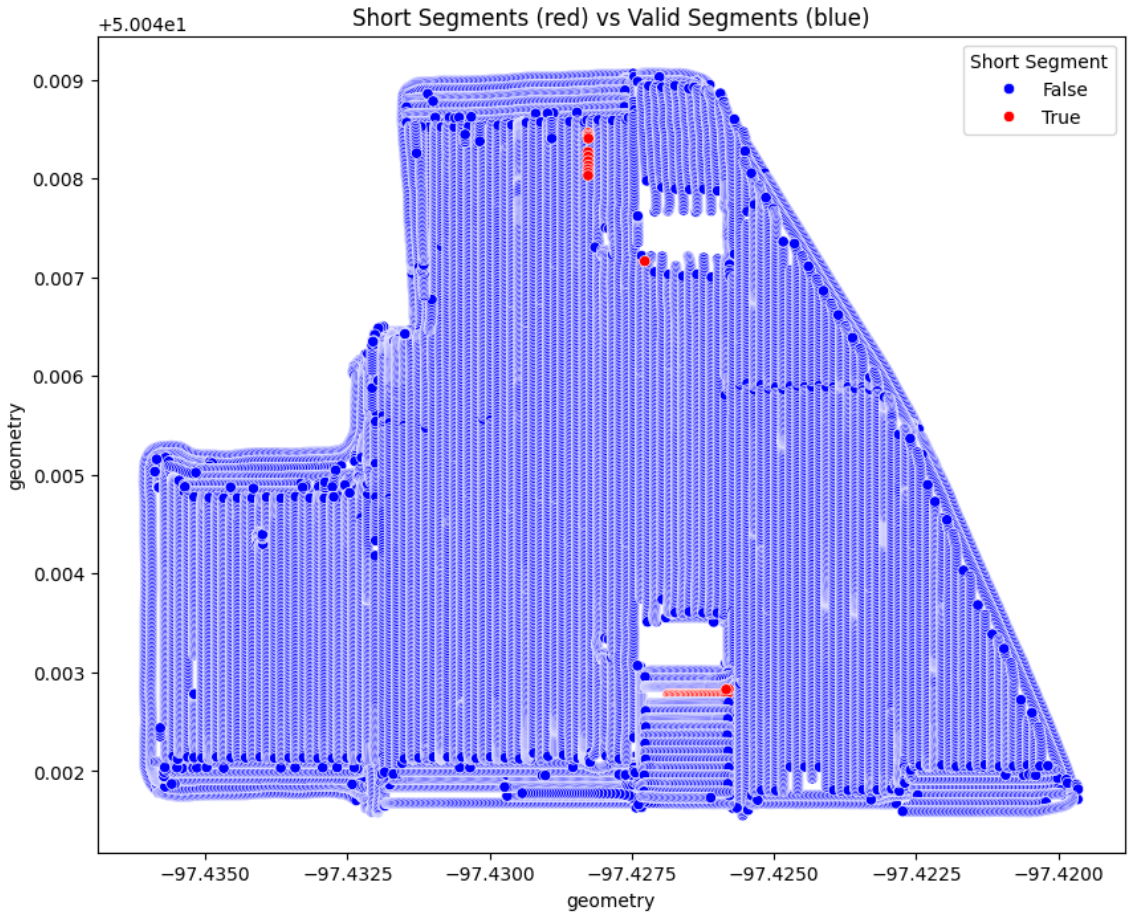

Figure 3.4: Short segment vs Valid segment (Manitoba Home)

study removing then the first 24s and the last 10s.

3.2.3 Forward and Backward Measurement Filter

The continuous measurement of yield and moisture can introduce errors, influencing the accuracy of the recorded data. Erroneous yield measurements often occur when the ground speed and distance traveled are low or when the high yield is recorded over a short distance. Many methods are have been applied to remove outliers values such as applying biological limits filter [11] or excluding ±3 standard deviation from the mean.

However, the effectiveness of some filtering methods depends on the distribution of the data (more efficient for normal distribution) and do not consider the spatial variation within the yield map. To address this limitation, an alternative filtering method has been applied. This approach compares each value with its neighboring values and sets a threshold beyond which that value is considered as an outlier. This method implies selecting a range of backward and forward observations, computing the mean of this observations and comparing the current observation with theses mean values using a predefined threshold . [58] set a threshold of ±30%. During this study, a window of four (04) values in both backward and forward direction was used, with a threshold of ±30%. This algorithm is applied to filter the moisture, the speed and the yield measurements.

3.2.4 Overlapping Area

During the harvester's movement, an harvested area can be delineated using the swath width and the distance between the current location and the next point. However, the cutting bar may sometimes overlap an area already marked as harvested, what can lead to erroneous measurements. Its therefore important to remove overlapping points to obtain a more precise data without any ambiguity. To do so, we applied the method proposed by [10] to define the harvested area. We constructed

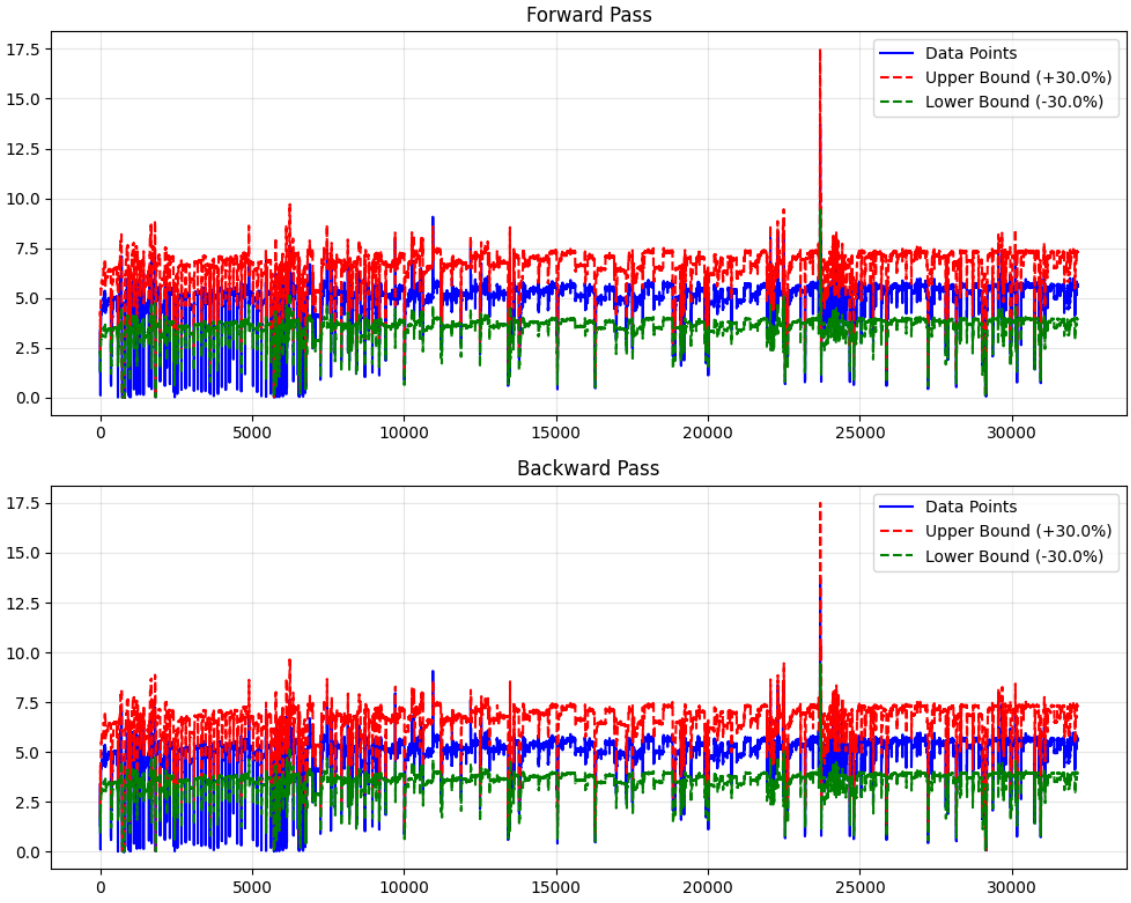

Figure 3.5: Speed filtering using backward and forward measurements filter

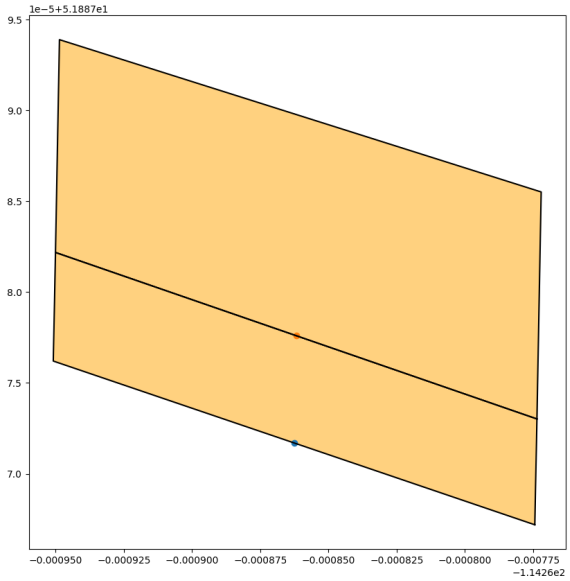

Figure 3.6: Harvested area between two points

a polygon using the bearing angle between two consecutive points to determine the travel direction, while setting the point coordinated as the center of the swath. This is illustrated in Figure 3.6. A polygon was then generated from the edge points using the GeoPandas library [31]. All geospatial manipulations were performed using the WGS84 ellipsoid coordinate system [99].

3.2.5 Isolation Forest

Isolation forest [44] is an unsupervised machine learning technique to detect outliers. It is based on the assumption that because anomalies are few and different from other data, they can be isolated using few partitions. Trees are constructed by splitting the data recursively until all data points are isolated. The technique is continued until a forest of decision trees is formed. [10] used it in his yield data processing. For this study we used a contamination parameters of 0.05 , 100 estimators and random state 42.

3.2.6 Z-Score Filter

Z score is a popular statistical concept and is sometimes used to remove outliers from datasets [10, 58, 107]. It measures how many standard deviations a data point is from the mean and can be used as a threshold to filter out extreme values. It is important to notice that the effectiveness of this method is most pronounced in normally distributed data. A Distribution analysis was done during this study and as observed by [58] the raw yield distribution was found to be non-normal. However, at the end of the processing pipeline, the data distribution became closer to normal distribution, allowing us then to apply the z-score filter. The distribution of the original versus filtered yield data (before applying the z-score filter) is shown in Figure 3.7, while the overall yield data processing pipeline is summarized in Figure 3.8.

3.3 Orthomosaics Imagery Data

UAV flights were conducted over each study field to capture high-resolution multispectral imagery. Raw images were stitched into full-field orthomosaics using the Pix4Dmapper software [69]. The quality of the orthomosaics varied across sites due to differences in camera specifications, acquisition dates, and flight conditions.

For the two Manitoba fields (Ed's and Home), imagery was acquired on July 29, 2024 (DOY 211).

- For Ed's, this corresponds to 59 days after planting and 46 days before harvest.
- For Home, this represents 80 days after planting and 63 days before harvest.

Both orthomosaics were captured with the MicaSense RedEdge-P multispectral camera [76], which provides five raw spectral bands (red, green, blue, near-infrared, and red-edge). The resulting spatial resolutions were 3.43 cm for Ed's and 3.56 cm for Home.

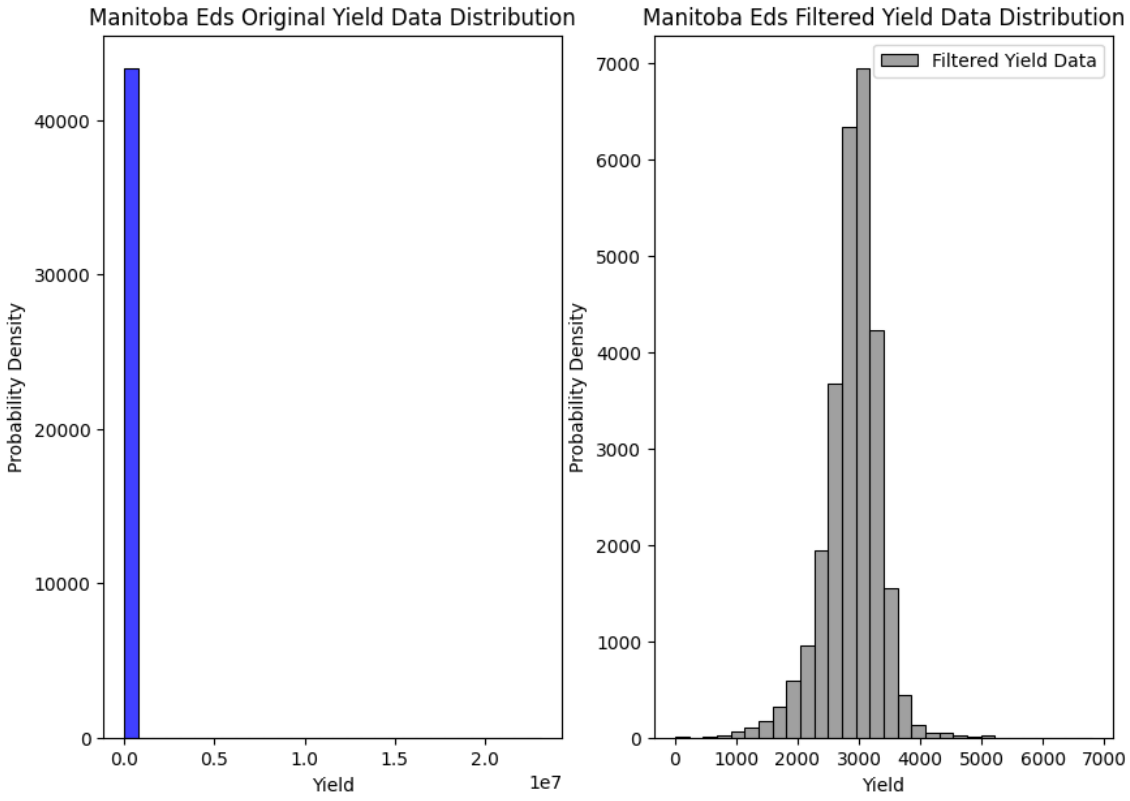

Figure 3.7: Original yield distribution vs Filtered yield distribution (before applying z-score filter)

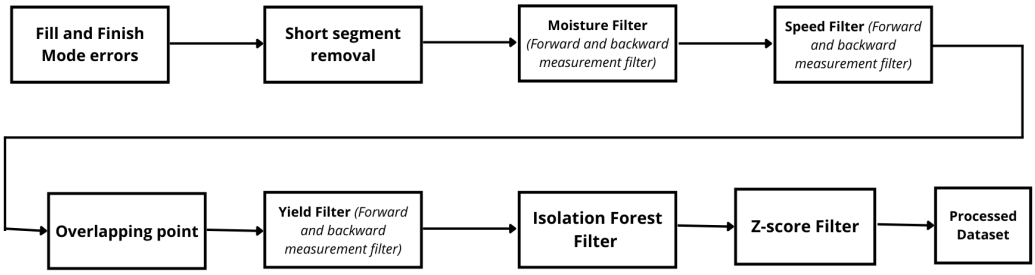

Figure 3.8: Yield Data Processing Pipeline

For the Alberta Precision Planter (Alb PP) field, imagery was acquired slightly earlier, on July 24, 2024 (DOY 206), corresponding to 55 days after planting and 84 days before harvest. Data were collected with the MicaSense RedEdge-MX Dual camera [75], also providing five multispectral bands (red, green, blue, near-infrared, and red-edge), with a spatial resolution of 2.96 cm.

3.4 Dataset

3.4.1 Feature Extraction

To transform raw UAV orthomosaic imagery into usable inputs for machine learning models, we generated a fishnet grid overlay across each field image. Following [55], who demonstrated the utility of aggregated pixel statistics for yield prediction and [73] who used a fishnet grid, we adopted a tiling approach in which each cell represents

Figure 3.9: [76] MicaSense RedEdge-P camera

Figure 3.10: [75] MicaSense RedEdge-MX Dual Camera

a spatial unit of approximately $5\,\mathrm{m} \times 5\,\mathrm{m}$. Tiles border data were extracted and store as geopandas [31] polygons.

Each tile was cropped from the orthomosaic with `rasterio` library. To ensure robustness, empty or invalid tiles were discarded. Pixel values within each valid tile were normalized to $[0, 1]$, and stored as multi-band arrays to support subsequent feature extraction and modeling.

Within each tile, we extracted descriptive statistics from every spectral band (Blue, Green, Red, Red Edge, and NIR). Specifically, we computed the mean, minimum and maximum of reflectance values, resulting in a feature vector characterizing the spectral distribution of each band.

In addition to raw spectral features, we derived vegetation indices that are widely has been proved to be good crop yield predictors [55]. Two indices were computed:

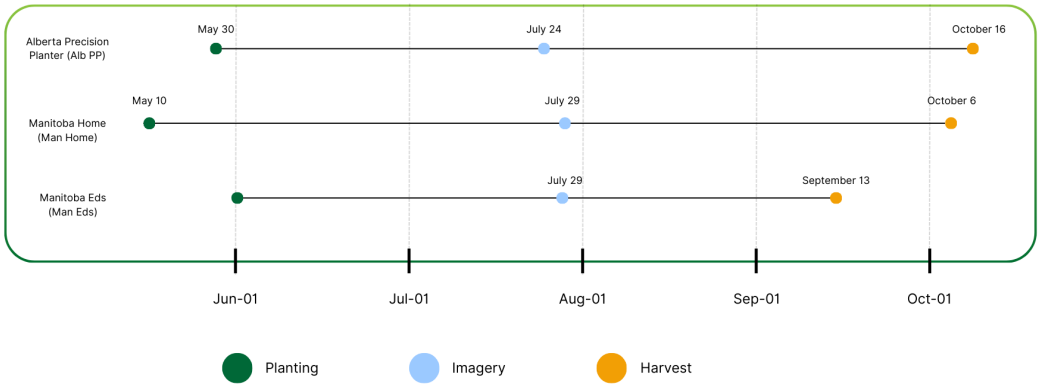

Figure 3.11: Timeline of planting, imagery acquisition, and harvest for the three fields.

- Normalized Difference Vegetation Index (NDVI), defined as :

$$\text{NDVI} = \frac{\rho_{\text{NIR}} - \rho_{\text{Red}}}{\rho_{\text{NIR}} + \rho_{\text{Red}}}$$

.

- Normalized Difference Red Edge Index (NDRE), defined as :

$$\text{NDRE} = \frac{\rho_{\text{NIR}} - \rho_{\text{RedEdge}}}{\rho_{\text{NIR}} + \rho_{\text{RedEdge}}}$$

.

For each index, tile-level aggregations (mean, min, max) were calculated using dedicated functions.

This procedure gave three distinct feature datasets:

1. **All-bands dataset:** statistical aggregates from each spectral band.
2. **NDVI dataset:** tile-level NDVI aggregates only.
3. **NDRE dataset:** tile-level NDRE aggregates only.

3.4.2 Data Fusion with Yield Measurements

To establish a supervised learning dataset, remote sensing features must be fused with ground-truth yield data. Yield measurements were collected via combine harvesters equipped with yield monitors, stored as georeferenced shapefiles. These point-based observations were integrated with the tiled imagery.We extracted and stored the raster tiles geospatial boundaries as polygons using `GeoPandas` [31]. Each tile's extent was represented as a polygon geometry, enabling spatial operations such as intersection with yield shapefile points. This vector representation ensured that image tiles and ground-truth yield measurements could be accurately aligned in a common coordinate reference system.

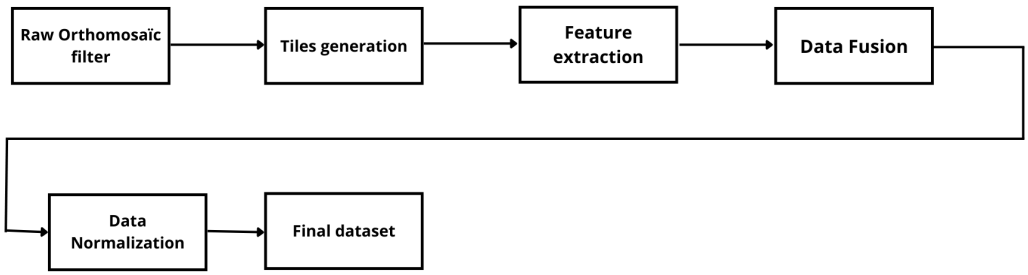

Figure 3.12: Dataset generation pipeline

For each tile polygon, yield points falling within its boundaries were identified through spatial joins. When multiple yield points overlapped a tile, their values were averaged to obtain a representative yield. This value was then assigned as the target variable for that tile. Tiles with no overlapping yield points were excluded from the dataset.

The resulting fused dataset, which pairs tile-level feature vectors with corresponding yield values, is illustrated in Figure 3.13.To facilitate cross-field experiments, features and targets were normalized using a configurable scheme, supporting either min-max scaling or z-score standardization of both input features and yield targets. For this study, both features and target were normalized using min-max normalization. The fused dataset pipeline is represented in Figure 3.12

blue_mean	blue_min	blue_max	green_mean	green_min	green_max		target	polygon	field_name
0.02526739	0.01686617	0.036084358	0.080017544	0.03387356	0.13068141		48.6903	POLYGON ((688533.3794000001	Alb_pp
0.03102661	0.01854997	0.045020625	0.07423267	0.0315308	0.1140618		39.82095	POLYGON ((688543.37164 5751	Alb_pp
0.02931512	0.01826987	0.042500257	0.082597576	0.041059442	0.1259739		21.4752	POLYGON ((688558.3600000001	Alb_pp
0.02891853	0.01792079	0.04210438	0.0812802	0.041632574	0.13045828	• • • •	35.660125	POLYGON ((688568.35224 5751	Alb_pp
0.03079423	0.02118346	0.045491066	0.07991135	0.04752804	0.12388313		26.25252	POLYGON ((688578.3444800001	Alb_pp
0.02787241	0.02022334	0.03746129	0.08186876	0.05074729	0.1224042		30.889575	POLYGON ((688593.33284 5751	Alb_pp
0.02711523	0.01761264	0.03600489	0.078550346	0.038166568	0.1195572		34.46932	POLYGON ((688603.3250800001	Alb_pp
0.02753723	0.01886581	0.036034398	0.07582929	0.047158666	0.10987919		31.87425	POLYGON ((688618.31344 5751	Alb_pp

Figure 3.13: Sample of final all-bands fused dataset

3.4.3 Training and Testing Dataset

To rigorously evaluate the generalization ability of machine learning models for crop yield prediction, multiple training and testing strategies were implemented. These strategies control how tiles from different fields, blocks, or clusters are allocated to training and validation, thereby simulating real-world scenarios.

3.4.3.0.1 Intra-field baseline. In the simplest setting, each field was treated independently. Tile-level samples were randomly split into training and test subsets (80/20), ensuring that both sets were drawn from the same spatial domain. This baseline represents the best-case scenario in which models are trained and tested under nearly identical field conditions.

3.4.3.0.2 Inter-field baseline. To evaluate cross-field transferability, training and testing were conducted across multiple fields. Each field dataset was splited into training and test subsets (80/20). Training subsets were merge together to form a unique training dataset, while testing was performed separately on each target field test subset. This strategy allows assessment of whether models trained on diverse environments can generalize to unseen fields.

3.4.3.0.3 Leave-One-Field-Out (LOFO). In this setting, models were trained on all fields except one, which was reserved entirely for testing. This scenario mimics real-world deployment where yield predictions must be generated for a new field with no prior training data. This allow to assess the model ability to generalize on unseen field.

3.4.3.0.4 Leave-One-Block-Out (LOBO). To address spatial autocorrelation within fields, we implemented a block-based cross-validation strategy. Fields were divided into square blocks (e.g., 25, 50, or 100 m) using tile polygons. At each iteration, all tiles from one block were used as the test set, while the remaining

blocks served as training data. This setup minimizes spatial dependence between training and test data, showing a more conservative and realistic estimate of model performance.

3.4.3.0.5 Leave-One-Cluster-Out (LOCO). Beyond regular grids, we also explored clustering-based cross-validation. Tiles were grouped into spatial or feature-based clusters using methods, in fact for this study we used K-means clustering. In each fold, one cluster was hold for testing, and the remaining clusters were used for training. This design evaluates the robustness of models under heterogeneous spatial or spectral conditions within a field.

These complementary training and testing setups allowed us to systematically compare intra-field accuracy, inter-field transferability, and robustness to spatial autocorrelation. Collectively, they provide a comprehensive framework for assessing the practical usefulness of yield prediction models in precision agriculture using only UAV imagery as feature.

3.5 Models

3.5.1 Overview

In this study, two supervised regression models were implemented to predict crop yield from UAV-derived features: a Random Forest Regressor (RF) and a Linear Regression model (LR). Both models were chosen to provide complementary perspectives: RF as a powerful non-parametric ensemble method capable of modeling non-linear feature–yield relationships, and LR as a simpler parametric baseline used for it interpretability and generalization ability.

3.5.2 Preprocessing and Normalization

Before training, all features and yield targets were normalized to ensure stable optimization and comparability across fields. Normalization was handled by a dedicated `NormalizationConfig` class supporting a **Min–max scaling**, mapping values to the $[0, 1]$ range. Normalization parameters were always fit on the training dataset and applied consistently to the test set. Predicted yields were subsequently denormalized before evaluation for interpretability purposes.

3.5.3 Random Forest Regressor

The RF model was implemented using `scikit-learn`'s `RandomForestRegressor`. Key hyperparameters included the number of trees ($n\_estimators = 250$), maximum depth (max_depth = 20), and max_features = $\sqrt{d}$ where d is the number of features. These parameters were selected to balance predictive power and computational cost.

RF is well-suited to high-dimensional and heterogeneous datasets such as UAV imagery because it aggregates multiple randomized decision trees, thereby capturing non-linear dependencies while reducing overfitting risk.

3.5.4 Linear Regression

The LR model was implemented as an Ordinary Least Squares regression using `scikit-learn`'s `LinearRegression`. This model assumes a linear relationship between features and yield, offering a transparent baseline against which more complex models can be compared. While less expressive than RF, LR provides valuable insights into the direction and magnitude of feature–yield associations through its regression coefficients.

3.5.5 Evaluation Metrics

Model performance was evaluated on both training and testing sets using the following metrics:

- **Root Mean Square Error (RMSE)** – measures overall predictive accuracy.
- **Mean Absolute Error (MAE)** – provides a scale-independent measure of error magnitude.
- **Mean Absolute Percentage Error (MAPE)** – evaluates relative prediction error as a percentage of true yield (with a small ϵ to avoid division by near-zero yields).
- **Coefficient of Determination (R^2)** – quantifies the proportion of yield variance explained by the model.
- **Pearson correlation coefficient (r)** – measures the linear association between predicted and observed yields, robust to scale differences.

Metrics were calculated using denormalized predictions to ensure interpretability in yield units. All results were logged and stored automatically for each experiment fold.

Chapter 4

Experimental Results

4.1 Experimental Setup

4.1.1 Hardware Configuration

All experiments in this study were performed locally. The machine used was an Apple iMac running macOS Ventura (version 13.1), equipped with the following specifications:

- Processor: Intel Core i7, 3.2 GHz, six physical cores. The six (06) core processor allowed us to run parallel experiments especially, for the random forest models.
- Graphics Processing Unit (GPU): Radeon Pro 555X with 2 GB of dedicated memory. This specification, while not exactly built for deep learning, was good enough for processing orthomosaics in QGIS and basic raster visuals. Many machine learning tasks in this work (Random Forest and Linear Regression) rely more on the CPU than on the GPU.
- Memory (RAM): 16 GB DDR4 at 2667 MHz. This is enough for the task required for this study : extracting tiles from large orthomosaics and running

experiments on large datasets. For tasks requiring more memory, data were processed in batches to avoid overloading system resources.

- Storage: 2 TB external hard drive. UAV orthomosaic imagery and yield shapefiles were stored on an external hard drive with sufficient capacity, since these can be large files. The 2 TB storage was also used for files generated by the code, such as experimental log files, graphs, generated datasets, and model outputs. Directories composed of structured folders were used to organize the data in a way that allows reproducibility.

4.1.2 Software Environment

Orthomosaic generation of UAV imagery was performed using Pix4D, a commercial software specialized in drone photogrammetry mapping. Pix4D was used to process raw UAV images into georeferenced orthomosaics, which then served as the primary input for this study.

Geospatial data were managed using QGIS LTR (Long-Term Release), an open-source Geographic Information System (GIS).

Pre-processing, feature extraction, and model training pipelines were developed in Python 3. We used Visual Studio Code (VS code) integrated development environment. VS Code was selected for its integrated features, which support Python code editing and execution.

The following libraries were used :

- `numPy` and `pandas` for numerical operations and data manipulation.
- `geoPandas` and `Shapely` for geospatial data processing and management.
- `rasterio` for raster reading, tiling, and pixel-level value extraction.
- `scikit-learn` for model training, evaluation metrics, and cross-validation.

- `matplotlib` and `seaborn` for data visualization, exploratory analysis, and experimental result plotting.
- `PyYAML` for configuration management, ensuring that experimental settings were consistently stored and reusable.

4.1.3 Experimental Configuration and logging

To enable reproducibility, we used configuration files (config.yaml), which stored hyperparameters, normalization strategies, and dataset paths. This allowed experiments to be reproduced under identical conditions without manual reconfiguration.

Experiments produced log files that made it possible to interpret results, track execution progress, and visualize outputs. The logs generated included:

- Training logs (.txt files), which contain hyperparameters, start times, evaluation metrics (RMSE, MAE, MAPE, r2, correlation), and additional runtime information.
- Prediction outputs (.csv files), which saved true and predicted yield values for both training and test sets, enabling statistical analysis and visualization after experiments.

Intermediate and final outputs were organized in structured directories (e.g., `/logs`, `/predictions`, `/results`).

4.2 Numerical Results

This section discusses experimental results addressing challenges in UAV-based crop yield prediction, focusing especially on evaluation methodologies. These experiments aim to achieve the following objectives :

1. develop and compare machine learning models (Linear Regression, Random Forest) for yield prediction using UAV multispectral imagery;
2. construct and evaluate tile-based datasets by fusing UAV orthomosaics with yield data across diverse agricultural conditions;
3. implement and assess multiple evaluation strategies : traditional non-spatial strategies (intra-field, inter-field) with spatial-based strategies (lofo, LOBO, LOCO).
4. compare the LOBO and LOCO methodologies
5. compare predictive performances of vegetation indices and raw multispectral bands.
6. analyze the impact of spatial autocorrelation on model performance and transferability.

We used three (03) agricultural fields located in different geographic regions, carrying different soil types, climatic conditions, and management practices. UAV multispectral orthomosaics were combined with harvester yield data to create suitable datasets for the experiments that are to be held. Model performance was evaluated using various complementary metrics, including coefficient of determination (R^2), mean absolute percentage error (MAPE), correlation coefficient (R), root mean squared error (RMSE), mean absolute error (MAE).

The results are organized progressively from what we call baseline scenarios (non-spatial intra-field random splitting) to scenarios that incorporate spatial constraints.

4.2.1 Intra-Field Baseline

4.2.1.1 Dataset

The intra-field baseline experiment evaluated model performance when both training and testing datasets were drawn from the same field. The data were split using the

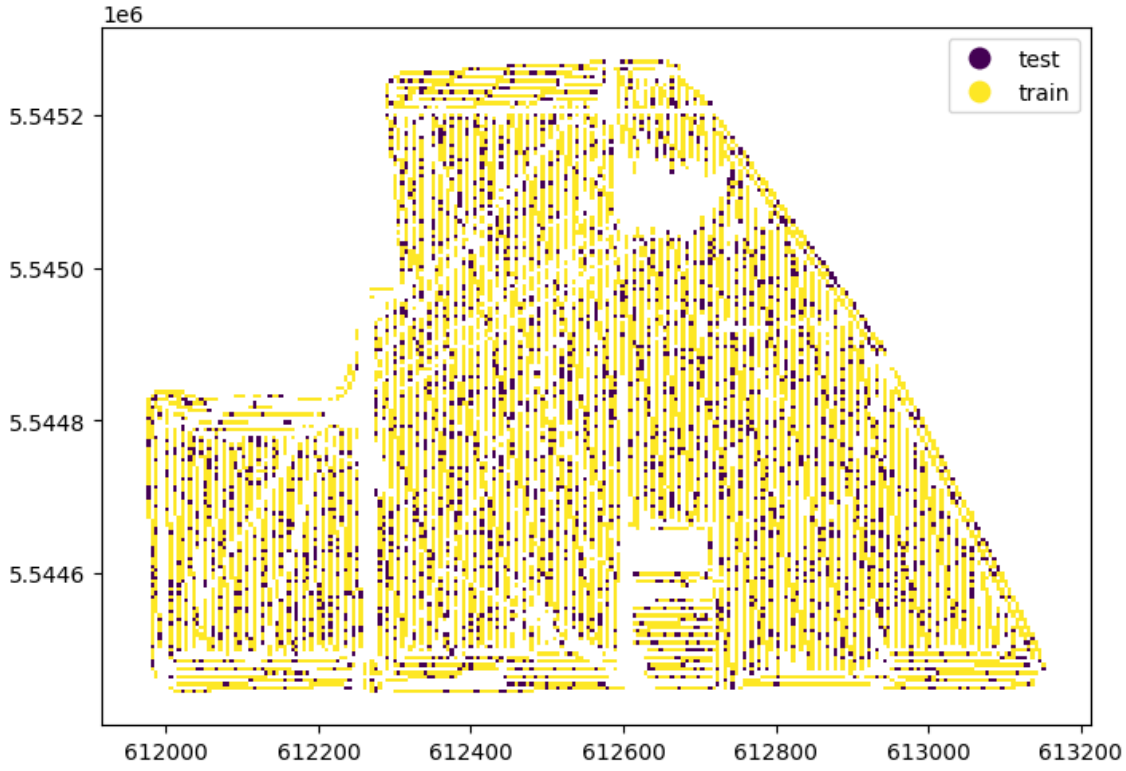

Figure 4.1: Spatial distribution of training and testing points for the Man Eds intra-field baseline experiment.

`train_test_split` function from scikit-learn [84], with 80% allocated to training and 20% for testing. This approach is widely used in machine learning evaluations; however, it doesn't take account of spatial correlation, as training and testing points may be geographically close, as illustrated in Figure 4.1. Such proximity leads to overoptimistic performance estimates by allowing the model to exploit local autocorrelation patterns [55].

4.2.1.2 Results & Discussions

The results for each field are summarized in Table 4.1. Several important observations can be drawn from it:

Table 4.1: Intra-field baseline results

		RMSE		MAE		MAPE		R^2		R	
Field	Feature	LR	RF	LR	RF	LR	RF	LR	RF	LR	RF
Alb_pp	all_bands	7.690	7.290	5.927	5.473	0.190	0.176	0.374	0.437	0.614	0.662
	ndre	8.080	8.428	6.251	6.462	0.204	0.210	0.309	0.248	0.561	0.516
	ndvi	8.380	8.505	6.499	6.500	0.210	0.209	0.257	0.234	0.512	0.508
Man_eds	all_bands	6.994	6.428	5.282	4.744	0.102	0.092	0.214	0.336	0.464	0.581
	ndre	7.252	7.228	5.547	5.490	0.108	0.106	0.155	0.161	0.395	0.409
	ndvi	7.277	7.069	5.548	5.327	0.107	0.103	0.149	0.197	0.387	0.450
Man_home	all_bands	13.328	12.598	9.740	9.244	0.080	0.076	0.475	0.531	0.690	0.730
	ndre	14.026	14.208	10.591	10.550	0.087	0.086	0.418	0.403	0.647	0.637
	ndvi	14.899	14.789	11.023	11.031	0.091	0.090	0.344	0.354	0.589	0.599

4.2.1.2.1 Random Forest consistently outperformed Linear Regression. In most fields and feature sets, Random Forests (RF) produced higher R^2 values and lower errors (like RMSE, MAE and MAPE). Depending on the field and features applied to, its R^2 performance ranged from 16 to 53%. This makes sense, since RF has an ability to model complex non-linear relationships between spectral reflectance and yield data that linear regression (LR) may fail to capture at a certain point.

4.2.1.2.2 Predictability differs by field. Man Home was the most efficient field, producing the best score performance with the highest R^2 (0.53) and the lowest MAPE (∼7%). This reflects a stable relationship between spectral features and the yield. Alb PP come to the second position in terms of variance explained, achieving a decent R^2 score, its best R^2 score was `0.43` with RF. In contrast, Man Eds produced lower R^2 values but achieved lower MAPE, indicating more precise predictions despite limited variance explained. Man Eds models performing lower R^2 values but decent relatively low MAPE shows the importance of combining multiple metrics to interpret results. If the goal study is the precision of prediction, Man eds would present decent results, but it falls short in capturing yield variability across the field. It is important to have a balanced trade-off between precision and variance explained.

4.2.1.2.3 Multispectral bands outperformed vegetation indices. Models trained with all the raw bands as features (red, green, blue, red edge, near-infrared) achieved better results than those trained on only vegetation indices, such as NDVI or NDRE, both for LR and RF. For example, with LR on the Man Home field, the MAE was 9.740 when using all bands, compared to 14.208 with NDRE and 14.789 with NDVI.

4.2.1.2.4 Linear Regression performed best with NDRE features. Linear Regression applied to NDRE features gives its best performances. NDRE is the feature holding the most linear pattern, occasionally allowing LR to outperform RF (e.g., Man Home: $R^2 = 0.418$ vs. 0.403 and Alb PP MAPE = 20% vs. 21%).

4.2.1.3 Experiment Conclusion

Based on the discussion in Section 4.2.1.2, the following conclusions can be drawn:

- **Random Forest is the strongest intra-field baseline model**.
- **Field-specific variability matters**. Even if there are distinguishable patterns across almost all field performances, the results vary most depending on the field. This implies that field-specific parameters, such as crop type, data acquisition parameters (e.g., camera specifications, imagery acquisition dates, planting dates, etc.) matter in order to achieve good performance.
- **Combined spectral bands outperform vegetation indices.** Using all multispectral bands provided better results than vegetation indices (NDVI, NDRE), demonstrating the value of richer spectral information for yield prediction.
- **Linear Regression works well with NDRE features**. LR models paired with NDRE inputs occasionally rivalled or surpassed RF, showing that NDRE has a stronger linear link with yield than NDVI. The red-edge spectral region

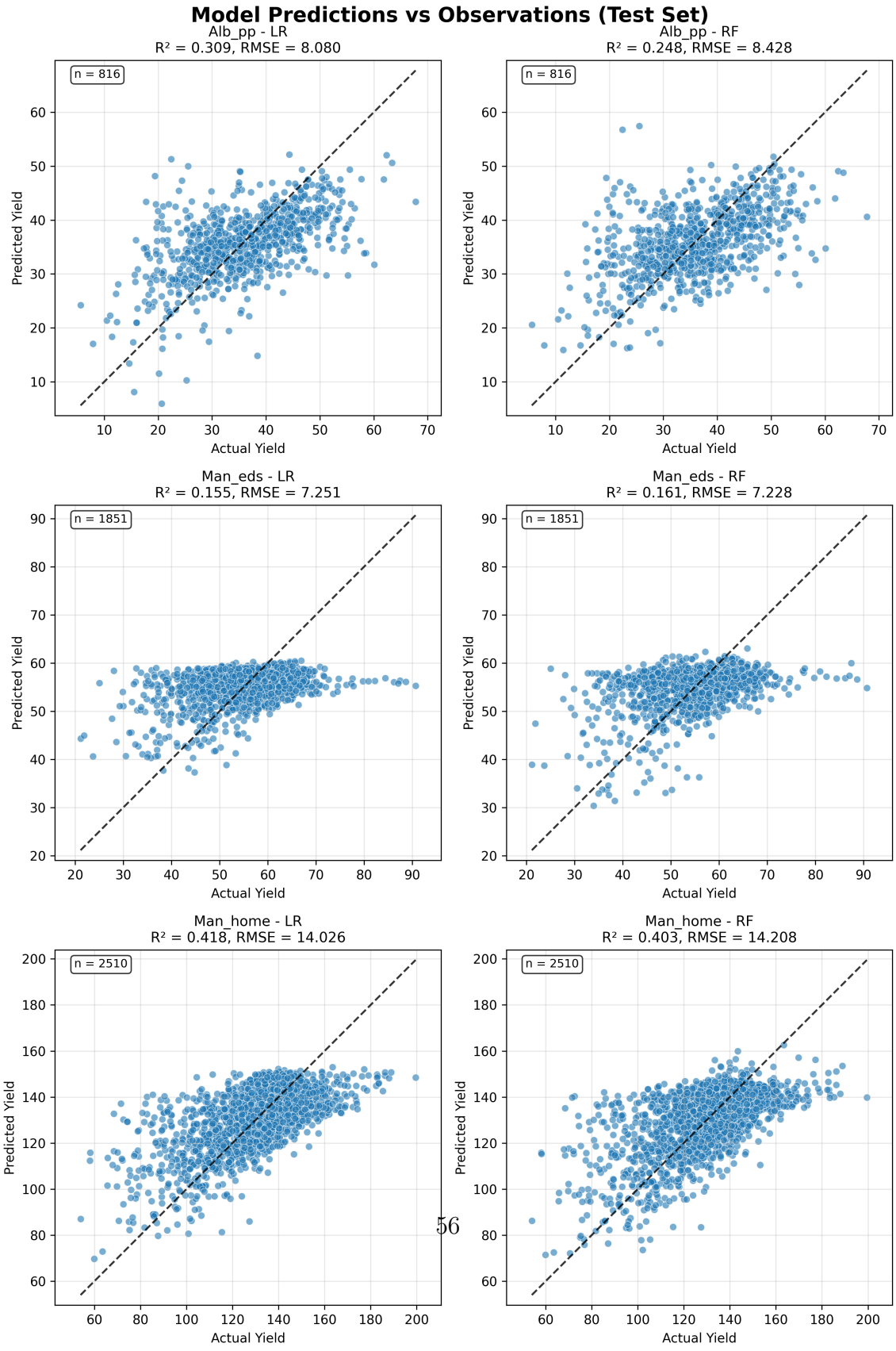

Figure 4.2: Predicted vs. actual yield values in the intra-field baseline using NDRE.

provides relevant information that aligns well with linear modelling assumptions, since the main difference between NDVI and NDRE is the substitution of the red band with the red-edge band in the formula.

4.2.2 Inter-Field Baseline

4.2.2.1 Dataset

In this experimental setup, each field dataset was first split into training (80%) and testing (20%) subsets using the same procedure as in the intra-field baseline. The three training subsets were then merged to form a combined training dataset, while each field's testing subset was kept separate. This configuration allowed models to be trained on a broader, multi-field dataset while still being evaluated independently on each field. The purpose of this setup is to assess whether pooling information across fields improves generalization performance within individual fields.

4.2.2.2 Results & Discussions

The results are presented in Table 4.2. Several key observations can be drawn:

4.2.2.2.1 Random Forest demonstrates cross-field generalization. RF achieved higher R^2 values and lower error metrics compared to LR across all fields. For example, in the challenging Man Eds field, RF with all bands reached $R^2 = 0.30$ and a correlation of $R = 0.55$, whereas LR completely collapsed with $R^2 = -23.03$. This highlights RF's ability to capture non-linear patterns that transfer across fields, even when spectral–yield relationships vary substantially.

4.2.2.2.2 Linear Regression struggled with negative R^2 values. LR failed to generalize across most fields, producing highly negative R^2 scores (e.g., -23.03 in Man Eds with multispectral bands, -15.32 in Man Eds with NDVI, and -4.33 in

Man Home with multispectral bands).

4.2.2.2.3 NDVI consistently underperformed. NDVI produced the lowest predictive power, often with extreme negative R^2 values (e.g., -1.06 in Alb PP and -15.32 in Man Eds with LR).

4.2.2.2.4 Field-dependent differences remain evident. Man Home benefited most from inter-field training, achieving $R^2 = 0.51$ with RF and all bands, while Alb PP and Man Eds showed weaker improvements. These differences highlight that even when pooling data across multiple fields, predictability is still shaped by local factors such as crop type, soil properties, and environmental conditions.

4.2.2.3 Experimental Conclusion

From the inter-field baseline experiments discussed in 4.2.2.2, the following lessons can be drawn:

- **RF is more robust to cross-field variability when LR prediction fails to generalize across heterogeneous environments** .
- **NDVI is not suitable for cross-field generalization**.
- **Field-specific conditions dominate generalization.** Even when trained on multiple fields, performance remained field-dependent, with Man Home being the most predictable and Man Eds the least. This shows the importance of incorporating field context (e.g., management, soil, crop type) when attempting multi-field generalization.

Table 4.2: Inter-field baseline results

		RMSE		MAE		MAPE		R^2		R	
Field	Feature	LR	RF	LR	RF	LR	RF	LR	RF	LR	RF
Alb_pp	all_bands	10.726	7.308	8.319	5.484	0.267	0.176	−0.218	0.435	0.414	0.659
	ndre	8.842	8.573	6.651	6.532	0.219	0.215	0.172	0.222	0.557	0.513
	ndvi	13.935	8.556	10.762	6.529	0.336	0.211	−1.056	0.225	0.515	0.500
Man_eds	all_bands	38.677	6.617	37.865	4.884	0.724	0.095	−23.032	0.297	0.101	0.546
	ndre	8.983	7.478	6.885	5.651	0.133	0.109	−0.296	0.102	0.332	0.367
	ndvi	31.874	8.613	29.909	5.971	0.565	0.118	−15.322	−0.192	0.336	0.239
Man_home	all_bands	42.470	12.914	39.350	9.464	0.286	0.078	−4.331	0.507	0.579	0.713
	ndre	14.961	14.488	11.407	10.795	0.093	0.088	0.338	0.380	0.596	0.624
	ndvi	26.930	16.097	24.210	11.840	0.178	0.097	−1.143	0.234	0.591	0.570

4.2.3 Leave-One-Field-Out (LOFO)

4.2.3.1 Dataset

The LOFO experiments were designed to evaluate model performance when predicting yield in a completely unseen field, i.e., one excluded from training. This configuration evaluates cross-field generalization under a field with different environmental and geographical parameters, such as soil composition, geographic location, climate, yield distribution, and even crop type (corn versus canola). For each run, the training set consisted of either one or both of the other fields, while the excluded field served exclusively as the test set. For example, when evaluating the Alb_PP (Alberta Precision Planter) field, models were trained using Man Eds and Man Home data.

This strategy has been adopted in precedent studies [33, 55], but typically with geographically relatively close fields from the same site, where cross-validation results were being averaged across folds to output a single result. This work applies LOFO to fields separated by large distances and characterized by diverse conditions, providing sufficient grounds for evaluating field transferability

4.2.3.2 Results & Discussions

The results, reported in Tables 4.3 and 4.4, show that models perform very poorly when transferred across fields. Even with the richer combination of all the spectral bands as features set, generalization falls short:

4.2.3.2.1 Severe negative R^2 values. When transferring from Alb PP to Man Home, RF yielded $R^2 = -38.18$ and LR $R^2 = -26.92$. Similarly, transferring from Man Home to Alb PP resulted in catastrophic values ($R^2 = -42.72$ for RF and $R^2 = -714{,}473$ for LR), reflecting a complete failure to capture yield variability under domain change.

4.2.3.2.2 High relative errors. MAPE values frequently exceeded 70% and, in some cases, 200% (e.g., Man_Home → Alb PP), indicating that models made large absolute errors despite sometimes regressing towards mean yield levels.

Table 4.3: LOFO test results for `all_bands` feature type (lower is better for RMSE/MAPE; higher is better for R^2).

Source	Target	FeatureType	RMSE(LR)	RMSE(RF)	MAPE(LR)	MAPE(RF)	R^2(LR)	R^2(RF)
Alb pp	Man eds	all_bands	22.476	40.287	0.369	0.704	-6.953	-24.554
Alb pp	Man home	all_bands	99.087	117.395	0.731	0.876	-26.915	-38.184
Alb pp+Man eds	Man home	all_bands	78.567	87.680	0.568	0.643	-16.550	-20.858
Alb pp+Man home	Man eds	all_bands	76.170	41.731	1.427	0.778	-90.347	-26.418
Man eds	Alb pp	all_bands	2468.560	23.762	65.623	0.790	-61814.195	-4.728
Man eds	Man home	all_bands	86.301	88.735	0.637	0.652	-20.176	-21.387
Man eds+Man home	Alb pp	10500.603	26.080	325.244	0.868	-1118500.137	-5.900	
Man home	Alb pp	8392.461	65.646	237.842	2.143	-714473.300	-42.715	
Man home	Man eds	30.997	50.546	0.538	0.943	-14.127	-39.226	

4.2.3.3 Experimental Conclusion

Based on the discussion in Paragraphs 4.2.3.2.1 and 4.2.3.2.2, the following conclusions can be drawn:

Table 4.4: LOFO test results for `ndre` feature type (lower is better for RMSE/MAPE; higher is better for R^2).

Source	Target	FeatureType	RMSE(LR)	RMSE(RF)	MAPE(LR)	MAPE(RF)	R^2(LR)	R^2(RF)
Alb pp	Man eds	ndre	11.367	12.840	0.169	0.192	-1.034	-1.596
Alb pp	Man home	ndre	40.986	82.514	0.285	0.604	-3.776	-18.358
Alb pp+Man eds	Man home	ndre	29.292	72.400	0.196	0.521	-1.440	-13.903
Alb pp+Man home	Man eds	ndre	9.200	13.156	0.137	0.199	-0.333	-1.725
Man eds	Alb pp	ndre	19.894	21.104	0.657	0.689	-3.014	-3.518
Man eds	Man home	ndre	61.520	70.198	0.441	0.503	-9.761	-13.011
Man eds+Man home	Alb pp	9.299	21.057	0.253	0.687	0.123	-3.498	
Man home	Alb pp	51.129	61.500	1.650	1.999	-25.518	-37.367	
Man home	Man eds	34.894	41.810	0.641	0.777	-18.170	-26.522	

- **Cross-field transfer is extremely limited.** Neither LR nor RF could generalize yield–spectral relationships to unseen fields, with R^2 values often highly negative and MAPE exceeding 70–200%.
- **Field-specific factors dominate.** Large differences in crop type, soil conditions, and acquisition parameters undermine model transferability, emphasizing the need for domain adaptation or fine-tuning strategies in practical multi-field yield prediction.

4.2.4 Leave-One-Block-Out (LOBO)

4.2.4.1 Dataset

The LOBO configuration is designed as an intra-field spatial evaluation strategy to assess the ability of models to generalize across distinct spatial regions within the same field. Unlike the random intra-field baseline split, LOBO takes into account the spatial dependence of data: each field is divided into square blocks of fixed size, as illustrated in Figure 4.3, and the model is trained on $(n-1)$ blocks before being tested on the remaining unseen block. This procedure is repeated, allowing every block to be tested until each block has served once as the test set. The resulting metrics are then aggregated (mean, median, standard deviation, minimum, maximum) for

interpretation purposes.

Experiments were conducted with different block sizes of 50, 100, 200, 300, 400, and 500 m, to help analyze the effect of the block size on prediction accuracy and model performance globally.

4.2.4.2 Results & Discussions

The LOBO experiments produced lower performance compared to the intra-field baseline, showing the difficulty of spatial generalization (Figure 4.4) . Field-level trends and block-size effects can be summarized as follows:

4.2.4.2.1 Alb pp and Man eds: poor performances. Both Alb pp and Man eds displayed poor results under LOBO. R^2 values were predominantly negative across most block sizes, suggesting that models failed to capture yield variability when evaluated on spatially independent blocks. Alb pp showed occasional improvement, reaching a modest $R^2 = 0.13$ at 300 m and slight gains at 450 m, but results remained unstable. Man eds performed even worse, with negative R^2 across all scales. This implies that the spectral information provided for these fields was insufficient for the models to capture block heterogeneity.

4.2.4.2.2 Man home: more stable and predictable. Conversely, Man home performances are stable under LOBO. Here, R^2 values improved according to block size, plateauing around 250–300 m. The best performance was achieved at 400 m with $R^2 = 0.36$. RMSE values increased with block size, as expected, but remained within reasonable bounds. Linear Regression (LR) paired with NDRE features slightly outperformed Random Forest (RF) demonstrating that even in that spatial configuration, LR is well suited to NDRE features.

4.2.4.2.3 Block-size trade-offs. Block size has a strong influence on model performances:

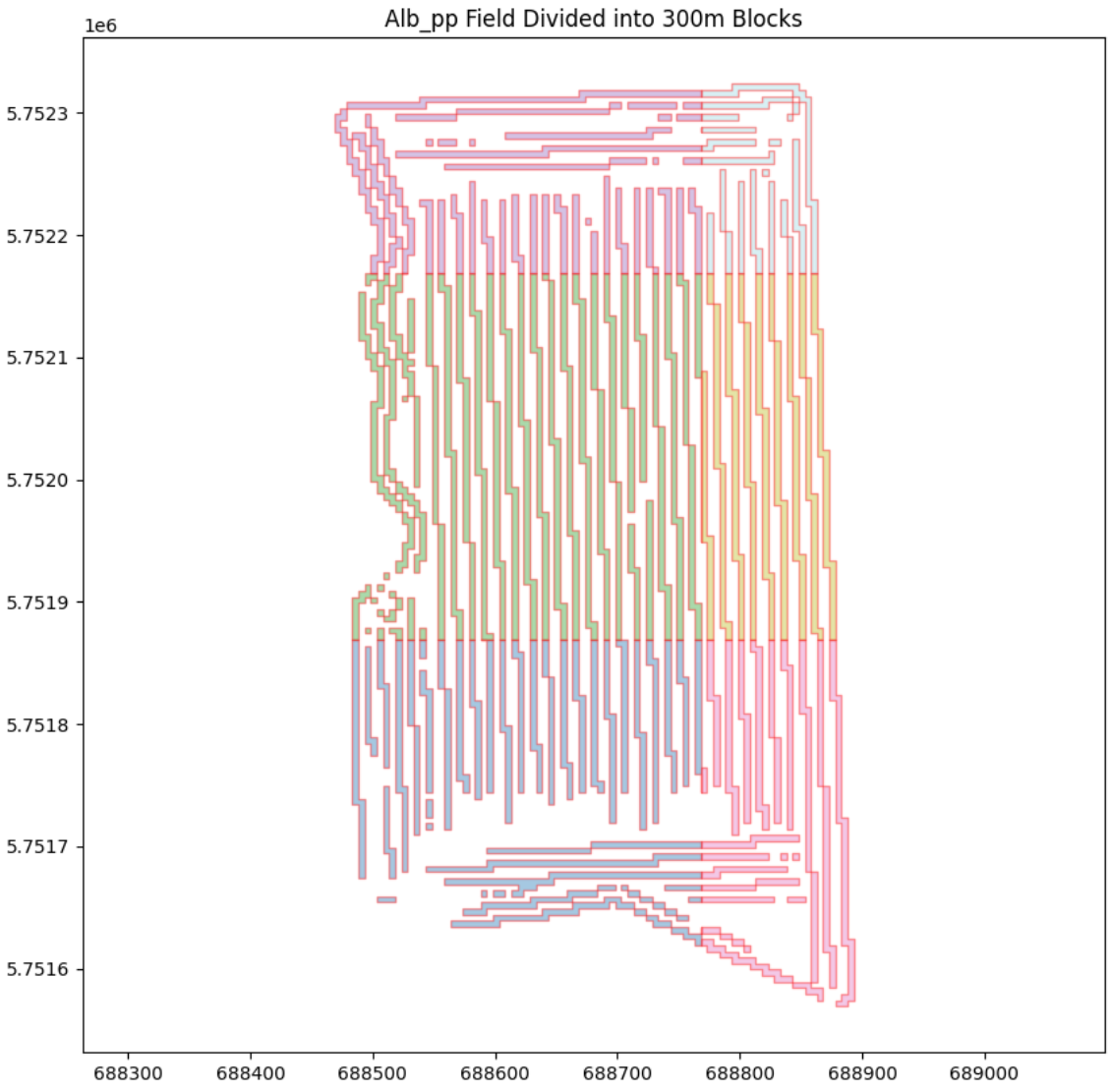

Figure 4.3: Spatial distribution of training and testing points for the Alb PP LOBO experiment.

- Small blocks (<200 m) give low RMSE values due to limited within-block variability, but negative R^2, indicating failure to capture variance.
- Large blocks (≥300 m) have improved R^2 but higher RMSE and MAPE values due to increased heterogeneity within blocks.

This confirms that larger blocks do not automatically lead to better generalization. Instead, each field has an optimal block size—around 200–300 m for Man home—where variance explanation (R^2) was maximized without excessively increasing error metrics.

4.2.4.3 Experimental Conclusion

Based on the discussion in Section 4.2.4.2, the following conclusions can be drawn:

- Small blocks underestimate prediction difficulty by reducing variability.
- Large blocks increase prediction errors due to blocks heterogeneity.
- An optimal block size exists, showing that the way data are distributed across blocks can lead to better performance. This aspect could be studied more deeply.
- Field factors dominate over block-size effects, with some fields enabling efficient prediction and others not at all even with different block sizes.

Overall, the LOBO experiment really underscored how much spatial autocorrelation affects yield prediction. Using block-based validation gave us a more realistic testing scenario compared to random splits—but it also exposed some of the limits of traditional models like Linear Regression and Random Forest. They just don't do well at capturing the detailed spatial patterns in yield across a field.

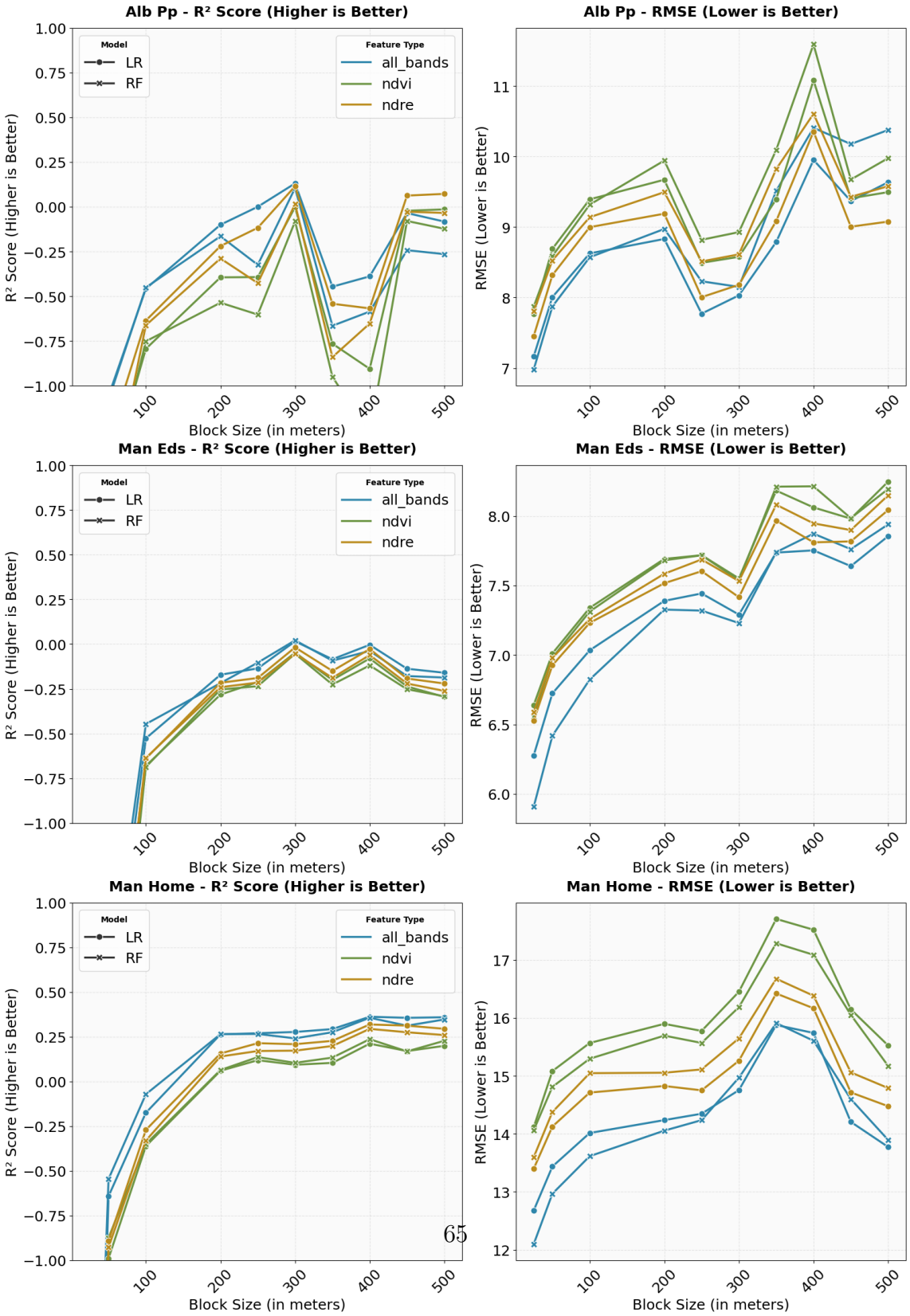

Figure 4.4: LOBO R^2 and RMSE mean results across block sizes.

4.2.5 Leave-One-Cluster-Out (LOCO)

4.2.5.1 Dataset

he LOCO setup is another way of doing intra-field spatial validation, but instead of dividing data into fixed blocks like LOBO, it groups observations into clusters (Figure 4.5). To build those clusters, we used the centroid coordinates of each polygon and ran K-means with several different values of k $(3, 5, 8, 10, 15, 20)$. For each run, the model was trained on $(k-1)$ clusters and tested on the one left out. This setup allows us to assess how performance varies across spatial partitioning, from few clusters to finer many clusters.

4.2.5.2 Results & Discussions

The LOCO results (Figure 4.6) show some clear trends across fields and feature types:

4.2.5.2.1 Effect of cluster number. Model performance improved when the number of clusters was small, with higher R^2 values. As the number of clusters increased, performance generally declined. This indicates that excessive fragmentation reduces the representativeness of training data for the left-out cluster, making generalization more difficult.

4.2.5.2.2 Linear regression (LR) outperforms Random Forest (RF). Linear Regression (LR) often outperformed Random Forest (RF) under LOCO, particularly with vegetation index features. This is a contrast compared with intra-field baseline setups, where RF dominated. This implies that RF is sometimes overoptimistic when spatial considerations are not taken into account, unlike LR.

4.2.5.2.3 Multispectral bands have the best results. Multispectral bands still carry the best results across fields, followed by NDRE and then NDVI. This

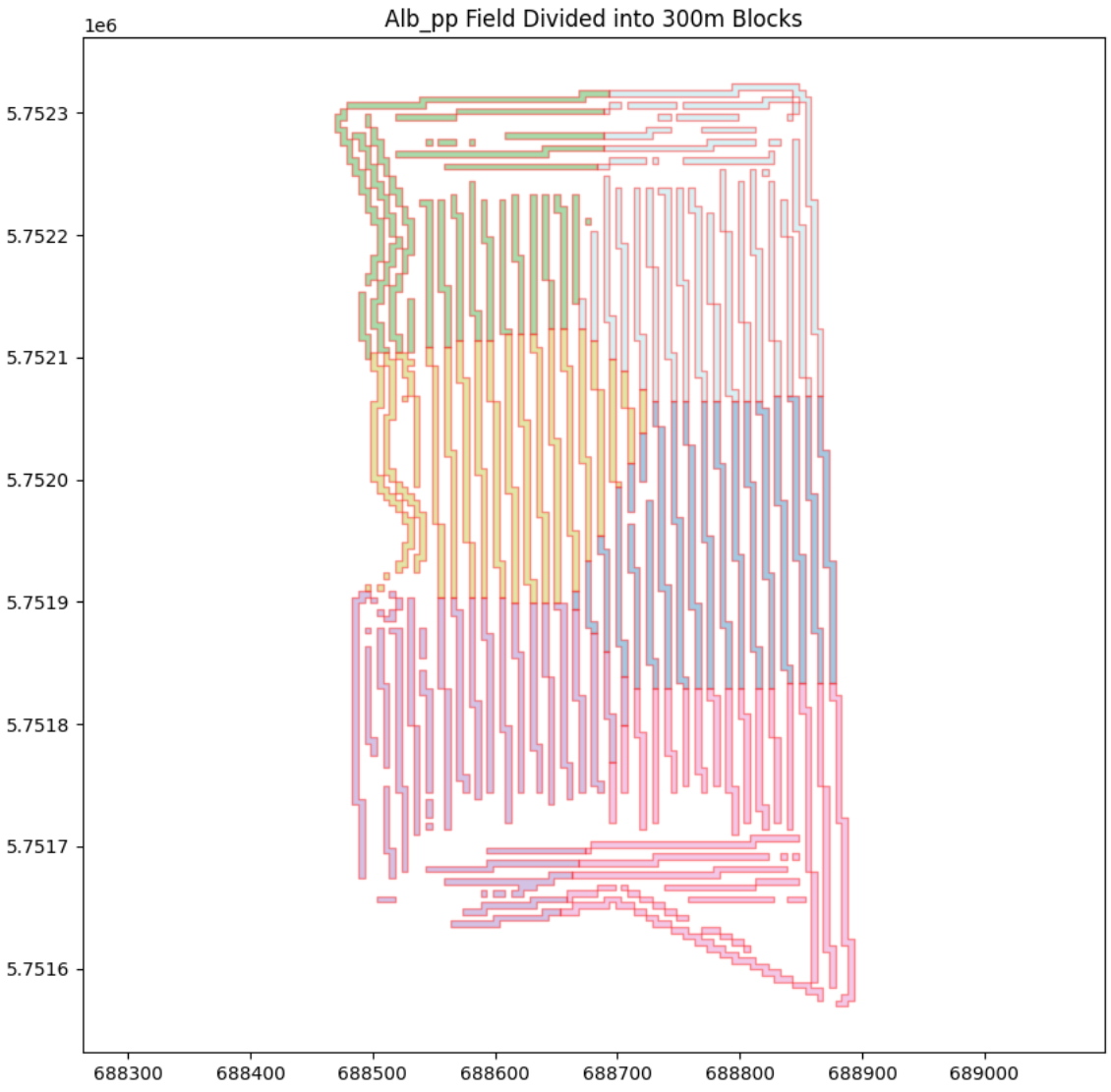

Figure 4.5: Spatial distribution of training and testing points for the Alb PP LOCO experiment.

trend confirms the advantage of using richer spectral information, especially under spatially constrained validation.

4.2.5.2.4 Field-specific performance. Performance varied strongly by field:

- **Man home**: Best overall performance, reaching $R^2 = 0.41$ at 3 clusters. Results declined gradually with increasing cluster count, but remained more stable than in other fields.
- **Alb pp**: Achieved modest success with $R^2 = 0.12$ at 5 clusters, but performance was inconsistent and dropped considerably at higher k.
- **Man eds**: Best performance occurred with 3 clusters ($R^2 = 0.18$), but performance dropped quickly beyond this point, highlighting sensitivity to cluster number.

4.2.5.3 Experimental Conclusion

Based on the discussion in Section 4.2.5.2, the following conclusions can be drawn:

- Few clusters provide more stable and interpretable predictions than a large number of clusters. Unlike LOBO, a discernible trend is present between performance and the number of clusters across fields.
- Linear models (LR) adapt better than non-linear RF under cluster-based CV, especially with vegetation indices.

4.2.6 Comparison of LOBO and LOCO

Both LOBO and LOCO were designed to address the issue of spatial autocorrelation, but they take different approaches to enforcing spatial independence. Their comparison shows some differences in performance patterns and methodological implications.

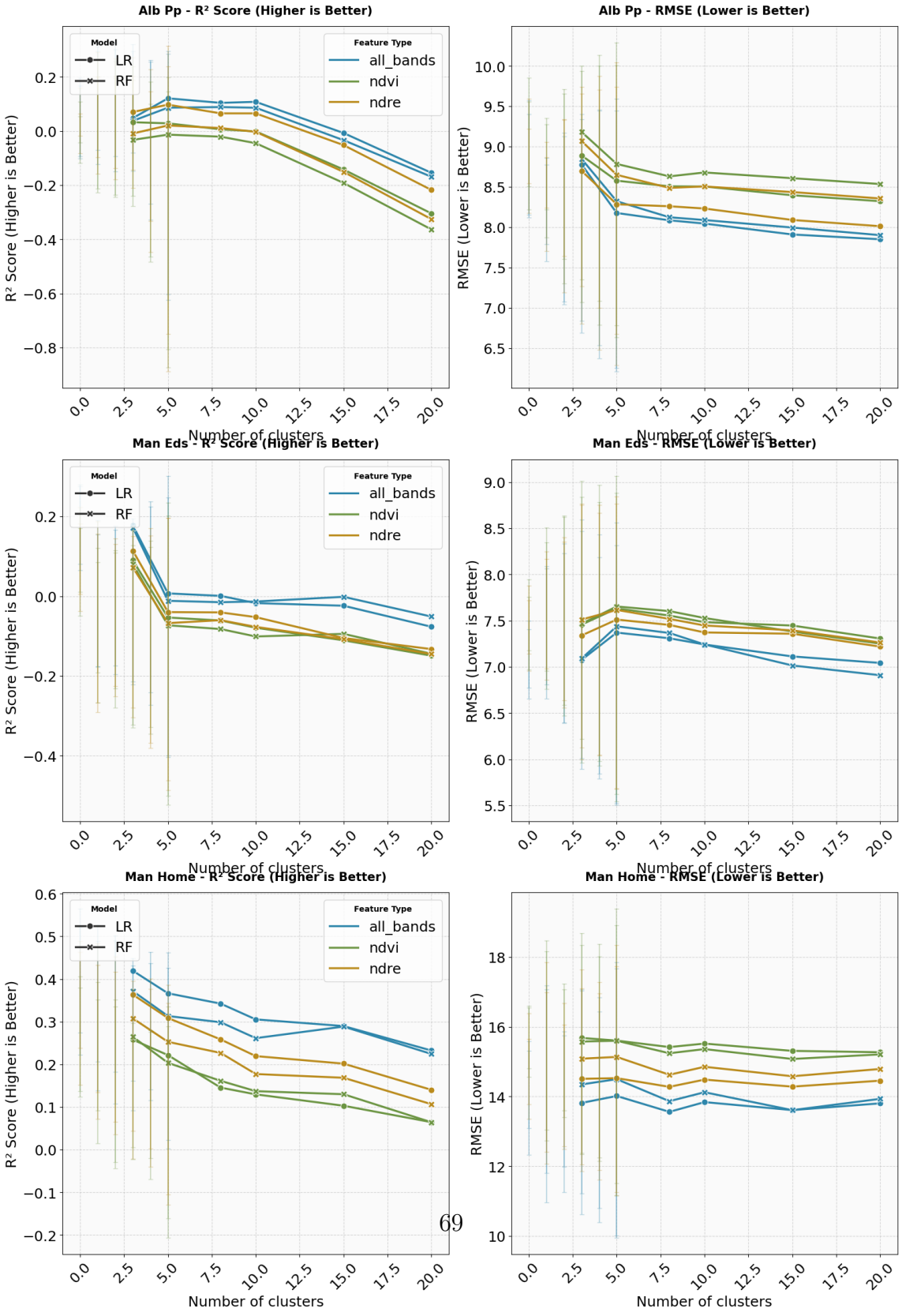

Figure 4.6: LOCO R^2 and RMSE mean results across different numbers of clusters.

4.2.6.0.1 Overall performance. LOCO generally have higher R^2 values and more stable performance than LOBO. For instance, Man home achieved $R^2 = 0.41$ under LOCO (3 clusters) compared to $R^2 = 0.36$ under LOBO (400 m). Alb pp and Man eds, which performed poorly in LOBO (mostly negative R^2 values), showed modest but more consistent improvements under LOCO (e.g., Alb pp: $R^2 = 0.12$ at 5 clusters; Man eds: $R^2 = 0.18$ at 3 clusters).

4.2.6.0.2 Effect of partitioning strategy. LOBO relies on fixed square blocks, which may split heterogeneous areas arbitrarily. LOCO, in contrast, groups spatially similar polygons through clustering. In LOBO, no discernible pattern emerged between performance and block size, though the results hinted that there might be an optimal size worth exploring. In contrast, LOCO revealed a clearer trend, with performance varying with the number of clusters. These findings highlight that the effectiveness of spatial validation depends strongly on how data are grouped. Choosing the right partitioning strategy really matters. This aspect could be the subject of further investigation.

4.2.6.0.3 Model behavior. In LOBO, Random Forest and Linear Regression performed similarly, with RF occasionally showing instability. In LOCO, where model performance are better, LR most often outperformed RF, particularly when using vegetation indices. This suggests that LOCO favors models that capture broad linear relationships, while RF can be over-optimistic by exploiting spatial autocorrelation.

4.2.6.0.4 Field dependence. No matter how we ran the tests, field-specific effects were clear. Both methods confirmed strong field-specific effects: Man home outperformed the other fields, regardless of validation strategy. Alb pp and Man eds were much harder to predict, although LOCO helped reduce some of the instability observed in LOBO.

In summary, while LOBO highlights the limitations of current models under strict

spatial independence, LOCO suggested that the performance of spatially aware experimentation can depend on the methodology used to partition the field. Together, these two strategies offer complementary insights into the impact of spatial autocorrelation on crop yield prediction.

4.2.7 Overall Takeaways

The experimental results across all configurations reveal many key insights into UAV-based crop yield prediction:

- **Model choice matters, but is not enough.** Random Forest outperformed Linear Regression in non-spatial setups, confirming its ability to capture non-linear yield–spectral relationships. However, both models struggled under spatially independent settings (LOFO, LOBO). Even under improved spatial split methodologies, LR sometimes performed better than RF, suggesting that RF can occasionally be over-optimistic. This shows that increasing model complexity alone cannot overcome domain shifts or spatial autocorrelation.
- **Spectral information is a major asset.** Multispectral bands (RGB, red-edge, NIR) outperformed vegetation indices. NDRE occasionally allowed LR to perform competitively, but NDVI was the weakest feature set, especially under inter-field and spatial validation. Richer spectral information is therefore essential for robust prediction.
- **NDRE showed the strongest linear link with yield**. Linear Regression achieved its best results when NDRE was used as the input.
- **Field-specific variability dominates outcomes.** Man Home was the most predictable field, achieving the highest R^2 values across nearly all experiments. Alb pp and Man Eds, in contrast, were harder to model, with frequent negative R^2 in spatially independent settings. These results emphasize the strong influence of local parameters like soil, crop, and management conditions.

- **Field transferability remains a challenge**. Using only UAV spectral information to achieve model transfer across fields remains difficult and requires additional parameters

- **Non-spatial experiments overestimate performance.** Intra- and inter-field baselines yielded promising results (RF with $R^2 \sim 0.5$), but performance dropped sharply when spatially independent methods (LOFO, LOBO, LOCO) were applied. This confirms that random splits take advantage spatial autocorrelation, providing over-optimistic estimates that do not reflect real-world generalization.

- **Spatial autocorrelation remains the central challenge.** Once we enforced spatial independence, predictive power dropped sharply—no matter which features or models were used. This shows just how sensitive yield prediction is to spatial effects, and addressing autocorrelation (through domain adaptation, fine-tuning, or spatially aware models) is critical for practical deployment.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

This research explored the potential of UAV-based multispectral imagery for crop yield prediction across multiple agricultural fields. Data were collected from three fields: two in Manitoba—Ed's and Home—growing canola and corn respectively, and one in Alberta (Alb PP) growing canola. Orthomosaics were generated from UAV imagery and fused with yield data to construct a comprehensive dataset. This dataset was used to train and evaluate two machine learning models, namely Linear Regression (LR) and Random Forest (RF).

The main contributions of this thesis can be summarized as follows:

1. **Development of a geospatial yield data processing pipeline**: A complete pipeline was developed for cleaning, fusing, and spatially aggregating yield data with UAV multispectral imagery into 5 m $\times$ 5 m tiles. This pipeline, built upon the work of [11, 10, 58], ensures that yield data are free from erroneous variables and exhibit distributions closer to normality.

2. **Implementation of spatially aware evaluation strategies**: Multiple spatial validation strategies were implemented and compared, including Leave-

One-Block-Out (LOBO) and Leave-One-Cluster-Out (LOCO) for assessing intra-field generalization, and Leave-One-Field-Out (LOFO) for evaluating cross-field transferability. These spatially aware approaches were compared against traditional non spatial baseline methodologies involving random train–test splits. **It was revealed that random train and test split results are overly optimistic, as they exploit spatial autocorrelation between data points, particularly for RF models.**

3. **Key findings and insights**:

   (a) **Spectral information richness**: Models trained on combined raw spectral bands outperformed those using vegetation indices alone. Among vegetation indices, NDRE paired with Linear Regression produced the most consistent results across spatial configurations. **This finding has practical implications: in cost sensitive applications, UAVs equipped with only Red Edge and NIR bands can still provide valuable predictive performance while reducing sensor and processing costs.**

   (b) **Field transferability**: Cross-field transferability remained poor when models relied solely on UAV spectral information, with LOFO performance substantially degraded compared to within field baselines. **This limitation highlights the need to integrate complementary data sources—such as soil properties, topography, and management practices to achieve robust cross-field generalization.**

   (c) **Intra-field variability**: The spatial partitioning strategy significantly influenced intra-field spatial generalization performance. LOCO outperformed LOBO, producing more stable and accurate predictions. This finding demonstrates that **(1) the choice of spatial sampling design critically affects model evaluation performance, and (2) predicting yield in unobserved zones from data collected in other field areas is feasible when an appropriate spatial splitting strategy is used to identify the most representative regions.**

(d) **Performance variability across fields**: Model performance varied substantially across fields, with Man Home consistently achieving the best results, followed by Alb PP and Man Eds. These differences could not be attributed solely to crop type (two canola fields showed different performance patterns), suggesting that data collection protocols, field management history, and sensor configuration variability significantly impact model accuracy. **This underscores the critical need for standardized data acquisition and processing procedures across sites.**

5.1.1 Limitations

While this research provides valuable insights into UAV-based yield prediction, several limitations must be acknowledged :

- **Temporal coverage:** Only one UAV image were used per field, which presents significant constraints. Without prior knowledge of field conditions at the time of acquisition or verification of image processing quality, the dataset's representativeness remains uncertain. Single date captures are vulnerable to transient weather conditions, phenological timing variations, and sensor artifacts that may not reflect the field's true productive potential. Multi temporal imagery would pro

- **Data standardization:** UAV data were collected under variable conditions, and metadata were incomplete for some fields (e.g., drone's flight eight). The absence of standardized data collection protocols across fields introduced uncontrolled variability into the analysis. This lack of standardization complicates the interpretation of cross-field performance differences and limits the reproducibility of findings.

- **Limited data:** For this study only UAV imagery data were used. The absence of complementary data sources including soil properties, environmental variables and management practices restricted the models' ability to predict

crop yield more accurately by using determining factors. Relying exclusively on spectral data constrained model transferability across fields.

- **Sample size constraints:** The analysis was done using data from three fields (two canola, one corn) due to data quality issues. Two additional fields were excluded because of unreliable imagery quality reducing the diversity of the dataset.

5.1.2 Recommendations for UAV Data Collection

Based on the findings of this study, the following guidelines are recommended for operational UAV-based yield prediction systems:

- **Timing:** Images should be acquire during the mid-season growing stage when spectral information is the most meaningful for prediction tasks.
- **Temporal coverage:** Acquire at least three images spaced throughout the mid-season period. This would help since weather condition are varying from one date to another.
- **Sensor selection:** Prioritize multispectral sensors with Red Edge and NIR bands as minimum requirements; full spectral coverage (Blue, Green, Red, Red Edge, NIR) is preferred when feasible.
- **Spatial coverage:** Ensure complete field coverage to avoid sampling bias and to produce representative training data across all yield zones.
- **Metadata documentation:** Systematically record flight parameters including altitude, ground speed, weather conditions (temperature, wind speed, cloud cover) and standardize the imagery collection protocol based on these parameters.
- **Quality control:** Implement immediate post flight quality checks to verify imagery integrity, adequate coverage.

5.2 Future Work

While this research produced insightful results, several limitations suggest directions for future investigations. First, future studies should integrate additional data sources such as weather variables, soil characteristics, and management practices to enhance model transferability and robustness. This study relied exclusively on spectral information derived from UAV imagery, which proved insufficient for consistent field-to-field generalization. Second, the inclusion of temporal aspects through UAV imagery acquired at multiple timestamps could enrich the dataset, particularly during the mid-season, as previous studies have demonstrated that this period offers the most informative spectral signatures for yield prediction and this can help to capture data in different weather conditions. Third, pushing further the analysis of spatial partitioning techniques is recommended to better understand their influence on model evaluation, in this study we only analyzed two of them, but there lot more to explore. The results of this work have shown that spatial partitioning significantly affect model performance and interpretation.

In addition, exploring more advanced machine learning and deep learning models, such as Convolutional Neural Networks (CNNs), could allow for a more effective capture of spatial and spectral dependencies within UAV data. The impact of spatial resolution should also be investigated by experimenting with different tile sizes to assess their influence on model accuracy and computational cost. This study employed a uniform 5 m $\times$ 5 m grid, but alternative resolutions could help identify optimal trade-offs between predictive performance and efficiency. Finally, expanding the dataset to include more fields and crop types would provide greater statistical diversity and enable broader generalization across agricultural environments, as the present study was limited to three fields (two canola and one corn).

Bibliography

[1] *1.4. Support Vector Machines*. scikit-learn. URL: https://scikit-learn/stable/modules/svm.html.

[2] *A Friendly Introduction to [Deep] Neural Networks — KNIME*. URL: https://www.knime.com/blog/a-friendly-introduction-to-deep-neural-networks.

[3] A. D. Beck, S. W. Searcy, and J. P Roades. "YIELD DATA FILTERING TECHNIQUES FOR IMPROVED MAP ACCURACY". In: *Applied Engineering in Agriculture* 17.4 (2001). ISSN: 1943-7838.

[4] Nikos Alexandratos. "World Agriculture towards 2030/2050: the 2012 revision". In: ().

[5] Selcuk Arslan and Thomas S Colvin. "An Evaluation of the Response of Yield Monitors and Combines to Varying Yields". In: ().

[6] Liana Baghdasaryan et al. "Deep density estimation based on multi-spectral remote sensing data for in-field crop yield forecasting". In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). ISSN: 2160-7516. June 2022, pp. 2013–2022.

[7] Jordan Steven Bates et al. "Estimating Canopy Density Parameters Time-Series for Winter Wheat Using UAS Mounted LiDAR". In: *Remote Sensing* 13.4 (Feb. 15, 2021), p. 710. ISSN: 2072-4292.

[8] Lefteris Benos et al. “Machine Learning in Agriculture: A Comprehensive Updated Review”. In: *Sensors* 21.11 (Jan. 2021). Publisher: Multidisciplinary Digital Publishing Institute, p. 3758. ISSN: 1424-8220.

[9] Sourav Bhadra et al. “End-to-end 3D CNN for plot-scale soybean yield prediction using multitemporal UAV-based RGB images”. In: *Precision Agriculture* 25.2 (Apr. 1, 2024), pp. 834–864. ISSN: 1573-1618.

[10] Bhavesh Bisht. “Yield Prediction Using Spatial and Temporal Deep Learning Algorithms and Data Fusion”. In: (Nov. 24, 2023). In collab. with My University. Publisher: Université d’Ottawa / University of Ottawa.

[11] Simon Blackmore and Mark Moore. “Remedial Correction of Yield Map Data”. In: ().

[12] Jon A. Brandt and David A. Bessler. “Price forecasting and evaluation: An application in agriculture”. In: *Journal of Forecasting* 2.3 (1983). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/for.3980020306, pp. 237–248. ISSN: 1099-131X.

[13] Hannah Burdett and Christopher Wellen. “Statistical and machine learning methods for crop yield prediction in the context of precision agriculture”. In: *Precision Agriculture* 23.5 (Oct. 2022), pp. 1553–1574. ISSN: 1385-2256, 1573-1618.

[14] Natural Resources Canada. *Spatial Resolution, Pixel Size, and Scale*. Last Modified: 2025-02-10. Jan. 28, 2008. URL: `https://natural-resources.canada.ca/maps-tools-publications/satellite-elevation-air-photos/spatial-resolution-pixel-size-scale`.

[15] Liege Cheung et al. “Using a novel clustered 3D-CNN model for improving crop future price prediction”. In: *Knowledge-Based Systems* 260 (Jan. 25, 2023), p. 110133. ISSN: 0950-7051.

[16] *Ciba Foundation Symposium 210 - Precision Agriculture: Spatial and Temporal Variability of Environmental Quality*. 1st ed. _eprint: https://onlinelibrary.wiley.com/doi/pdf/ John Wiley & Sons, Ltd.

[17] "Climate Trends and Global Crop Production Since 1980". In: *ResearchGate* (Aug. 6, 2025).

[18] A. Colin Cameron and Frank A.G. Windmeijer. "An R-squared measure of goodness of fit for some common nonlinear regression models". In: *Journal of Econometrics* 77.2 (Apr. 1997), pp. 329–342. ISSN: 03044076.

[19] *Coordinate system*. In: *Wikipedia*. Page Version ID: 1264043777. Dec. 20, 2024.

[20] *Coordinate systems, map projections, and transformations—ArcGIS Pro — Documentation*. URL: https://pro.arcgis.com/en/pro-app/latest/help/mapping/properties/coordinate-systems-and-projections.htm.

[21] Stetson Doggett. *What Is An Orthomosaic? Orthomosaic Maps & Orthophotos Explained*. dronegenuity. Nov. 27, 2019. URL: https://www.dronegenuity.com/orthomosaic-maps-explained/.

[22] *e-Reader — Internet of Things for Agriculture 4.0 — Impact and Challen*. URL: https://www-taylorfrancis-com.proxy.bib.uottawa.ca/reader/read-online/63e511d0-3e89-407d-83e4-3fa6717de35d/book/epub?context=ubx.

[23] *Enhanced vegetation index*. In: *Wikipedia*. Page Version ID: 1300941948. July 17, 2025.

[24] Matheus A. Ferraciolli, Felipe F. Bocca, and Luiz Henrique A. Rodrigues. "Neglecting spatial autocorrelation causes underestimation of the error of sugarcane yield models". In: *Computers and Electronics in Agriculture* 161 (June 2019), pp. 233–240. ISSN: 01681699.

[25] *Figure 5. Raster and Vector data types as representative of 'real...* ResearchGate. URL: https://www.researchgate.net/figure/Raster-and-Vector-data-types-as-representative-of-real-world-geographic-space_fig4_269335995.

[26] *File:Random forest explain.png - Wikipedia*. Nov. 11, 2021. URL: https://commons.wikimedia.org/wiki/File:Random_forest_explain.png.

[27] Patrick Filippi, Si Yang Han, and Thomas F.A. Bishop. "On crop yield modelling, predicting, and forecasting and addressing the common issues in published studies". In: *Precision Agriculture* 26.1 (Dec. 7, 2024), p. 8. ISSN: 1573-1618.

[28] Patrick Filippi et al. "An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning". In: *Precision Agriculture* 20.5 (Oct. 1, 2019), pp. 1015–1029. ISSN: 1573-1618.

[29] Flavio Furukawa et al. "Corn Height Estimation Using UAV for Yield Prediction and Crop Monitoring". In: *Unmanned Aerial Vehicle: Applications in Agriculture and Environment*. Ed. by Ram Avtar and Teiji Watanabe. Cham: Springer International Publishing, 2020, pp. 51–69. ISBN: 978-3-030-27156-5 978-3-030-27157-2.

[30] *Geographic coordinate system*. In: *Wikipedia*. Page Version ID: 1278199138. Mar. 1, 2025.

[31] *GeoPandas 1.0.1 — GeoPandas 1.0.1+0.g747d66e.dirty documentation*. URL: `https://geopandas.org/en/stable/index.html`.

[32] Yahui Guo et al. "Integrated UAV-Based Multi-Source Data for Predicting Maize Grain Yield Using Machine Learning Approaches". In: *Remote Sensing* 14.24 (Jan. 2022). Number: 24 Publisher: Multidisciplinary Digital Publishing Institute, p. 6290. ISSN: 2072-4292.

[33] Luthfan Nur Habibi, Tsutomu Matsui, and Takashi S. T. Tanaka. "Critical evaluation of the effects of a cross-validation strategy and machine learning optimization on the prediction accuracy and transferability of a soybean yield prediction model using UAV-based remote sensing". In: *Journal of Agriculture and Food Research* 16 (June 1, 2024), p. 101096. ISSN: 2666-1543.

[34] Jiawei Han, Micheline Kamber, and Jian Pei. "12 - Outlier Detection". In: *Data Mining (Third Edition)*. Ed. by Jiawei Han, Micheline Kamber, and Jian Pei. The Morgan Kaufmann Series in Data Management Systems. Boston: Morgan Kaufmann, Jan. 1, 2012, pp. 543–584. ISBN: 978-0-12-381479-1.

[35] Jiawei Han, Micheline Kamber, and Jian Pei. "2 - Getting to Know Your Data". In: *Data Mining (Third Edition)*. Ed. by Jiawei Han, Micheline Kamber, and Jian Pei. The Morgan Kaufmann Series in Data Management Systems. Boston: Morgan Kaufmann, Jan. 1, 2012, pp. 39–82. ISBN: 978-0-12-381479-1.

[36] Jiawei Han, Micheline Kamber, and Jian Pei. "3 - Data Preprocessing". In: *Data Mining (Third Edition)*. Ed. by Jiawei Han, Micheline Kamber, and Jian Pei. The Morgan Kaufmann Series in Data Management Systems. Boston: Morgan Kaufmann, Jan. 1, 2012, pp. 83–124. ISBN: 978-0-12-381479-1.

[37] Jichong Han et al. "Prediction of Winter Wheat Yield Based on Multi-Source Data and Machine Learning in China". In: *Remote Sensing* 12.2 (Jan. 9, 2020), p. 236. ISSN: 2072-4292.

[38] Fares M. Howari and Habes Ghrefat. "Chapter 4 - Geographic information system: spatial data structures, models, and case studies". In: *Pollution Assessment for Sustainable Practices in Applied Sciences and Engineering*. Ed. by Abdel-Mohsen O. Mohamed, Evan K. Paleologos, and Fares M. Howari. Butterworth-Heinemann, Jan. 1, 2021, pp. 165–198. ISBN: 978-0-12-809582-9.

[39] *https://arxiv.org/pdf/1511.08861*. URL: https://arxiv.org/pdf/1511.08861.

[40] *https://arxiv.org/pdf/2402.01295*. URL: https://arxiv.org/pdf/2402.01295.

[41] *https://link-springer-com.proxy.bib.uottawa.ca/content/pdf/10.1023/A%3A1013887801918.pdf*. URL: https://link-springer-com.proxy.bib.uottawa.ca/content/pdf/10.1023/A%3A1013887801918.pdf.

[42] *IDB - Search results for ndvi*. URL: https://www.indexdatabase.de/search/?s=ndvi.

[43] *Introduction to Recurrent Neural Networks*. GeeksforGeeks. Section: Misc. URL: https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/.

[44] *Isolation forest*. In: *Wikipedia*. Page Version ID: 1268845920. Jan. 11, 2025.

[45] *IsolationForest*. scikit-learn. URL: `https://scikit-learn/stable/modules/generated/sklearn.ensemble.IsolationForest.html`.

[46] Ramesh Janipella, Vikash Gupta, and Rucha V. Moharir. "Chapter 8 - Application of Geographic Information System in Energy Utilization". In: *Current Developments in Biotechnology and Bioengineering*. Ed. by Sunil Kumar, Rakesh Kumar, and Ashok Pandey. Elsevier, Jan. 1, 2019, pp. 143–161. ISBN: 978-0-444-64083-3.

[47] Graham R. Jeffries et al. "Mapping sub-field maize yields in Nebraska, USA by combining remote sensing imagery, crop simulation models, and machine learning". In: *Precision Agriculture* 21.3 (June 2020), pp. 678–694. ISSN: 1385-2256, 1573-1618.

[48] Jig Han Jeong et al. "Random Forests for Global and Regional Crop Yield Predictions". In: *PLOS ONE* 11.6 (2016). Publisher: Public Library of Science, e0156571. ISSN: 1932-6203.

[49] Zhe Jiang. "A Survey on Spatial Prediction Methods". In: *IEEE Transactions on Knowledge and Data Engineering* 31.9 (Sept. 2019), pp. 1645–1664. ISSN: 1558-2191.

[50] Sri Charan Kakarla et al. "Applications of UAVs and Machine Learning in Agriculture". In: *Unmanned Aerial Systems in Precision Agriculture*. Ed. by Zhao Zhang et al. Vol. 2. Series Title: Smart Agriculture. Singapore: Springer Nature Singapore, 2022, pp. 1–19. ISBN: 978-981-19-2026-4 978-981-19-2027-1.

[51] Nantachai Kantanantha, Nicoleta Serban, and Paul Griffin. "Yield and Price Forecasting for Stochastic Crop Decision Planning". In: *Journal of Agricultural, Biological, and Environmental Statistics* 15.3 (2010). Publisher: [International Biometric Society, Springer], pp. 362–380. ISSN: 1085-7117.

[52] Sarah Kentsch et al. “Analysis of UAV-Acquired Wetland Orthomosaics Using GIS, Computer Vision, Computational Topology and Deep Learning”. In: *Sensors* 21.2 (Jan. 2021). Number: 2 Publisher: Multidisciplinary Digital Publishing Institute, p. 471. ISSN: 1424-8220.

[53] Saeed Khaki and Lizhi Wang. “Crop Yield Prediction Using Deep Neural Networks”. In: *Frontiers in Plant Science* 10 (May 22, 2019). Publisher: Frontiers. ISSN: 1664-462X.

[54] Patrick Killeen, Iluju Kiringa, and Tet Yeap. “Corn Grain Yield Prediction Using UAV-based High Spatiotemporal Resolution Multispectral Imagery”. In: *2022 IEEE International Conference on Data Mining Workshops (ICDMW)*. 2022 IEEE International Conference on Data Mining Workshops (ICDMW). ISSN: 2375-9259. Nov. 2022, pp. 1054–1062.

[55] Patrick Killeen et al. “Corn Grain Yield Prediction Using UAV-Based High Spatiotemporal Resolution Imagery, Machine Learning, and Spatial Cross-Validation”. In: *Remote Sensing* 16.4 (Feb. 14, 2024), p. 683. ISSN: 2072-4292.

[56] Thomas van Klompenburg, Ayalew Kassahun, and Cagatay Catal. “Crop yield prediction using machine learning: A systematic literature review”. In: *Computers and Electronics in Agriculture* 177 (Oct. 1, 2020), p. 105709. ISSN: 0168-1699.

[57] Chung-Ming Kuan and Halbert White. “Artificial neural networks: an econometric perspective”. In: *Econometric Reviews* 13.1 (Jan. 1994), pp. 1–91. ISSN: 0747-4938, 1532-4168.

[58] G. Lyle, B. A. Bryan, and B. Ostendorf. “Post-processing methods to eliminate erroneous grain yield measurements: review and directions for future development”. In: *Precision Agriculture* 15.4 (Aug. 2014), pp. 377–402. ISSN: 1385-2256, 1573-1618.

[59] Michael J. Mahoney et al. *Assessing the performance of spatial cross-validation approaches for models of spatially structured data*. Mar. 13, 2023. arXiv: 2303.07334[stat].

[60] Kanichiro Matsumura. "Unmanned Aerial Vehicle (UAV) for Fertilizer Management in Grassland of Hokkaido, Japan". In: *Unmanned Aerial Vehicle: Applications in Agriculture and Environment*. Ed. by Ram Avtar and Teiji Watanabe. Cham: Springer International Publishing, 2020, pp. 39–50. ISBN: 978-3-030-27156-5 978-3-030-27157-2.

[61] Md. Suruj Mia et al. "Multimodal Deep Learning for Rice Yield Prediction Using UAV-Based Multispectral Imagery and Weather Data". In: *Remote Sensing* 15.10 (May 10, 2023), p. 2511. ISSN: 2072-4292.

[62] Petteri Nevavuori, Nathaniel Narra, and Tarmo Lipping. "Crop yield prediction with deep convolutional neural networks". In: *Computers and Electronics in Agriculture* 163 (Aug. 1, 2019), p. 104859. ISSN: 0168-1699.

[63] *Normalized Difference Red Edge Index*. In: *Wikipedia*. Page Version ID: 1297709338. June 28, 2025.

[64] *Normalized difference vegetation index*. In: *Wikipedia*. Page Version ID: 1296825487. June 22, 2025.

[65] Ibrahim A. Onour and Bruno S. Sergi. "Modeling and forecasting volatility in global food commodity prices". In: *Agricultural Economics (Zemědělská ekonomika)* 57.3 (Mar. 31, 2011), pp. 132–139. ISSN: 0139570X, 18059295.

[66] Sudhanshu Sekhar Panda, Daniel P. Ames, and Suranjan Panigrahi. "Application of Vegetation Indices for Agricultural Crop Yield Prediction Using Neural Network Techniques". In: *Remote Sensing* 2.3 (Mar. 1, 2010), pp. 673–696. ISSN: 2072-4292.

[67] J. L. Ping and A. Dobermann. "Processing of Yield Map Data". In: *Precision Agriculture* 6.2 (Apr. 2005), pp. 193–212. ISSN: 1385-2256, 1573-1618.

[68] *Pix4D*. In: *Wikipedia*. Page Version ID: 1306930496. Aug. 20, 2025.

[69] *PIX4Dmapper: Professional photogrammetry software for drone mapping*. Pix4D. URL: `https://www.pix4d.com/product/pix4dmapper-photogrammetry-software`.

[70] Jonne Pohjankukka et al. “Estimating the Prediction Performance of Spatial Models via Spatial k-Fold Cross Validation”. In: *International Journal of Geographical Information Science* 31.10 (Oct. 3, 2017), pp. 2001–2019. ISSN: 1365-8816, 1362-3087. arXiv: 2005.14263[stat].

[71] Mengjia Qiao et al. “Crop yield prediction from multi-spectral, multi-temporal remotely sensed imagery using recurrent 3D convolutional neural networks”. In: *International Journal of Applied Earth Observation and Geoinformation* 102 (Oct. 1, 2021), p. 102436. ISSN: 1569-8432.

[72] Panagiotis Radoglou-Grammatikis et al. “A compilation of UAV applications for precision agriculture”. In: *Computer Networks* 172 (May 2020), p. 107148. ISSN: 13891286.

[73] Nitin Rai et al. “A case study on canola *(Brassica napus* L.*)* potential yield prediction using remote sensing imagery and advanced data analytics”. In: *Smart Agricultural Technology* 9 (Dec. 1, 2024), p. 100698. ISSN: 2772-3755.

[74] Rahul Raj et al. “Precision Agriculture and Unmanned Aerial Vehicles (UAVs)”. In: *Unmanned Aerial Vehicle: Applications in Agriculture and Environment*. Ed. by Ram Avtar and Teiji Watanabe. Cham: Springer International Publishing, 2020, pp. 7–23. ISBN: 978-3-030-27156-5 978-3-030-27157-2.

[75] *RedEdge-MX Dual Camera System - Integration Guide*. MicaSense Knowledge Base. Jan. 26, 2022. URL: https://support.micasense.com/hc/en-us/articles/360037369993-RedEdge-MX-Dual-Camera-System-Integration-Guide.

[76] *RedEdge-P Integration Guide*. MicaSense Knowledge Base. Mar. 11, 2025. URL: https://support.micasense.com/hc/en-us/articles/4410824602903-RedEdge-P-Integration-Guide.

[77] Jonathan Richetti et al. “A methods guideline for deep learning for tabular data in agriculture with a case study to forecast cereal yield”. In: *Computers and Electronics in Agriculture* 205 (Feb. 1, 2023), p. 107642. ISSN: 0168-1699.

[78] Jonathan Richetti et al. “A methods guideline for deep learning for tabular data in agriculture with a case study to forecast cereal yield”. In: *Computers and Electronics in Agriculture* 205 (Feb. 1, 2023), p. 107642. ISSN: 0168-1699.

[79] *Root mean square deviation*. In: *Wikipedia*. Page Version ID: 1305777427. Aug. 14, 2025.

[80] Nidhi Sahai. *Convolutional Neural Network: Layers, Types, & More*. Blogs & Updates on Data Science, Business Analytics, AI Machine Learning. Jan. 8, 2024. URL: https://www.analytixlabs.co.in/blog/convolutional-neural-network/.

[81] Verónica Saiz-Rubio and Francisco Rovira-Más. “From Smart Farming towards Agriculture 5.0: A Review on Crop Data Management”. In: *Agronomy* 10.2 (Feb. 2020). Number: 2 Publisher: Multidisciplinary Digital Publishing Institute, p. 207. ISSN: 2073-4395.

[82] Eryk lewinson Says. *A Comprehensive Overview of Regression Evaluation Metrics*. NVIDIA Technical Blog. Apr. 20, 2023. URL: https://developer.nvidia.com/blog/a-comprehensive-overview-of-regression-evaluation-metrics/.

[83] Raí A. Schwalbert et al. “Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil”. In: *Agricultural and Forest Meteorology* 284 (Apr. 15, 2020), p. 107886. ISSN: 0168-1923.

[84] *scikit-learn: machine learning in Python — scikit-learn 1.7.1 documentation*. URL: https://scikit-learn.org/stable/.

[85] *Spatial Data - an overview — ScienceDirect Topics*. URL: https://www.sciencedirect.com/topics/engineering/spatial-data.

[86] *Spatial without Compromise · QGIS Web Site*. URL: https://qgis.org/.

[87] Spatial Zone. *What Is GIS? A Guide to Geographic Information Systems*. Apr. 23, 2023.

[88] S. Sunoj et al. “Corn Grain Yield Prediction and Mapping from Unmanned Aerial System (UAS) Multispectral Imagery”. In: *Remote Sensing* 13.19 (Jan. 2021). Number: 19 Publisher: Multidisciplinary Digital Publishing Institute, p. 3948. ISSN: 2072-4292.

[89] Namgiri Suresh et al. “Crop Yield Prediction Using Random Forest Algorithm”. In: *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*. 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS). Vol. 1. ISSN: 2575-7288. Mar. 2021, pp. 279–282.

[90] Nicoleta Tantalaki, Stavros Souravlas, and Manos Roumeliotis. “Data-Driven Decision Making in Precision Agriculture: The Rise of Big Data in Agricultural Systems”. In: *Journal of Agricultural & Food Information* 20.4 (Oct. 2, 2019), pp. 344–380. ISSN: 1049-6505, 1540-4722.

[91] Danilo Tedesco-Oliveira et al. “Convolutional neural networks in predicting cotton yield from images of commercial fields”. In: *Computers and Electronics in Agriculture* 171 (Apr. 1, 2020), p. 105307. ISSN: 0168-1699.

[92] Haifeng Tian et al. “A Novel Spectral Index for Automatic Canola Mapping by Using Sentinel-2 Imagery”. In: *Remote Sensing* 14.5 (Jan. 2022). Number: 5 Publisher: Multidisciplinary Digital Publishing Institute, p. 1113. ISSN: 2072-4292.

[93] W. R. Tobler. “A Computer Movie Simulating Urban Growth in the Detroit Region”. In: *Economic Geography* 46 (June 1970), p. 234. ISSN: 00130095.

[94] Claudia Vallentin et al. “Delineation of management zones with spatial data fusion and belief theory”. In: *Precision Agriculture* 21.4 (Aug. 2020), pp. 802–830. ISSN: 1385-2256, 1573-1618.

[95] Sergio Vélez et al. “Speeding up UAV-based crop variability assessment through a data fusion approach using spatial interpolation for site-specific management”. In: *Smart Agricultural Technology* 8 (Aug. 1, 2024), p. 100488. ISSN: 2772-3755.

[96] Anna X. Wang et al. “Deep Transfer Learning for Crop Yield Prediction with Remote Sensing Data”. In: *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*. COMPASS ’18: ACM SIGCAS Conference on Computing and Sustainable Societies. Menlo Park and San Jose CA USA: ACM, June 20, 2018, pp. 1–5. ISBN: 978-1-4503-5816-3.

[97] Lipo Wang. *Support Vector Machines: Theory and Applications*. Secaucus, UNITED STATES: Springer, 2005. ISBN: 978-3-540-32384-6.

[98] Seth J. Wenger and Julian D. Olden. “Assessing transferability of ecological models: an underappreciated aspect of statistical validation”. In: *Methods in Ecology and Evolution* 3.2 (2012). _eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/1 210X.2011.00170.x, pp. 260–267. ISSN: 2041-210X.

[99] *WGS 84*. In: *Wikipédia*. Page Version ID: 219242552. Oct. 7, 2024.

[100] *What is a Neural Network? — IBM*. Oct. 6, 2021. URL: `https://www.ibm.com/think/topics/neural-networks`.

[101] *What is a Recurrent Neural Network (RNN)? — IBM*. Oct. 4, 2021. URL: `https://www.ibm.com/think/topics/recurrent-neural-networks`.

[102] *What is LSTM - Long Short Term Memory?* GeeksforGeeks. Section: Computer Subject. URL: `https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/`.

[103] *World population prospects 2019: highlights*. OCLC: 1110010089. New York: United Nations, 2019. ISBN: 978-92-1-148316-1.

[104] Rui Xu and Don Wunsch. *Clustering*. Newark, UNITED STATES: John Wiley & Sons, Incorporated, 2008. ISBN: 978-0-470-38278-3.

[105] Jinru Xue and Baofeng Su. “Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications”. In: *Journal of Sensors* 2017.1 (2017), p. 1353691. ISSN: 1687-7268.

[106] Qi Yang et al. “Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images”. In: *Field Crops Research* 235 (Apr. 2019), pp. 142–153. ISSN: 03784290.

[107] *Z score for Outlier Detection - Python*. GeeksforGeeks. Section: Machine Learning. URL: https://www.geeksforgeeks.org/z-score-for-outlier-detection-python/.

[108] Naiqian Zhang, Maohua Wang, and Ning Wang. "Precision agriculture—a worldwide overview". In: *Computers and Electronics in Agriculture* 36.2 (Nov. 2002), pp. 113–132. ISSN: 01681699.