UCLouvain

AGRO LOUVAIN

Faculté des bioingénieurs

Developing Wheat Crop Yield Estimation Method for Spain from Remotely Sensed Metrics Using Artificial Intelligence

Auteur : Abdullah Toqeer

Promoteur : Prof. Pierre Defourny

Lecteur(s) : Prof. Xavier Draye

Dr. Julien Radoux

Année académique: 2022-2023

Mémoire de fin d'études présenté en vue de l'obtention du diplôme de Master SAIV – MSc Erasmus+ GEM : Geo-information Science and Earth Observation for Environmental Modelling and Management

Acknowledgments

I would like to express my gratitude to my supervisor, Professor Pierre Defourny, for his invaluable teaching, guidance, and support throughout the master's thesis journey. I am also grateful to Pierre Houdmont and Quentin Deffense for frequently meeting me and providing guidance throughout the master thesis. I am thankful to the European Space Agency Project Sen4Stat and the National Statistics Office of Spain for sharing the data for research. I am also grateful to Earth and Life Institute of Université catholique de Louvain for providing the service of virtual machines to carry out research.

I want to take this opportunity to thank the European Education and Culture Executive Agency, GEM Erasmus Mundus Program, and its educational and industrial partners, especially the University of Tartu and Université catholique de Louvain, for providing me with this great opportunity of studying double masters on scholarship. I want to acknowledge my family and friends for their unwavering support and motivation throughout the journey.

Lastly, I wish my jury members, Professor Xavier Draye and Dr. Julien Radoux, and the readers of this thesis, as much enjoyment in reading it as I had in writing it.

Table of Contents

List of Figures……vii

List of Table……..viii

List of Equations…...……viii

Acronyms……ix

Chapter 1: Introduction 1

Chapter 2: Literature Review 4

2.1 Crop Yield Estimation Before Remote Sensing 4

2.2 Crop Yield Estimation with Remote Sensing 5

2.3 Crop Yield Estimation with Remote Sensing and Artificial Intelligence 9

Chapter 3: Aims and Objectives 20

3.1 General Context 20

3.2 Objectives 21

Chapter 4: Materials 22

4.1 Study Area 22

4.2 Climate Conditions of Spain in the 2019 Growing Season 24

4.3 Data Sources and Data Preparation 25

4.3.1 In-Situ Data 25

4.3.1.1 In-situ Data Preparation 25

4.3.2 Meteorological data 28

4.3.3 Remote Sensing Data 28

4.3.4 Features Extraction 29

Chapter 5: Methods 31

5.1 Selection of Models 31
5.1.1 Multivariate Regression Model 31
5.1.1.1 Least Absolute Shrinkage and Selection Operator Model 31
5.1.2 Machine Learning Model 32
5.1.2.1 Decision Tree Regression 32
5.1.2.2 Random Forest Regression 33
5.1.2.3 Support Vector Regression 34
5.1.3 Deep Learning Model 34
5.1.3.1 Multilayer Perceptron 34
5.2 Optimal Hyperparameters of Models 36
5.3 Model Training and Verification 36
5.4 Model Validation 38
5.5 Model Evaluation Metrics 38
5.5.1 Nash–Sutcliffe Efficiency 39
5.5.2 R Squared 39
5.5.3 Mean Absolute Error 40
5.5.4 Root Mean Squared Error 40
5.5.4 Relative Root Mean Squared Error 41
5.5.5 Root Mean Squared Logarithmic Error 41
5.6 Important Feature Identification 42
Chapter 6: Results 44
6.1 Multivariate Regression Model 44
6.1.1 Least Absolute Shrinkage and Selection Operator Model 44
6.1.1.1 Model Training Checking 44
6.1.1.2 Model Verification 45

6.1.1.3 Model Validation 45
6.2 Machine Learning Model 47
6.2.1 Decision Tree Regression Model 47
6.2.1.1 Model Training Checking 47
6.2.1.2 Model Verification 47
6.2.1.3 Model Validation 47
6.2.2 Random Forest Regression Model 49
6.2.2.1 Model Training Checking 50
6.2.2.2 Model Verification 50
6.2.2.3 Model Validation 50
6.2.3 Support Vector Regression Model 52
6.2.3.1 Model Training Checking 52
6.2.3.2 Model Verification 52
6.2.3.3 Model Validation 52
6.3 Deep Learning Model 54
6.3.1 Multilayer Perceptron Model 54
6.3.1.1 Model Training Checking 56
6.3.1.2 Model Verification 56
6.3.1.3 Model Validation 56
6.4 Comparison of Models 58
6.5 Identification of the Appropriate Remotely Sensed Features 60
6.5.1 Individual Feature Importance 60
6.5.2 Feature Group Importance 61
Chapter 7: Discussion 63
7.1 Multivariate Regression Model 63

7.2 Machine Learning Model 64
7.3 Deep Learning Model 66
7.4 Remote Sensing Feature Importance 67
Chapter 8: Conclusion 71
References 73
Appendix 90

List of Figures

Figure 1: Study Area Map 22
Figure 2: Crop Calendar of Spain for Wheat Crop 23
Figure 3: Frequency of Observed Yield Before Cleaning the Data 26
Figure 4: Frequency of Observed Yield After Cleaning the Data 27
Figure 5: Growth Stages of Crop Based on LAI Time Series 29
Figure 6: Sen4Stat Feature Extraction Flow Chart 30
Figure 7: Simple Neuron Unit 35
Figure 8: 10-fold Cross Validation Approach 37
Figure 9: Methodology of the Current Study 43
Figure 10:Result of LASSO Model Verification Phase 46
Figure 11: Result of LASSO Model Validation Phase 46
Figure 12: Result of Decision Tree Model Verification Phase 48
Figure 13: Result of Decision Tree Model Validation Phase 49
Figure 14: Result of Random Forest Model Verification Phase 51
Figure 15: Result of Random Forest Model Validation Phase 51
Figure 16: Result of Support Vector Regression Model Verification Phase 53
Figure 17: Result of Support Vector Regression Model Validation Phase 54
Figure 18: Model Architecture of Multilayer Perceptron Model 55
Figure 19: Result of Multilayer Perceptron Model Verification Phase 57
Figure 20: Result of Multilayer Perceptron Model Validation Phase 57
Figure 21: Model Performance Based on R^2 and NSE 58
Figure 22: Model Performance Based on MAE, RMSE, RRMSE & RMSLE 59
Figure 23: Random Forest Feature Importance by Individual Features 61
Figure 24: Random Forest Feature Importance by Group 62
Figure 25: Random Forest Feature Importance by Phenological Stages 69

List of Table

Table 1: Wheat Yield Records for 2019 24

List of Equations

Equation 1 32
Equation 2 35
Equation 3 39
Equation 4 39
Equation 5 40
Equation 6 40
Equation 7 41
Equation 8 41

Acronyms

AI	Artificial Intelligence
ANN	Artificial Neural Network
BNN	Bayesian Neural Network
BPNN	Back Propagation Neural Network
BRT	Boosted Regression Tree
BV-NET	Biophysical Variable Neural Network
CAP	Common Agricultural Policy
CNN	Convolutional Neural Network
CP-ANNs	Counter-Propagation Artificial Neural Networks
DACM	Deep Learning Adaptive Crop Model
DF	Decision Forest
DNN	Deep Neural Network
DT	Decision Tree
DTR	Decision Tree Regression
EF	Nash-Sutcliffe Model Efficiency
ENET	Elastic Net
ERA5 Land	Fifth Generation European Reanalysis Dataset for Land
ESA	European Space Agency
EVI	Enhanced Vegetation Index
FAO	Food And Agriculture Organization
GARVI	Green Atmospherically Resistant Vegetation Index
GCVI	Green Chlorophyll Vegetation Index
GLCM	Gray Level Cooccurrence Matrix
GLM	Generalized Linear Regression Model
GNDVI	Green Normalized Difference Vegetation Index
GP	Deep Gaussian Process Component
GPR	Gaussian Process Regression
GWRFR	Geographically Weighted Random Forest Regression
JRC	Joint Research Centre

LAI	Leaf Area Index
LASSO	Least Absolute Shrinkage and Selection Operator Regression
LR	Linear Regression
LST	Land Surface Temperature
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MARS	Monitoring Agricultural Resources
MLP	Multilayer Perceptron
MLPNN	Multilayer Perceptron Neural Network
MLR	Multiple Linear Regression
MOB	Model-Based Recursive Partitioning
MODIS	Moderate Resolution Imaging Spectroradiometer
MPPE	Mean Percentage Prediction Error
MSE	Mean Squared Error
MWDRVI	Mixed Wide Dynamic Range Vegetation Index
NDRE	Normalized Difference Red Edge
NDRE1	Normalized Difference Red Edge Based on Band 5
NDRE2	Normalized Difference Red Edge Based on Band 6
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NN	Neural Network
nRMSE	Normalized Root Mean Square Error
NSE	Nash–Sutcliffe Efficiency
PCA	Principal Component Analysis
PE	Percent Error
PLS	Partial Least Square
PLSR	Partial Least Squares Regression
PRISMA	Precursore Iperspettrale Della Missione Applicativa
R^2	R-Squared
REIP	Red Edge Inflection Point

RF	Random Forest
RFR	Random Forest Regression
RMSE	Root Mean Square Error
RMSLE	Root Mean Squared Logarithmic Error
RR	Ridge Regression
RRMSE	Relative Root Mean Square Error
SAFY	Simple Algorithm for Yield Estimation
Sen4Stat	Sentinels For Agricultural Statistics
SKNs	Supervised Kohonen Networks
SLR	Stepwise Linear Regression
SMLR	Stepwise Multiple Linear Regression
SR	Simple Regression
SRM	Simple Regression Model
SVM	Support Vector Machine
SVR	Support Vector Regression
TBVIs	Two-Band Vegetation Indices
VTCI	Vegetation Temperature Condition Index
WARM	Water Accounting Rice Model
WDRVI	Wide Dynamic Range Vegetation Index
Willmott's *d*	Index Of Agreement
XGBoost	Extreme Gradient Boosting
XY-Fs	XY-Fused Networks

Chapter 1: Introduction

Agricultural production is an important part of the environment, society, and the economy worldwide, especially the strategic crops directly impacting food security (Abdul-Jabbar et al., 2023; Ruiter et al., 2017). Recent research shows that the global demand for food will continue to increase for the next several decades, and the world population is anticipated to reach an estimated 9 billion people by 2050 (Godfray et al., 2010). With this trend, around 70% increase in food production is needed (FAO, 2009). To accurately assess the extent to which this high demand can be met, it is essential to monitor crop yields across the globe. Crop yield estimation involves predicting the harvestable crop amount from a given crop field area. Crop yield estimation at different scales is important in agricultural monitoring and management, providing critical information to farmers, policymakers, and other stakeholders (Dhillon et al., 2023a; Vannoppen & Gobin, 2022).

The early estimation of crop yield at the field level provides farmers with a baseline of what to expect. Thus, it helps them make better management and planning decisions for risk and premium insurance, ideal planting and harvesting time, and future input costs (Johnson, 2014; Nuske et al., 2014). The farmers use crop yield data to assess different management practices on crop yield and identify any anomalies resulting from climate change or diseases during the growing season (Vannoppen & Gobin, 2022). These factors directly impact the business model of the farmers (Barriguinha et al., 2022). Moreover, the crop yield data is also important for agriculture insurance companies to evaluate the risk of extreme weather events on cropping systems and yield (Vannoppen & Gobin, 2022).

Accurate crop yield assessment is important for effective agricultural planning, ensuring food security, and making trade policies. Stakeholders use this information to make informed and strategic decisions regarding the food and feed stocks of the country (Doraiswamy et al., 2010; Vannoppen & Gobin, 2022). At the same time, policymakers use this information to formulate food policies, regulate food prices, and trade policies (Dhillon et al., 2023a; Fritz et al., 2019).

The rapid progress in remote sensing and Artificial Intelligence (AI) has revolutionized the field of crop monitoring and yield estimation at various spatial and temporal scales (Dhillon et al.,

2023b), providing cost-effective, sustainable, and comprehensive solutions for agricultural monitoring and decision-making (Abdul-Jabbar et al., 2023; A. Prasad et al., 2006). AI algorithms have shown significant results in predicting crop yields using remotely sensed data (Khan et al., 2022; Pang et al., 2022; You et al., 2017). These algorithms learn complex and nonlinear relationships between remotely sensed metrics and crop yield, resulting in highly accurate predictions (Tian et al., 2021; Wang et al., 2022; Zhu et al., 2022).

In recent years, increased population and climate change have significantly threatened agricultural productivity and food security, limiting cropland and water resources (Dhillon et al., 2023b; Xie et al., 2017). Changes in temperature and precipitation patterns have significantly impacted agricultural production (Xie et al., 2017). The accurate and timely crop yield estimation can help mitigate climate change's impacts by providing early warning signals to farmers and other stakeholders. This information is also valuable for identifying cases of regional crop failure, which may require humanitarian aid to prevent famine in vulnerable populations, especially in third-world countries that cannot afford to import food easily (Padilla et al., 2012).

Wheat is a major crop in the European Union with a high economic value and has a valuable role in various sectors such as human consumption, animal feed, and biodiesel production (Alarcón-Segura et al., 2022; Dhillon et al., 2023a). Moreover, estimating yield accurately, mapping the croplands, and studying the spatial factors that influence these crops are particularly important in the European Union, where the requirements for getting subsidies from Common Agricultural Policy (CAP) are linked to land use and agricultural practices. All the members of the European Union must verify the declared field area, crop type, and compliance with required polices (EU Regulation 2013; Segarra et al., 2020).

Despite the importance of crop yield estimation in food security, agricultural sustainability, and economic growth (Vannoppen & Gobin, 2022), accurate and timely estimation at different scales and finding optimal temporal and spatial resolution remains a challenge that needs to be studied (Dhillon et al., 2023b). Crop yields are influenced by multiple factors, including weather, soil conditions, pests, and diseases, making prediction a complex task. However, remote sensing and AI have offered new opportunities for improving crop yield estimation accuracy and timeliness.

This study aims to develop a wheat yield estimation method for Spain using remotely sensed metrics and artificial intelligence. This method could be used with other approaches to improve food security, sustainable agriculture, and trade policies, making an important contribution to agricultural monitoring and management.

Chapter 2: Literature Review

This chapter will provide an overview of the current state of the art scientific knowledge about crop yield estimation methods across various scales. Section 2.1 will describe the traditional methods and approaches historically employed or still in use for crop yield estimation. Section 2.2 will explore the utilization of remote sensing derived data in crop yield estimation. Finally, in section 2.3, the utilization of different artificial intelligence techniques, such as machine learning and deep learning, combined with the remote sensing data, will be discussed for crop yield estimation purposes.

2.1 Crop Yield Estimation Before Remote Sensing

The traditional methods and approaches for monitoring crops and yield estimation are primarily based on field surveys (Mehdaoui & Anane, 2020). Crop cut is one of the oldest yield estimation methods. Developed in India in the 1950s, it quickly gained widespread acceptance and became one of the standards for crop yield estimation (Fermont & Benson, 2011). The methodology involves the estimation of yield in one or more sub-parcels of the field, and the total yield is then determined by calculating the overall production per total harvested area of the sub-parcels. This method has been used as a standard for years and recommended by international organizations, including the United Nations Food and Agriculture Organization (FAO). The downside of this approach is it doesn't fully incorporate the heterogeneity of crop conditions within the field. A study in Bangladesh showed that the crop cut estimate yield was 20 percent more than the actual yield (Diskin, 1999). Moreover, this method doesn't consider the post-harvest yield (Sapkota et al., 2016).

The farmer's estimate is another conventional method utilized in the past. It involves interviewing the farmers to obtain their estimation either before the harvest (pre-harvesting) or through retrospective recall after the harvest (post-harvesting) for their farm or parcel. The farmer's prediction is based on their previous year's experiences, where they compare crop performance to previous years (Sapkota et al., 2016). This approach is time and cost-efficient as it doesn't require field measurements. A refined version of the farmer's estimate method is the crop card method, in

which the farmer is provided with a survey form where they can document the total quantity of crop for different crops in each harvest. This form can then be used to estimate the total harvest yield for different crops (Sapkota et al., 2016).

The expert assessment method is also used to estimate crop yield. Field agronomists and researchers give opinions about yield by visually evaluating the crop condition (Dumanski & Onofrei, 1989). This assessment is then combined with the field measurements and empirical formulas to estimate yield (Sapkota et al., 2016). The downside of this method is the limited availability of experts with both practical and technical experience who can accurately assess the performance of multiple crops under variable conditions on a large scale.

Another method of estimating the crop yield is by using grain weight. This is performed by counting the number of pods in the measuring frame for different areas of the field and then taking the average. Additionally, the number of grains present in those pods is also counted. These measurements contribute to estimating the crop yield based on the weight of those grains. Compared to the crop cut method, the grain weight approach does not require harvesting and weighing the production (Sapkota et al., 2016). Among the various methods mentioned, the whole field harvest method is considered the most precise, as it involves harvesting the complete field in the trial area. Due to its accuracy, it is used as a standard to check the effectiveness of the other methods (Sapkota et al., 2016).

To study heterogeneity and incorporate the spatial variability factor into these traditional yield estimation methods, soil sampling and testing in the laboratory are needed to evaluate the plant health, nutrients content in the soil, etc., which make these approaches costly (Khanal et al., 2018; Imanni et al., 2022). Moreover, these traditional methods are time consuming on large scales, i.e., regional, country, and global levels (Mulder et al., 2011; Imanni et al., 2022; Yang et al., 2014).

2.2 Crop Yield Estimation with Remote Sensing

Recent research conducted by Khanal et al. (2018), Peng et al. (2015), Stevens et al. (2013), and Yao et al. (2016) has highlighted the potential of satellite earth observation data to replace

traditional approaches to monitor crop and yield. These studies show that by assimilating the georeferenced field data on soil and crop conditions with spectral information acquired through remote sensing techniques, it is possible to estimate various crops and soil properties. In contrast to remote sensing metrics, meteorological variables have been used much earlier to understand and monitor crops. Among them, temperature and precipitation are the main ones for crop yield estimation. Along with these variables, some proxy parameters like Land Surface Temperature (LST) are also collected and used in crop yield estimation methods. Johnson (2014) used LST and Normalized Difference Vegetation Index (NDVI) metrics in the regression tree model to predict the corn and soybean yield at the county level in the United States. NDVI correlated positively with crops, and Daytime LST was negatively correlated with corn yield. Moreover, nighttime LST did not correlate with both crops.

Weather, crop monitoring, and vegetation indices are combined to predict crop yield at local, country, and regional levels (Prasad et al., 2006; Rojas, 2007). In the past, many studies have been done on crop yield mapping using remote sensing data from satellites (Khanal et al., 2018; Lobell et al., 2015), aircraft (Yang et al., 2014), and unmanned vehicles (Geipel et al., 2014; Shi et al., 2016). The remote sensing derived indices and biophysical variables are very effective in yield estimation, especially the NDVI, Enhanced Vegetation Index (EVI), and Leaf Area Index (LAI) (Kuwata & Shibasaki, 2015; Mkhabela et al., 2011; Peng et al., 2019; Zhang et al., 2022). These metrics are based on the reflectance from vegetation and are sensitive to changes in vegetation cover and density. This makes these remote sensing metrics effective in capturing variations in crop growth and development, which are important factors in determining crop yield.

LAI is an important biophysical variable for crop yield estimation as it reflects crop growth status by calculating the photosynthetic rate of the crops. LAI is a measurement that quantifies the ratio of leaf area to the corresponding unit of ground area, and it has been used in many models to estimate crop yield (Xie et al., 2017). The LAI values are proportional to the photosynthetic radiations intercepted by the crop (Zhang et al., 2022). Peng et al. (2019) used the LAI recorded at four different growth stages to estimate the yield of oilseed rape. The LAI values observed at the initiation of the stem elongation stage were found to be most closely related to the oilseed rape yield.

Mkhabela et al. (2011) explored the use of NDVI for spring wheat, canola, barley, canola, and field peas yield prediction in the Canadian Prairies. Based on Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) performance metrics, the difference in predicted and actual yield was within ±10% for all the crops, and the prediction can be made one to two months before harvesting.

Kuwata & Shibasaki. (2015) used the EVI and climate data with deep learning and Support Vector Machine (SVM) models to estimate the corn yield in the USA. Based on the ten-fold cross validation approach and RMSE performance metric, the deep learning model was more efficient, and EVI was strongly related to the yield.

Water Accounting Rice Model (WARM) was used by Gilardelli et al. (2019) to estimate the rice crop yield at a sub-field level in the Pavia province in Northern Italy from 2014 to 2016. The model's performance was assessed using MAE and Relative Root Mean Square Error (RRMSE) performance metrics. The model showed an MAE and RRMSE of 820 kg/ha and 15.7%, respectively. Furthermore, LAI derived from Sentinel–2A, Landsat-7, and Landsat-8 was assimilated within the model. The model's performance significantly improved with an MAE and RRMSE of 660 kg/ha and 13.8%, respectively.

Silvestro et al. (2017) conducted a comparative analysis of AquaCrop and Simple Algorithm for Yield Estimation (SAFY) models to estimate the wheat yield in China at a field level from 2012-2015. LAI and Canopy cover were assimilated in the models, and model performance was assessed using RMSE and RRMSE. The SAFY model outperformed AquaCrop with RMSE and RRMSE of 1090 kg/ha and 18%, respectively. The AquaCrop model showed an RMSE and RRMSE of 1120 kg/ha and 24%, respectively.

LAI assimilation in the SAFY model was also done by Manivasagam et al. (2021) to estimate the within-field yield of wheat crop in Israel for 2017-2018. The LAI was derived from PlanetScope and Sentinel-2 and was fused while assimilating in the model. The model performance was assessed with R-squared (R^2) and RMSE metrics. The model showed a moderate performance with an R^2 and RMSE of 0.45 and 690 kg/ha, respectively.

SAFY was also used by Chahbi Bellakanji et al. (2018) to estimate the yield of barley and wheat cereals at a field level in the region of Kairouan, Tunisia, from 2010 to 2012. The model performance was assessed with RMSE. The SAFY model predicted the barley yield more accurately with an RMSE of 537 kg/ha, followed by the wheat yield with an RMSE of 795 kg/ha.

A comparative study between the SAFY crop growth model and Simple Regression Model (SRM) was done by Gaso et al. (2019) for yield prediction of winter wheat at the field level in western and southern Uruguay for two growing seasons in 2013 and 2014. The SRM model was trained with NDVI at the anthesis stage. SAFY model was trained by assimilating LAI, including two equations, one representing the relationship between nitrogen content of leaf and light use efficiency and the second equation to estimate the evolution of the harvest index during the grain filling period. The model's performance was assessed with RMSE. The SRM was significantly better, with an RMSE of 966 kg/ha, whereas the SAFY showed poor performance, with an RMSE of 1532 kg/ha.

Several studies have observed an improvement in the accuracy of crop yield estimation methods by incorporating weather data along with the remote sensing metrics of LAI, NDVI, and EVI (Cai et al., 2019; Peng et al., 2018; Prasad et al., 2020). Despite extensive research in the field (Cai et al., 2019; Mkhabela et al., 2011; Peng et al., 2018; Peng et al., 2019; Prasad et al., 2020; Xie et al., 2017), there is still a need for further exploration of the potential applications of earth observation data in crop monitoring and yield prediction. Additionally, there is a need to continue improving existing methods to achieve more accurate results. The accuracy of yield prediction mainly depends on remote sensing data's availability, quality, and temporal resolution (Blasch et al., 2015), along with the different approaches used in developing the model (Forkuor et al., 2017). The decision to use which kind of data and method for a particular situation mainly depends on the objective, estimation scale, data availability, and desired precision level (Sapkota et al., 2016). The emergence of online resources, particularly cloud-based platforms such as Google Earth Engine (Gorelick et al., 2017) and Sentinel hub, have facilitated easier and more accessible access to earth observation data and its preprocessing on the online platform (Prasad et al., 2020; Imanni et al., 2022).

2.3 Crop Yield Estimation with Remote Sensing and Artificial Intelligence

Multiple regression models are among the most frequently used and long-established methods for crop monitoring and yield estimation (Geipel et al., 2014; Lobell et al., 2015). Shastry et al. (2017) conducted a study to predict wheat, maize, and cotton crop yield using crop yield and weather data. Multiple regression techniques, such as quadratic, pre-quadratic, interactions, linear polynomial, Generalized Linear Regression Model (GLM), and Stepwise Linear Regression (SLR), were used. The performance of these models was assessed using three performance metrics: R^2, RMSE, and Mean Percentage Prediction Error (MPPE). The analysis showed that the GLM outperformed the other models for wheat yield prediction, pure quadratic models performed better for maize yield prediction, and SLR demonstrated the highest level of accuracy in cotton yield prediction among all the models evaluated. These findings suggest that the choice of the most suitable model for yield prediction can vary depending on the specific crop, highlighting the importance of selecting appropriate modeling approaches tailored to each crop type.

Aravind et al. (2022) recently used six different models, including Stepwise Multiple Linear Regression (SMLR), Artificial Neural Network (ANN), Least Absolute Shrinkage and Selection Operator Regression (LASSO), Elastic Net (ENET), Principal Component Analysis (PCA) alone and together with ANN and SMLR. The goal was to estimate wheat yield using crop yield and weather data of more than 30 years in five Indian cities (Amritsar, Hisar, Ludhiana, Patiala, and IARI, New Delhi). 70% of the data was used to train the models, and the remaining 30% for validation based on three performance metrics, i.e., R^2, RMSE, and Normalized Root Mean Square Error (nRMSE). The results highlighted that LASSO and ENET were the best among the other models used for predicting wheat yield in Hisar, Ludhiana, Amritsar, and IARI New Delhi, as their nRMSE values were less than 10%, and 10-20% for Patiala, indicating that they were not overfitting followed by PCA-SMLR, SMLR, PCA-ANN, and ANN.

Tang et al. (2022) used LASSO and Back Propagation Neural Network (BPNN) methods to forecast the winter wheat yield in the North China Plain by utilizing climate, crop growth, soil moisture, flag leaf photosynthesis, and crop yield data from 2018-2021. The model performance was assessed with R^2 and RMSE evaluation metrics. The results showed that BPNN was more

accurate with R^2 values ranging from 0.66-0.94 and RMSE ranging from 302-684 kg/ha, while for LASSO, they were 0.23-0.78 and 548-1023 kg/ha, respectively. Based on these results, a BPNN model was developed for wheat yield prediction using two growth indicators, LAI and dry matter, which can forecast yield 35 days before harvesting.

Cai et al. (2019) conducted a comparative analysis of different machine learning algorithms at the statistical division level in Australia for predicting wheat yield from 2000-2014. SVM, Random Forest (RF), and Neural Network (NN) were tested with the LASSO multiple linear regression model as the benchmark. The models used satellite (EVI, solar-induced chlorophyll fluorescence), climate, and crop yield data, and the performance of models was evaluated using the R^2 metric. The machine learning models performed better than the multiple linear regression model, with SVM having the highest value of R^2 of 0.75, followed by NN, RF, and LASSO. The models also effectively predicted the wheat yield two months before harvesting. Moreover, exploratory data analysis showed that the combined satellite and climate data provide the best performance compared with using only satellite or climate data.

Zhang et al. (2020) performed a comparative study to predict the winter wheat yield at the field level in 5 counties of Jiangsu Province, China, in 2016. The Linear Regression (LR), Partial Least Square (PLS), and PCA algorithms were used to estimate the yield and were trained on nine different remote sensing indices. The performance of algorithms was assessed using R^2 and RMSE performance metrics. PLS outperformed the other algorithms with R^2 of 0.74 and RMSE of 786.5 kg/ha, followed by PCA with R^2 and RMSE of 0.56 and 1067.3 kg/ha, respectively, and LR with R^2 and RMSE of 0.47 and 1342.7.5 kg/ha.

Similarly, Marszalek et al. (2022) used SLR and RF models to estimate the winter wheat yield at the field level in southern Germany. Satellite data, including time series and indices like normalized difference water index (NDWI), normalized difference red edge (NDRE), and red edge inflection point (REIP), were used along with climate and crop yield data from 2016-2018 to train the models. The range of yield values for the fields studied was 4.9-10.2 t/ha. The performance of models was evaluated based on three performance metrics, i.e., R^2, MAE, and RMSE. The

combination of all raw bands of Sentinel-2 along with evapotranspiration and precipitation data resulted in high performance with R^2 of 0.84, MAE of 0.49 t/ha, and an RMSE of 0.56 t/ha.

Kayad et al. (2019) compared the performance of the Multiple Linear Regression (MLR) model with two machine learning models (RF and SVM) to estimate the within-field variability of corn yield in Ferrara, North Italy, from 2016 to 2018. The models were trained using nine vegetation indices, including NDVI, Normalized Difference Red Edge Based on Band 5 (NDRE1), Normalized Difference Red Edge Based on Band 6 (NDRE2), Green Normalized Difference Vegetation Index (GNDVI), Green Atmospherically Resistant Vegetation Index (GARVI), EVI, Wide Dynamic Range Vegetation Index (WDRVI), Mixed Wide Dynamic Range Vegetation Index (MWDRVI), and Green Chlorophyll Vegetation Index (GCVI) derived from Sentinel-2. The model's accuracy was assessed using three performance metrics, i.e., R^2, MAE, and RMSE. The RF outperformed the other models with R^2 ranging from 0.35 – 0.6, MAE ranging from 620 kg/ha – 1100 kg/ha, and RMSE ranging from 900 kg/ha – 1500 kg/ha followed by SVM with R^2 (0.27 – 0.55), MAE (680 kg/ha – 1200 kg/ha), RMSE (970 kg/ha – 1700 kg/ha) and MLR with R^2 (0.2 – 0.55), MAE (720 kg/ha – 1350 kg/ha), RMSE (1000 kg/ha – 1800 kg/ha).

Lambert et al. (2017) used the LR approach to estimate the yield of multiple crops, including cotton, maize, millet, and sorghum, at the field level in the Koningue commune of southern Mali from 2014 to 2016. The model was trained using remote sensing metrics and LAI, and the performance was evaluated using RMSE. The most accurately predicted yield was for cotton with RMSE of 550 kg/ha, followed by millet with 630 kg/ha, sorghum with 640 kg/ha, and maize with 1050 kg/ha, respectively.

A drawback with the LR models is that they cannot provide the spatial distribution of yield prediction in the heterogeneous environment. Their estimation is based on the simple assumption that the relationship between the environmental variables and crop yield is linear and homogenous in the complete data set. However, in reality, the relationship between the environmental variables and crop yield is nonlinear and heterogenous (Hao et al., 2021; X. Zhang et al., 2018). To address this issue, machine learning techniques are primarily used these days to estimate crop yield using earth observation data (Chlingaryan et al., 2018). They efficiently model the nonlinear

relationships between environmental variables and crop yield (Chen et al., 2016; Jeong et al., 2016). The conventional machine learning methods use vegetation indices to incorporate the crop monitoring factor (Shammi & Meng, 2020; Stas et al., 2016). This approximation is justifiable as studies have shown that vegetation indices and biophysical variables such as LAI, NDVI, and EVI are correlated to crop yield (Mkhabela et al., 2011; Peng et al., 2019; Zhang et al., 2022).

Jeong et al. (2016) used the RF model to predict wheat, potato, and maize yield using climate and biophysical variables. MLR was used as a benchmark to check the effectiveness of RF. The global wheat dataset from 1997-2003, the maize dataset of all maize-producing counties of the United States from 1984-2013, and the potato and maize dataset of the Northeastern seaboard region for 1987, 1992, 1997, and 2002 were used to train these models. The model efficiency was assessed based on three performance metrics, i.e., RMSE, Nash-Sutcliffe model efficiency (EF), and index of agreement (Willmott's *d*). The RF outperformed the MLR predictions with an RMSE ranging from 6% to 14% and EF ranging from 0.92 to 0.99, while for MLR, they were 14% to 49% and -0.87 to 0.31, respectively. Also, it was noted that the model accuracy may decrease when predicting the extreme values or values outside the range of the training data.

Recently Random Forest Regression (RFR) technique was used by Pang et al. (2022) to predict the wheat crop yield at the regional and local scales in Australia for 2018. The models were trained using NDVI, meteorological variables, and yield data from three states, i.e., Victoria, New South Wales, and South Australia. The model performance was evaluated using four metrics, i.e., R^2, MAE, Mean Squared Error (MSE), and RMSE. The RFR model performed well on a regional scale with R^2 of 0.86 and RMSE of 0.18 t/ha. On a local scale, it performed well for Victoria and New South Wales with R^2 of 0.89 and 0.87 and RMSE of 0.15 t/ha and 0.07 t/ha, respectively. However, South Australia showed poor performance compared to the other states, with R^2 of 0.45 and RMSE of 0.45 t/ha. The results highlighted that the model was overpredicting the small values and underpredicting the high values.

Similarly, Roell et al. (2020) compared the RF and traditional soil-based models of historic yield potential for wheat using soil, climate, and topography variables for Denmark from 1992-2018. The model performance was assessed with four metrics, i.e., R^2, MSE, RMSE, and Lin’s

concordance correlation coefficient. The RF model performed better than the old soil-only model with R^2 of 0.26 and RMSE of 1256 kg/ha. While for soil only model, the R^2 was 0.20, and RMSE was 1982 kg/ha, which was significantly higher (726 kg/ha) than the RMSE of RF.

Hunt et al. (2019) estimated the within-field yield variability for the wheat crop using an RFR model in the United Kingdom in 2016. Five different vegetation indices from Sentinel-2, including NDVI, GCVI, GNDVI, simple ratio, and WDRVI, were used to train the model. A ten-fold cross validation approach was used, and the performance was evaluated using RMSE. The model showed an RMSE of 660 kg/ha. Furthermore, environmental data, including meteorological, soil moisture, and topography, were also combined with Sentinel-2 data. The model showed an improvement with an RMSE of 610 kg/ha.

Fieuzal et al. (2020) also used the RF model to estimate the wheat yield on a field level in Gers County in southwestern France from 2014 to 2017. The model was trained using Landsat-8 and Sentinel-2 derived NDVI and crop rotation maps. The model performance was assessed using two metrics, i.e., R^2 and RMSE. The model showed R^2 ranging from 0.40 - 0.60 and RMSE ranging from 813 kg/ha – 700 kg/ha for different years. Moreover, the model effectively predicted the yield three months before the harvest.

Marshall et al. (2022) conducted a study for field-level yield prediction at Bonifiche Ferraresi farm in Italy for wheat, corn, soybean, and rice crops of the year 2020. Multiple methods were used to predict the yield, including RF, Partial Least Squares Regression (PLSR), and two-band vegetation indices (TBVIs). The models were trained on the Sentinel-2 and Precursore Iperspettrale Della Missione Applicativa (PRISMA) spectral bands at the crop's vegetative, reproductive, and maturity stages. The model performance was evaluated using the R^2 metric. RF outperformed the other models and was able to explain 20% more variability in the yield compared to the PLSR and TBVIs.

Khan et al. (2022) compared multiple machine learning models such as Geographically Weighted Random Forest Regression (GWRFR), MLR, RFR, PLSR, Decision Tree Regression (DTR), and Support Vector Regression (SVR) to estimate corn yield on the county level in the United States.

The models were trained using climate, soil, vegetation indices, and yield data, and the performance of the models was evaluated based on R^2 and RMSE. GWRFR outperformed all other machine learning models with R^2 of 0.901 and RMSE of 0.764 MT/ha, followed by SVR (R^2=0.889, RMSE=0.811), RFR (R^2=0.828, RMSE=1.006), MLR (R^2=0.792, RMSE=1.107), PLSR (R^2=0.791, RMSE=1.111) and DTR (R^2=0.743, RMSE=1.231). The study found that GWRFR can better address the spatial heterogeneity as Moran's I value of the residuals was smaller than the other machine learning models.

In a study by Wang et al. (2016), fifteen vegetation indices were used to estimate the wheat yield in Jiangsu province of China from 2010-2014. RFR, SVR, and ANN models were used, and their performance was assessed with R^2 and RMSE metrics. RFR performed better than SVR and ANN with R^2 of 0.533 in the jointing stage, R^2 of 0.721 in the booting stage, and R^2 of 0.790 in the Anthesis stage, followed by SVR and ANN. Moreover, the model robustness of RFR and SVR was found to be equally good in all growth stages; however, the ANN model performed unstably on all growth stages of testing data.

In a similar study, Zeleke et al. (2021) used SVR to predict sugarcane yield in Wonji-Shoa, Ethiopia, and compared the results with MLR and Multilayer Perceptron Neural Network (MLPNN) models. Seven vegetation indices derived from Landsat 8 and Sentinel 2A from 2016 to 2019 were used to train these models, and their performance was assessed with ten-fold cross validation and R^2, RMSE, and MAE evaluation metrics. The SVR model outperformed the other models with R^2=0.74 and RMSE=12.95 t/ha, followed by MLPNN (R^2=0.69, RMSE=15.59 t/ha) and MLR (R^2=0.55, RMSE=17 t/ha).

Johnson et al. (2016) compared the nonlinear machine learning models, including Bayesian Neural Networks (BNN) and Model-Based Recursive Partitioning (MOB) with MLR for predicting wheat, barley, and canola yield using NDVI and EVI in the Canadian Prairies between 2000-2011. The model performance was assessed with MAE and skill score metrics. All models showed similar results for wheat and canola, with skill scores of 0.269 and 0.192, respectively. However, the nonlinear models (BNN, MOB) for barley showed higher performance than the MLR with a skill score of 0.248. The study suggested that non-linear models not showing a significant

advantage over MLR could be due to the short study period and the linear relationship between yield and NDVI or EVI.

Aghighi et al. (2018) evaluated the machine learning techniques, including Boosted Regression Tree (BRT), RFR, SVR, and Gaussian process regression (GPR), to estimate the silage maize yield in the northwest part of Iran. These models were trained on time series derived NDVI from Landsat 8 and yield data from 2013-2015. The models were evaluated with a five-fold cross validation technique, and their performance was assessed with correlation coefficient, RMSE, and MAE metrics. RFR was found to be most stable with standard deviation for RMSE and MAE of only 0.23 and 0.18, respectively, followed by BRT (std RMSE =0.47, std MAE=0.43), SVR (std RMSE =0.74, std MAE=0.57) and GPR (std RMSE =1.39, std MAE=0.96) respectively.

The use of deep learning techniques has become increasingly prevalent alongside traditional machine learning methods for crop monitoring (Pratama et al., 2020), crop type classification (Zhao et al., 2021), land use classification (Wan et al., 2022), plant disease detection (Mohanty et al., 2016), and crop yield prediction (Zhu et al., 2022). Multilayer Perceptron (MLP), Convolutional neural networks (CNNs), and Long Short-Term Memory (LSTM) are widely used in all these applications. Deep learning models can reveal the complex nonlinear relationships between variables by breaking them into small pieces (LeCun et al., 2015). One key difference between traditional machine learning and deep learning methods is the feature extraction process. In conventional machine learning, features are extracted using heuristics. While in deep learning models, multilayers of neural networks automatically extract features from the data. Deep learning models try to simulate the human brain by using multiple hidden layers in the neural network (J. Zhang et al., 2022).

You et al. (2017) have used the LSTM and CNN models for county-level soybean yield prediction in eleven counties in the United States. The models were trained using surface reflectance, LST, land cover type, and yield data from 2003 to 2015. The Deep Gaussian Process component (GP) was also incorporated to model the spatial-temporal structure of the data. The models were evaluated based on RMSE metric against baseline models, including Ridge regression (RR), Decision Trees (DT), and Deep Neural Network (DNN). The CNN-GP outperformed all other

models with an RMSE of 5.55 t/ha, followed by CNN (5.77 t/ha), LSTM-GP (5.83 t/ha), LSTM (6.27 t/ha), DT (7.73 t/ha), RR (7.96 t/ha) and DNN (8.32 t/ha).

In a similar study, LSTM and CNN were used by Mahdi et al. (2020) for district-level potato yield and precipitation prediction in Munshiganj, Bangladesh. The time series data from 2010 to 2017 was used for potato yield prediction and 30-year rainfall data from 1981 to 2019 for precipitation prediction. The GP component was also integrated into the models, and their performance was assessed using the RMSE metric. The CNN-GP performed better among other models used with an RMSE of 7.15 t/ha, followed by LSTM-GP (7.49 t/ha), CNN (7.80 t/ha), and LSTM (8.49 t/ha).

Yang et al. (2019) have explored CNN to estimate the yield in a rice field in Southern China. The model was trained using RGB and multispectral images acquired from unmanned aerial vehicles in 2017. The model performance was evaluated using three metrics, i.e., R^2, RMSE, and Mean Absolute Percentage Error (MAPE), and compared with the linear and quadratic regression models. The result shows that CNN trained with RGB and multispectral data performed significantly better than the regression models.

LSTM and CNN were combined by Sun et al. (2019) to estimate soybean yield at the county scale in the United States. The model was trained using crop growth, weather, LST, surface reflectance, and yield data from 2003 to 2015. The model performance was assessed with RMSE and Percent Error (PE) and compared with LSTM and CNN models alone. The CNN-LSTM model outperformed the other models in both in-season and end-of-season estimations with an RMSE of 329.53 kg/ha, followed by CNN (359.12 kg/ha) and LSTM (363.15 kg/ha).

Similarly, Sharma et al. (2020) also used the CNN-LSTM model to predict winter wheat yield in seven major wheat-growing states of India on the tehsil (block) level. The model was trained using raw satellite data from 2001 to 2011. The model accuracy was assessed using the RMSE metric, and performance was compared with baseline models, including Decision Forest (DF), DT, step regression, RR, and LSTM-GP. The CNN-LSTM outperformed all the baseline models by over 50%, followed by LSTM-GP, RR, DF, step regression, and DT. Moreover, it was observed that

the yield estimates could be further improved by incorporating contextual information, including the location of fields, water bodies, and urban areas.

Fernandes et al. (2017) used a stacking ensemble model with ANN to predict sugarcane yield at the municipality level for São Paulo State, Brazil. The model was trained using a time series NDVI derived from Moderate Resolution Imaging Spectroradiometer (MODIS) and yield data from 2003 to 2012. The model accuracy was assessed using R^2 and RRMSE metrics. The model showed an R^2 of 0.43 and an RRMSE of 8%. The model was able to forecast the yield three months prior to the harvest time with a smaller value of RMSE than the official survey at the state level.

In a recent study, Wang et al. (2022) used the LSTM model to predict winter wheat yield in the Henan province of China. The model was trained using time series LAI derived from MODIS and yield data from 2003 to 2016. The model performance was assessed using R^2, MAPE, and RMSE metrics, and the results were compared with baseline models, including RF, SVR, PLS, and Extreme Gradient Boosting (XGBoost). The LSTM outperformed all the baseline models with R^2 of 0.87 and RMSE of 522.3 kg/ha, followed by SVR (R^2=0.76, RMSE=725.8 kg/ha), RF (R^2=0.72, RMSE=656.6 kg/ha), XGBoost (R^2=0.72, RMSE=785.8 kg/ha), and PLS (R^2=0.7, RMSE=809.1 kg/ha). The study showed that traditional machine learning models didn't perform well because the ability of these models to analyze nonlinear relationships is not as good as LSTM, and they also don't consider the time correlation with yield.

Tian et al. (2021) used the LSTM model to estimate the wheat yield at Guanzhong Plain, China. The model was trained using meteorological data and remote sensing indices, such as Vegetation Temperature Condition Index (VTCI) derived from LST, NDVI, and LAI, and yield data from 2007 to 2017. The model performance was evaluated using RMSE and R^2 metrics, and the results were compared with baseline models, including SVM and BPNN. The LSTM model outperformed the baseline models with an R^2 of 0.83 and RMSE of 357.77 kg/ha, followed by BPNN with an R^2 of 0.42 and RMSE of 812.83 kg/ha, and SVM with an R^2 of 0.41 and RMSE of 867.83 kg/ha.

Bali & Singla. (2021) explored the LSTM to estimate wheat crop yield in the Ludhiana district of India. The model was trained using climate and yield data of 43 years ranging from 1970 to 2012.

The model accuracy was assessed using MAE, RMSE, and MSE metrics, and the results were compared with traditional machine learning models, including RF, MLR, and ANN. The LSTM outperformed all the other models with RMSE of 147.12 kg/ha and MAE of 60.50 kg/ha, followed by RF (RMSE = 540.88 kg/ha, MAE = 449.36 kg/ha), ANN (RMSE = 732.14 kg/ha, MAE = 623.13 kg/ha), and MLR (RMSE = 915.64 kg/ha, MAE = 796.07 kg/ha). MLR showed poor performance as the climate factors had nonlinear behavior, which is difficult to capture for MLR.

Pantazi et al. (2016) compared different neural networks, including Counter-Propagation Artificial Neural Networks (CP-ANNs), XY-fused Networks (XY-Fs), and Supervised Kohonen Networks (SKNs) to predict the within-field variation in wheat yield prediction using soil and crop growth data in a small field of UK for single cropping season of 2013. The accuracy of the models was evaluated, and SKN performed better than other models with an accuracy of 81.65%, followed by XY-F with an accuracy of 80.92 and CP-ANN with an accuracy of 78.3%. The study showed that SKN can classify the field area into potential yield zones.

Recently a novel approach of Deep Learning Adaptive Crop Model (DACM) has been proposed by Zhu et al. (2022) to estimate the soybean yield in ten states of the United States. The model uses crop growth feature extraction, Gray Level Cooccurrence Matrix (GLCM), and LSTM to incorporate crop monitoring and spatial heterogeneity. The model was trained using crop classification, remote sensing metrics, and yield data from 2003 to 2017. The model accuracy was assessed using four metrics, i.e., R^2, RMSE, MAPE, and mean percentage error. The results were compared with other machine learning and deep learning models, such as SVR, RFR, DNN, CNN, and LSTM. The DACM outperformed all the popular models with an RMSE of 296.304 kg/ha and R^2 of 0.805, followed by LSTM (R^2=0.781, RMSE=315.403 kg/ha), CNN (R^2=0.754, RMSE=333.627 kg/ha), DNN (R^2=0.739, RMSE=344.925 kg/ha), RFR (R^2=0.707, RMSE=365.235 kg/ha) and SVR (R^2=0.435, RMSE=507.334 kg/ha).

To summarize, it can be said that a lot of work has been done to estimate crop yield using different methods and metrics. However, very little of this work done is at a field level, especially which uses earth observation data and artificial intelligence. Moreover, there is no apparent convergence

or consensus about which approach or metrics are most suitable. Hence, no general or context-based (area, scale of study, type, and data quality) standard practices have yet been developed.

Chapter 3: Aims and Objectives

3.1 General Context

As established earlier, accurate crop yield estimation is important for farmers and policymakers to make informed decisions about agricultural practices, production, and food security. In general, national statistics offices are responsible for collecting the crop yield data. Most of them use uncertain data sources or traditional methods, such as eye estimation, farmer surveys, etc., to collect them. When using these uncertain data sources, even a small measurement error can lead to a significant difference in the estimated yield and crop production at a higher level. Indeed, when these few samples are aggregated and considered as a representation of the entire region or nation, the error can increase exponentially, leading to inaccurate estimates and potential misinterpretation of the actual crop yield.

This inaccurate crop yield estimation can have several negative consequences for farmers and policy makers. From a farmer's perspective, inaccurate yield estimates can lead to the inefficient use of resources, including labor, fertilizers, and pesticides. This inefficiency can lead to increased costs and reduced profits for farmers, affecting the sustainability of agricultural practices. From the policymaker’s perspective, inaccurate yield estimates can lead to misinformed food security and economic development decision-making. The inaccurate yield estimates can result in overestimating or underestimating specific crop production, leading to food security, trade and logistics, and economic development issues.

To summarize, improving the current methods used to estimate crop yield is important to progress toward accurate and reliable data collection. This can be achieved using modern technologies, such as remote sensing and artificial intelligence, which can provide accurate and timely data for crop yield estimation. These methods have the potential to revolutionize the way crop yield data is collected and analyzed so far, leading to more accurate estimates and informed decision-making.

3.2 Objectives

The study aims to develop a state of the art wheat crop yield estimation method at the field level that uses the remotely sensed features derived from satellite earth observation optical time series by leveraging artificial intelligence. In this context, the research question of this thesis is:

To what extent the yield estimation at the field level using machine learning and deep learning techniques is more accurate than the multivariate regression estimation methods?

To accomplish the goal of this study, the following steps are outlined:

1. Selection of appropriate machine learning and deep learning methods to deal with remote sensing time series.
2. Identification of the appropriate remotely sensed features that correlates to the crop yield.
3. Comparison of the results of different models, including multivariate regression models, machine learning models, and deep learning models based on a set of different performance metrics.

Chapter 4: Materials

This chapter will provide background information regarding the study area and the datasets used in this research. Section 4.1 provides a comprehensive description of the study area with context. Section 4.2 describes the climate conditions during the study period. Furthermore, section 4.3 delves into the data sources utilized and outlines the various data types employed. Finally, section 4.4 focuses on the quality assurance and quality control measures implemented on the in-situ data.

4.1 Study Area

The research is carried out in specific parts of multiple regions of Spain, including Castilla y Leon, located in the North-Western part of Spain, and Castilla-La-Mancha and Comunidad Valenciana, which are in the South-Eastern region of Spain, as shown in Figure 1. The study is carried out on a field level for the wheat crop for the year 2019. These specific areas were chosen because of the comprehensive availability of in-situ data. The small, black-colored polygons are the wheat fields in the study area, and their in-situ data were available. For Comunidad Valenciana, only one wheat field with in-situ data was available. The mean annual temperature of Castilla y Leon, Castilla-La-Mancha, and Comunidad Valenciana for 2019 was 12.27 °C, 14.34 °C, and 15.94 °C, respectively (World Bank, 2019). Similarly, the mean annual precipitation of Castilla y Leon, Castilla-La-Mancha, and Comunidad Valenciana for 2019 was 593.06 mm, 406.81 mm, and 460.13 mm, respectively (World Bank, 2019).

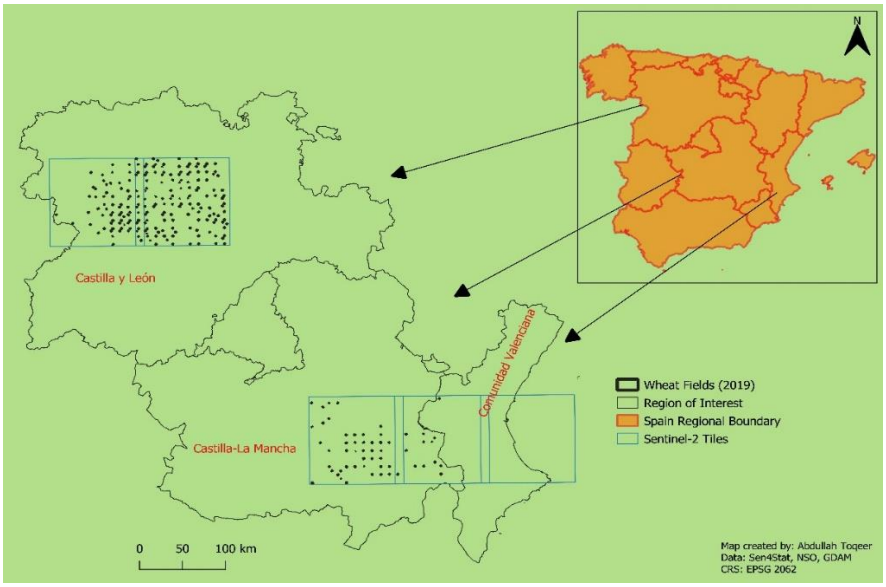

Figure 1: Study Area Map

In Spain, multiple crops are grown, including barley, corn, cotton, oats, rice, rye, sorghum, sunflower, and wheat (U.S. Department of Agriculture, 2023). These crops have different sowing, growing, and harvesting times based on their characteristics. The crop calendar for wheat is shown in Figure 2.

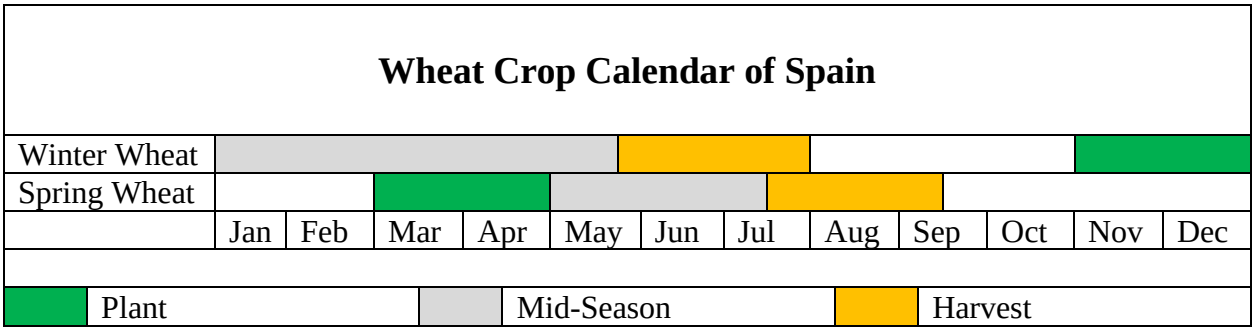

Figure 2: Crop calendar of Spain for Wheat Crop (Source: U.S Department of Agriculture, 2023)

Wheat (scientific name: Triticum aestivum L.) is an important crop in Spain, contributing to 3% of European wheat production. In Spain, most of the wheat production is being done in central and southern Spain, specifically in the regions of Castile (Castilla y Leon and Castilla-La-Mancha) and Andalusia (Segarra et al., 2020).

According to the Spanish Agricultural records, the average wheat yields for 2019 of Castilla y Leon, Castilla-La-Mancha, and Comunidad Valenciana for rainfed and irrigated land are shown in Table 1. As expected, the yield is significantly higher in irrigated fields than in rainfed fields. In Castilla y Leon, for rainfed and irrigated land, it was 2753 kg/ha and 4923 kg/ha, respectively. Similarly, in Castilla-La-Mancha, for rainfed and irrigated land, it was 1602 kg/ha and 5533 kg/ha, respectively. The wetter conditions of Castilla y Leon reduce the contrast between irrigated and rainfed wheat yield, unlike in Castilla-La-Mancha, where the irrigation triples the yield. For Comunidad Valenciana, it was 1446 kg/ha and 1922 kg/ha for dry and irrigated lands (Encuesta sobre Superficies y Rendimientos de Cultivos, 2019).

Table 1: Wheat Yield Records for 2019 (Source: Encuesta sobre Superficies y Rendimientos de Cultivos, 2019)

Area	Type of land	Number of samples (plots)	Yield (kg/ha)
Castilla y Leon	Rainfed	2368	2753
Castilla y Leon	Irrigated	202	4923
Castilla-La-Mancha	Rainfed	533	1602
Castilla-La-Mancha	Irrigated	40	5533
Comunidad Valenciana	Rainfed	8	1446
Comunidad Valenciana	Irrigated	22	1922

4.2 Climate Conditions of Spain in the 2019 Growing Season

The European Commission's Joint Research Centre (JRC) Monitoring Agricultural Resources (MARS) unit issues a monthly report about crop growing conditions and quantitative yield estimations within the European Union. According to JRC MARS, the months of January and February 2019 were marked as rain deficit in Southeastern Spain (JRC MARS Bulletin February, 2019; JRC MARS Bulletin January, 2019). This decreased water reserves in March across Spain to 58% of the reservoir capacity, which was ten percentage points lower than the 10-year average (JRC MARS Bulletin March, 2019). In April 2019, the drought conditions continued, which aggravated the risk of irrigation restrictions in Castilla y Leon, Comunidad Valenciana, and the eastern part of Castilla la Mancha, resulting in below average biomass formation for wheat crops in these areas (JRC MARS Bulletin April, 2019). By May 2019, the drought conditions were improved in Castilla y Leon. However, the soil moisture was still low, making the region dependent on further rain supply in the coming weeks (JRC MARS Bulletin May, 2019). In June 2019, the drought conditions increased in the southeast region as there was a rainfall deficit of 40-60 mm for May. This led to reduced yield overall, particularly for the winter crops in Castilla y Leon (JRC MARS Bulletin June, 2019). The dry conditions continued in July 2019, with a heatwave observed in the first week of the month (JRC MARS Bulletin July, 2019). The months of August and September were also drier and warmer than in the past. This led to low NDVI values in Castilla La Mancha and low biomass production in southern Castilla y Leon (JRC MARS Bulletin August, 2019; JRC MARS Bulletin September, 2019). The situation improved in November, with heavy rainfall in northeastern Spain (JRC MARS Bulletin November, 2019).

4.3 Data Sources and Data Preparation

The current study utilized three different types of data including in-situ, meteorological, and remote sensing. The in-situ data was acquired from the National Statistics Office of Spain. The meteorological data was taken from the Fifth Generation European Reanalysis Dataset for Land (ERA5 Land) database available online through the climate data store. The remote sensing data was taken from Sentinel-2 of the Copernicus program. The data description and preparation for each type are explained below:

4.3.1 In-Situ Data

The in-situ data used for training and validation was taken from the National Statistics Office of Spain for the year 2019. There were 849 fields in the study area from Castilla y Leon, Castilla-La-Mancha, and Comunidad Valenciana. The data shared was in shapefile format and had multiple pieces of information. Firstly, the field's number, shape, and location were present in the data. Secondly, the different yield estimation approaches used were available. These approaches include counting of spikes and grains, counting and weighing spikes, data provided by the farmers, eye estimation, weighing spikes, and other methods. Out of 849 yields, 34 were estimated with the count of spikes and grains approach, 42 with counting and weighing spikes, 28 with data provided by the farmer, 661 with eye estimation, 15 with weighing spikes, and 69 with other methods.

Thirdly, the type of field, which includes irrigated and rainfed, was mentioned. Lastly, the wheat yield value in kilogram per hectare was available. On average, the crop yield values of irrigated fields were significantly higher than the rainfed fields. Out of 849 fields, 720 were situated in rainfed areas and had an average yield of 2079 kg/ha, and the remaining 129 fields were in the irrigated area with an average yield of 5300 kg/ha.

4.3.1.1 In-situ Data Preparation

The quality check and control of the in-situ data is very important as different acquisition modes of yield values were used for different fields. According to Gozdowski et al. (2010) and Sudduth & Drummond (2007), around 10% to 50% spatial outliers or incorrect data exist in the raw yield data. In general, these outliers are removed by using statistical methods and setting a range of

biological potential of crop yield (Gozdowski et al., 2010; Simbahan et al., 2004). The outliers in the present study were identified by comparing the yield values with LAI, as LAI strongly influences crop yield. The LAI and crop yield are positively correlated in a linear way (Su et al., 2022). Similarly, J.a et al. (1994) reported that linear and quadratic relationships exist between LAI, crop yield, and defoliation.

The histogram method was used to identify the outliers in which the frequency of yield values was plotted, and all the values outside the set threshold were excluded. The threshold used to remove the outliers was from 1001 kg/ha to 6540 kg/ha. This threshold was set based on the relationship between LAI and yield values, as it was observed that LAI values outside this range did not exhibit a significant correlation with the corresponding LAI values. This approach led to the expulsion of 91 yield values (10.7 % of the total data). Some outliers were obvious, which, if used in the training data set, would significantly affect the model efficiency and yield prediction. For example, one field has a crop yield value of 1 kg/ha. A graph of the number of observations (frequency) and range of yield values, along with the density curve, is shown in Figure 3. The outliers range, i.e., below 1001 kg/ha and above 6540 kg/ha are marked in the green line. The frequency bars within the range are marked in blue and outside the range in red.

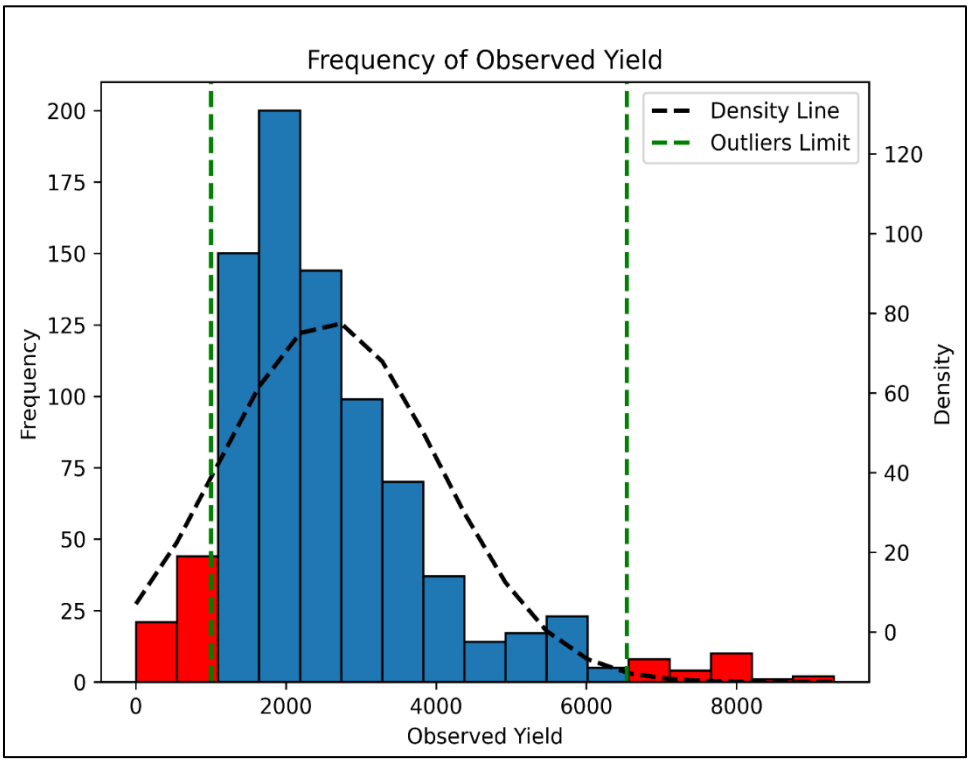

Figure 3: Frequency of Observed Yield Before Cleaning the Data

Similarly, within this acceptable yield range (1001 kg/ha – 6540 kg/ha), some fields still had very high yield but low LAI value or very low yield with high LAI values. Most of these outlier yield values were recorded using the eye estimation method, which explains why they are incorrect. Furthermore, 115 yield values (13.5 % of the total data) were excluded from the data by comparing within threshold yield values with LAI. In total, 206 yield values (24.2 %) were excluded from the in-situ data. Figure 4 shows the graph of the final frequency distribution of in-situ data. All the values are within the acceptable threshold. The minimum and maximum yield value in the data is 1100 kg/ha and 6500 kg/ha, respectively, and the mean yield of data is 2571 kg/ha. It is evident from the frequency distribution that most of the yield values (92.6 %) lie in the range of 1100 kg/ha – 4000 kg/ha, and only a few yield values (7.4 %) are available in the data from 4001 kg/ha – 6500 kg/ha.

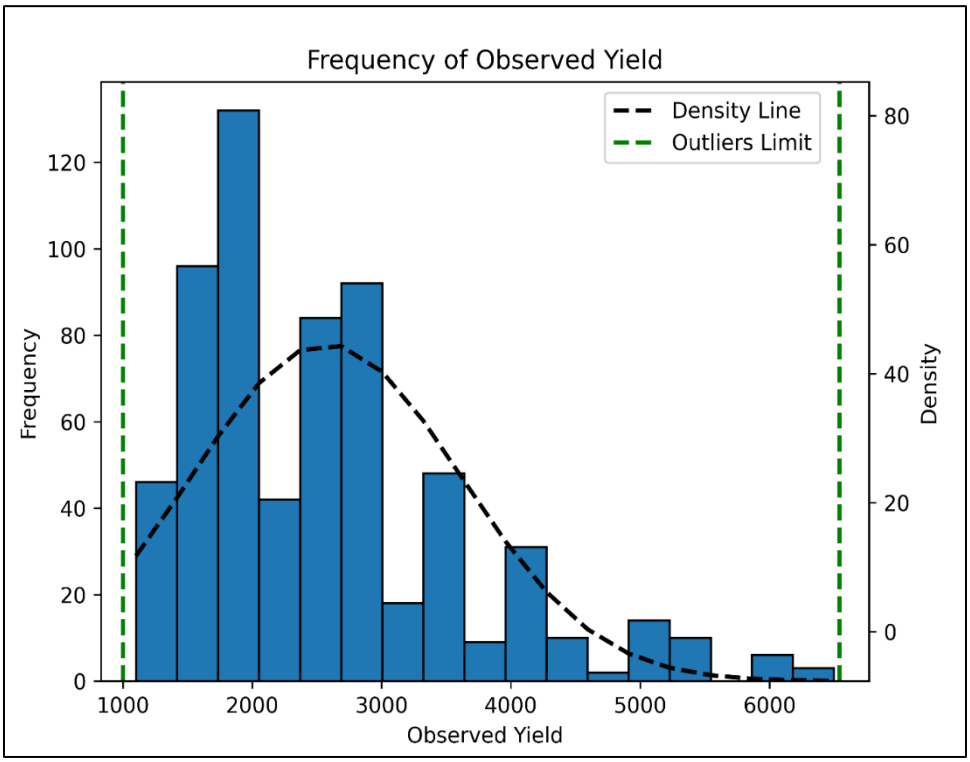

Figure 4: Frequency of Observed Yield After Cleaning the Data

After cleaning the data and removing the outliers, the data was split into training and validation data subsets. The split was randomly done using the 8:2 ratio, with 80% data in the training data subset and the remaining 20% data in the validation data subset. It is a common approach in machine learning and deep learning to train and validate the models. A similar 8:2 approach was

used by (Bali & Singla, 2021; Barriguinha et al., 2022; Forkuor et al., 2017; Marszalek et al., 2022) in crop yield prediction.

4.3.2 Meteorological data

The current study used temperature, precipitation, potential evapotranspiration, global radiation, and volumetric soil water data for 2019. The meteorological data was extracted from the ERA5-Land database available online on the climate data store of Copernicus. The data is available from 1981 to present on an hourly basis. The resolution of data is 0.1° latitude and 0.1° longitude, corresponding to a grid of approximately 9 km. The hourly data was converted into daily data. The key meteorological variables used were mean air temperature (°C), total global radiation ($KJ/m^2/day$), sum of precipitation (mm/day), potential evapotranspiration from a crop canopy (mm/day), and volumetric soil water (m^3m^{-3}) at four different depth levels which were 0 cm – 7 cm, 7 cm – 28 cm, 28 cm – 100 cm and 100cm – 289 cm (Sen4Stat, 2020). The European Space Agency (ESA) Sentinels for Agricultural Statistics (Sen4Stat) project acquired the meteorological data and processed it completely.

4.3.3 Remote Sensing Data

The remote sensing data used in the current study was taken from the Sentinel-2 satellite of the Copernicus program for 2019. Sentinel-2 L2A provides atmospherically corrected multispectral images in thirteen bands in visible, Near-Infrared, and Short-Wave Infrared ranges. Bands 2, 3, 4, and 8 are available in 10 meters spatial resolution. Bands 5, 6, 7, 8A, 11, and 12 are in 20 meters spatial resolution, and bands 1, 9, and 10 are in 60 meters spatial resolution. The temporal resolution is five days (Sentinel, n.d.).

The cloud screening on the satellite images was done using the Fmask algorithm. Furthermore, the mean LAI was computed for every field using cloud-free images and field boundaries. The LAI pixels whose center fits within the field's boundary, which were reduced by a negative buffer of 15 meters, were averaged. LAI values that were in-valid due to being derived from L2A data flagged with cloud, cloud-shadow, or due to being outside the acceptable range based on the quality assessment of the Biophysical Variable Neural Network (BV-NET) calibrated on wheat in

Belgium were excluded from the averaging process. For each field, only the LAI observations with more than 75% of the valid pixels were kept. An interpolation of the LAI observation time series using the Savitzky-Golay fitting algorithm was applied to generate an LAI estimation for each day of the season (Sen4Stat, 2020). The remote sensing data was also acquired and processed by the Sen4Stat project.

4.3.4 Features Extraction

The Savitsky Golay smoothed time series was used to derive the three growing periods (P_1, P_2, and P_3). The P_1, or emergence stage, covers the period from emergence to mid-growth level, starting from 10% of maximum LAI to 50% of maximum LAI. The P_2, or flowering stage, covers the period from mid-growth to the point where it reaches maximum LAI. The P_3, or senescence stage, covers the period from the maximum LAI point to when the LAI reaches 10% of the maximum again. An example of these growing phases based on smoothed LAI time series is shown in Figure 5 (Sen4Stat, 2020).

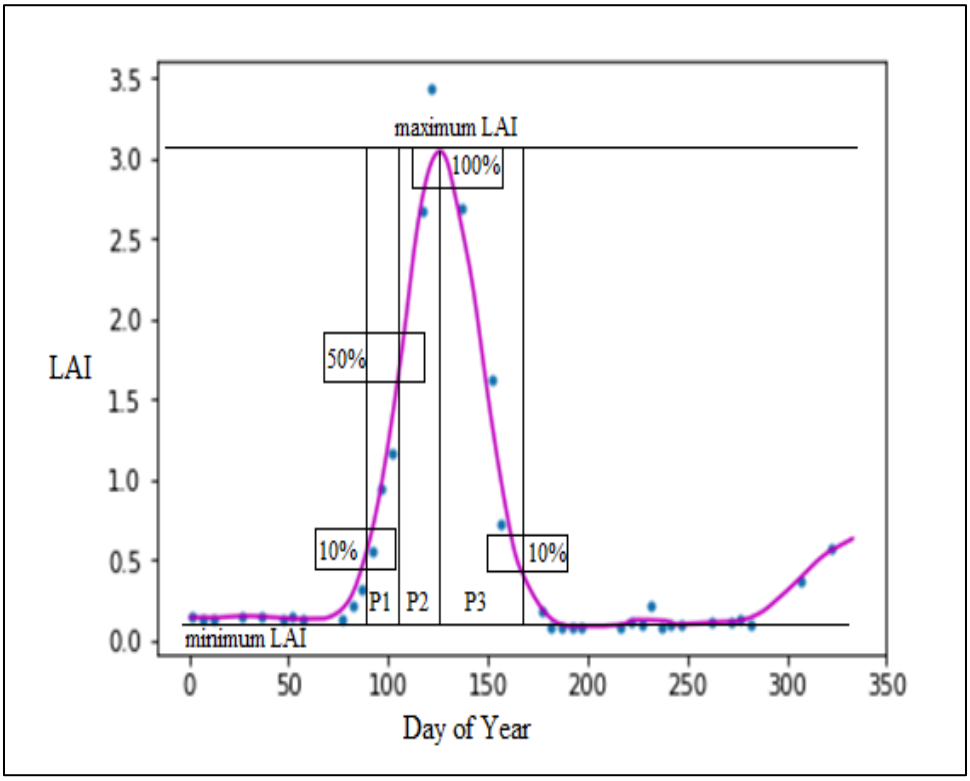

Figure 5: Growth Stages of Crop Based on LAI Time Series (Source: Sen4Stat, 2020)

On these three growing periods, features related to vegetation status (LAI) and weather data (temperature, potential evapotranspiration, precipitation, solar radiation, and volumetric soil water) were computed for every field. The SAFY simulated yield and parameters were also used as features. In total, 43 features corresponding to four groups of vegetation, weather, soil moisture, and crop growth model were computed. The complete list of 43 features and their description is attached in the Appendix. The complete workflow of the Sen4Stat project of computing 43 features is shown in Figure 6 (Sen4Stat, 2020). The Sen4Stat project processed the meteorological and remote sensing data and derived these features for further analysis.

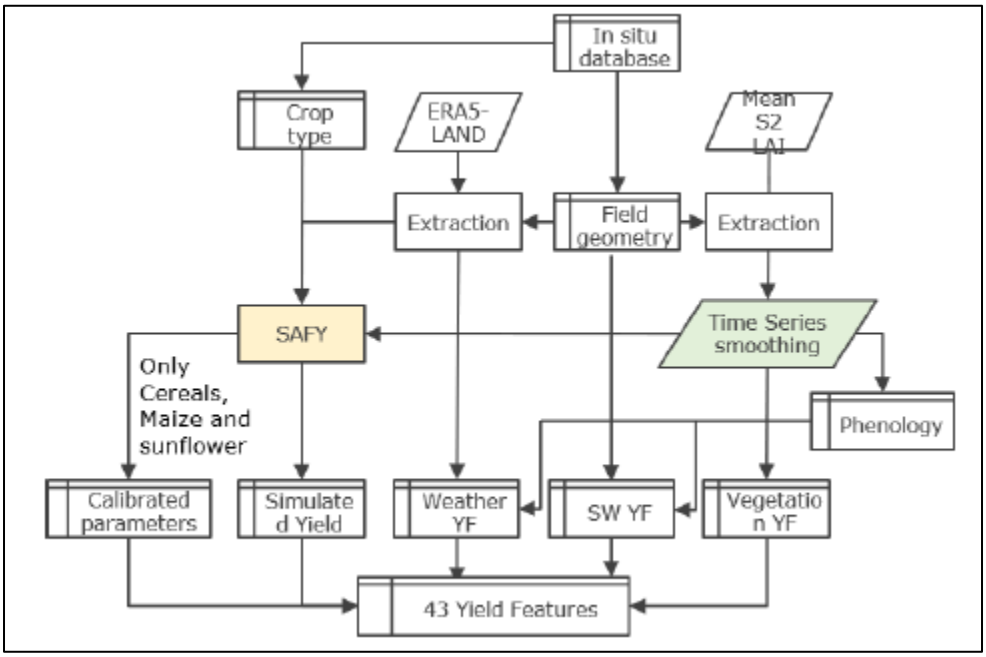

Figure 6: Sen4Stat Feature Extraction Flow Chart (source: Sen4Stat, 2020)

Chapter 5: Methods

This chapter presents a comprehensive overview of the methodology used for crop yield estimation at the field level and the identification of important remote sensing features. Section 5.1 describes the various multivariate regression, machine learning, and deep learning models used for crop yield estimation. In section 5.2, the methodology for determining the optimal hyperparameters of these models is discussed. Section 5.3 briefly describes the model training and verification approach used to identify the most suitable model based on the optimized hyperparameters. Furthermore, section 5.4 describes the model validation approach on independent data. Section 5.5 describes the different model evaluation metrics used to assess the models' performance. Lastly, Section 5.6 describes the approach used to identify the key remote sensing features important in crop yield estimation at the field level.

5.1 Selection of Models

Based on the literature review, the present study selects a set of various multivariate regression, machine learning, and deep learning models to estimate the wheat crop yield at the parcel level. For the multivariate regression model, the LASSO model was chosen. Similarly, DTR, RFR, and SVR were selected for machine learning models, and the MLP model was chosen for the deep learning model. The description of the models is as follows:

5.1.1 Multivariate Regression Model

5.1.1.1 Least Absolute Shrinkage and Selection Operator Model

LASSO is a widely used LR model for dealing with regression problems, especially when working with high-dimensional data. LASSO selects a subset of the predictor variables that are most important for predicting the target variable. It also shrinks the coefficients of the remaining predictor variables, which are unimportant, towards zero. LASSO achieves it by introducing a penalizing term to ordinary least squares objective function. This objective function is proportional to the sum of the absolute errors of the coefficients of the predictor variables. The penalizing term used in LASSO is shown in Equation 1 (Pedregosa et al., 2011).

$$min_w \frac{1}{2n_{samples}} ||X_w - y||\frac{2}{2} + \alpha||\omega||_1 \qquad \text{(Equation 1)}$$

Where α is a constant and $||\omega||_1$ is the *l*-norm of the coefficient vector. The working principle of LASSO is iteratively optimizing the objective function, reducing the sum of squared errors between the predicted and observed values, subject to the penalty term. The tuning parameter in the objective function, alpha, controls the strength of the penalty and determines the amount of shrinkage applied to the coefficients of predictor variables. The LASSO model used in the current study uses multiple hyperparameters, including "alpha" which controls the strength of penalty, "max_iter" which controls the total number of iterations, "tol" which is tolerance for optimization and "selection" which indicates the method used for predictor variable selection.

The input data is also scaled before training the model, as without scaling, the LASSO model may not be able to weigh the impact of each input feature in data properly. The data is scaled using the standard scaling technique, where the data is transformed with a unit standard deviation and zero mean.

5.1.2 Machine Learning Model

5.1.2.1 Decision Tree Regression

DTR is a simple and commonly used supervised machine learning algorithm for solving regression problems. It works by splitting the root node, which is the entire dataset, into the subsets. The splitting is based on minimizing the residual sum of squares, i.e., the deviation between the predicted and observed values of the target variable. In each split, the DT algorithm creates a decision node that splits the data into two homogeneous subsets, creating a tree-like structure with decision nodes and leaf nodes at every stage. The splitting will stop once the stopping criteria are met, which involves the different hyperparameters, including the maximum depth of the tree, minimum number of yield points required to split a decision node, and minimum number of yield points required to be at a leaf node, etc. (Pedregosa et al., 2011).

In decision tree regression, the algorithm uses a decision node to represent a splitting decision based on a specific predictor variable and a threshold value. Similarly, the algorithm uses leaf nodes to represent a range of input feature values, assigning a constant value to each leaf node. A

prediction is done for input by navigating the tree from a root node to a leaf node that matches the range of input feature values and returns the constant value of that node as the predicted value for the target variable. The DTR model used in the current study uses multiple hyperparameters, including “max_depth” which determines the maximum depth of the tree, and “min_samples_leaf” which specifies the minimum number of yield points required to be a leaf node. These hyperparameters serve as a stopping criterion for the tree formation of the model. Moreover, like LASSO, the input data for DTR is also scaled using a standard scaling technique to avoid any bias of the input features.

5.1.2.2 Random Forest Regression

RFR is another widely used supervised machine learning model to solve regression problems. The model is built on a large number of decision trees, where each DT is based on a randomly selected subset of training data and input features based on the bootstrap aggregation technique. Random selection helps reduce the correlation between the trees to improve the diversity, cover all input data aspects, and avoid overfitting. The prediction from each DT is then averaged to get a final prediction. This averaging of the predictions of all decision trees helps reduce variance in the predictions. A single DT works on the same principle as explained above in the decision tree regression model.

Similar to the DTR, RFR also uses a stopping criterion to finalize the model when the conditions are met. Multiple parameters including the total number of trees, maximum depth of the tree, minimum number of yield points required to split a decision node, and minimum number of yield points required to be a leaf node, etc. (Pedregosa et al., 2011). The RFR model used in the current study uses multiple hyperparameters, including “n_estimators” which controls the total number of trees in the forest, “max_depth” which controls the maximum depth of the tree, and “min_samples_leaf” which controls the minimum number of yield points required to be a leaf node. These hyperparameters prevent the overfitting of the trees and improve the model's overall performance.

The input data for RFR is also scaled using the standard scaling technique. If the input features have different scales, some may be given more importance than others, leading to a biased model.

5.1.2.3 Support Vector Regression

SVR is also a supervised machine learning model used for solving regression problems. It is based on the SVM algorithm and adapted for regression problems. The working principle of SVR is to find the best-fitted line with maximum data points. This best-fitted line is called a hyperplane. In simple LR models like LASSO, the goal is to minimize the sum of squared errors between the predicted and observed values. However, SVR aims to find the hyperplane within a threshold. This threshold refers to the margin of error between the predicted and observed values that is acceptable to us. SVR maps the predictor variables into a high dimensional space to identify the nonlinear relationships between the predictor variables and target variable. This mapping is done using a kernel function. SVR is more robust than the traditional LR models, as they can identify the nonlinear relationships between the predictor and target variables.

The SVR model used in the current study uses multiple hyperparameters, including “kernel” which specifies the kernel type used in the model to find the hyperplane, “C” which is the regularization parameter to have a balance between the model’s ability to fit the data and its generalization performance, “gamma” which is kernel coefficient, and “epsilon” which specifies the margin of error or tolerance acceptable to us.

The input data is also scaled before training the model as the SVR model uses distance metrics to find the support vectors, and different scales of predictor variables might affect the model's performance. The data is scaled using standard techniques similar to the one used for LASSO, RFR, and DTR models.

5.1.3 Deep Learning Model

5.1.3.1 Multilayer Perceptron

MLP is a widely used supervised artificial neural network for solving the regression and classification problems. It mainly consists of an input layer followed by one or more hidden layers and then an output layer. Each layer consists of one or more neurons or nodes that receive input from the previous nodes of the previous layer, process it using the activation function, and pass the output to the corresponding next layer nodes. These nodes are connected by weights (w) and

bias (b) which shows the strength of the connection between them. A simple linear neuron unit is shown in Figure 7. The equation of linear unit can be written as

$$y = wx + b \qquad \text{(Equation 2)}$$

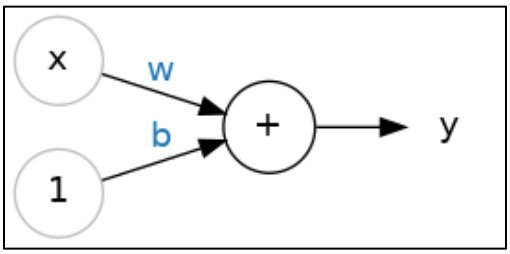

Figure 7: Simple Neuron Unit (Source Kaggle, n.d.)

The value of the weights is adjusted during the training process based on minimizing the difference between the predicted and observed values. Similarly, the bias is also a weight not connected with input data, so the node can modify its input without being dependent on input data. The MLP model uses a non-linear activation function to capture the data's nonlinearity to learn the complex and non-linear relationships. Similarly, a loss function is used to measure the difference between the model's observed and predicted values. This is important to know the progress of model learning and to avoid underfitting or overfitting the model.

The MLP model used in the current study uses multiple hyperparameters, including “activation” which specifies the activation function to learn nonlinear relationships, “input_shape” which specifies the number of predictor variables, “loss” which specifies the loss function to be used, such as MAE, RMSE, etc., “optimizer” which adjusts the weights to minimize the loss, “batch_size” which specifies the number of samples fed to the model during each iteration, “epochs” which specifies the number of times the entire dataset is used to train the model, “early_stopping” to avoid underfitting or overfitting and stop the training of batch if the loss is not minimizing further even the epochs are still left, “Dropout rate” that randomly shut some of the neurons in the layers to avoid any bias in the learning, and “BatchNormalization” to improve the speed of training process (Kaggle, n.d.).

Similar to all other models used, the input data is also scaled for the MLP model using a standard scaling technique, as it uses a gradient-based optimization algorithm to update the weights of neurons in training. If the input features have different scales, the gradients of the cost function with respect to each weight will also have a different magnitude. Due to this, some weights will be updated more quickly than others, which can potentially slow down the convergence of the gradient-based optimization algorithm. Scaling the input data will speed up the convergence of these optimization algorithms, improving the overall performance of the MLP model.

5.2 Optimal hyperparameters of models

The model's performances are greatly dependent on the hyperparameters, and wrong values of these parameters result in underfitting or overfitting, leading to poor performance. The optimal hyperparameters help tune the model to achieve the best results on the given data. The optimal hyperparameters in the current study for LASSO, DTR, RFR, SVR, and MLP were selected using the grid search technique. It works by defining multiple values for all the hyperparameters as a grid. The main concept is to train and evaluate the model using all possible combinations of hyperparameter values specified in the defined grid. The performance will be evaluated using cross validation techniques and evaluation metrics like MAE, RMSE, etc.

In the current study, the grid search technique is implemented using multiple hyperparameters, including “estimator” which specifies the model used for training and testing for hyperparameters, “param_grid” which specifies all the hyperparameters need to be tested and their multiple values as a list, “cv” which specifies the cross validation splitting strategy. A ten-fold cross validation strategy is used in grid search as a similar strategy is used to train the models. The “scoring” hyperparameter is used to specify the evaluation metric, which will evaluate the performance on the validation sets, and “verbose” is used to see the information like training time, the result of every combination, etc.

5.3 Model Training and Verification

In the model training and verification phase, the top 3 hyperparameters combinations (3 different models) with the best scoring value were selected from all possible combinations in the grid search

technique. The aim of this phase was to evaluate their performance and select the best model among them. The model training and verification were done using a k-fold cross validation approach with (k=10), i.e., a 10-fold cross validation technique. The training data subset was divided into ten equal parts, called folds. The experimental design of data splitting into training and validation subsets and further splitting of the training subset data into model training and verification data using 10-fold cross validation is shown in Figure 8.

All Data										
Training Data (80%)							Validation Data (20%)			
Training Data										
Model Training Data							Model Verification Data			
Split 1	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Split 2	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Split 3	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Split 4	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Split 5	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Split 6	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Split 7	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Split 8	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Split 9	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Split 10	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10

Figure 8: 10-fold Cross Validation Approach

Firstly, in the model training phase, the model stability was assessed by training and evaluating the model on the same data, specifically, the folds highlighted in grey in Figure 8 above for each split. The NSE and R^2 were calculated for each split, and the mean and standard deviation were computed on the result of all the ten folds. In the model verification phase, the model was trained on all ten splits by training 9 out of 10 folds and validating the remaining one fold. This process was repeated ten times, ensuring each fold was used for model evaluation once. The validation results from the ten runs were aggregated to get an overall estimate of the model's performance. A similar 10-fold cross validation technique was used by Forkuor et al. (2017), Khan et al. (2022), Khanal et al. (2018), Kuwata & Shibasaki (2015), Peng et al. (2019) and Zeleke et al. (2021) in the model training for the crop yield prediction.

5.4 Model Validation

After obtaining the final trained model through a ten-fold cross validation approach, which showed the highest stability and best values for evaluation metrics in training, the model was further validated using fully independent validation data. The validation data, which is 20% of the total data, was not involved in the model training and verification stages. This independent validation process aimed to assess the model's performance on unseen data and validate its generalization capabilities. All the model performance metrics were calculated on the predictions done in this validation step. A similar model validation and evaluation approach was used in crop yield prediction (Forkuor et al., 2017; Jeong et al., 2016).

5.5 Model Evaluation Metrics

The following evaluation metrics were used to evaluate the model's performances and compare their results with other models.

1. Nash–Sutcliffe Efficiency (NSE)
2. R squared
3. Mean Absolute Error
4. Root Mean Squared Error
5. Relative Root Mean Squared Error
6. Root Mean Squared Logarithmic Error (RMSLE)

These performance metrics are selected based on their ability to provide comprehensive insights into different aspects of the model's predictive capability. R^2 and NSE capture the overall performance and goodness of fit, whereas MAE, RMSE, and RRMSE provide information about the magnitude of the prediction errors. Altogether, these metrics provide complementary perspectives on different aspects of model performance, including accuracy, variability, and overall predictive capability. The description of these metrics and the formulas used to estimate them are as follows:

5.5.1 Nash–Sutcliffe Efficiency

NSE is a commonly used statistical approach to evaluate the model's performance. It compares the predicted crop yield values with the observed ones and determines how well the model predicts the observed data (Jeong et al., 2016). It is calculated as one minus the ratio of the sum of squared differences between the predicted and observed crop yield values normalized by the variance of observed crop yield values (Krause et al., 2005). The formula to calculate it is shown in Equation 3.

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{N}(\text{observed value} - \text{predictive value})^2}{\sum_{i=1}^{N}(\text{observed value} - \text{mean observed value})^2} \qquad \text{(Equation 3)}$$

The NSE value ranges from -∞ to 1. The NSE = 1 indicates that the observed and predicted values are perfectly matched. An NSE = 0 shows that the model is no better than using the mean of the observed values. The NSE value in the range -∞ < NSE < 0 indicates that the mean of observed values is better than the model's predicted values. The NSE value closer to 1 generally indicates that the model is accurate and can be used to predict that particular case (R Documentation, n.d.).

5.5.2 R Squared

R^2 is another commonly used statistical approach to evaluate the model's performance. It tells how well the observed data fit with the regression line. R^2 computes the total proportion of variance in the observed values that can be explained by the variance in the predicted values. R^2 is calculated as one minus the variance ratio of the residuals (the difference between the observed values and predicted values) by the variance of observed values (Pedregosa et al., 2011). The formula to calculate it is in Equation 4.

$$\text{R Squared} = 1 - \frac{\text{Variance (Observed values} - \text{Predicted values)}}{\text{Variance (Observed values)}} \qquad \text{(Equation 4)}$$

The value of R^2 ranges from 0 to 1, with $R^2 = 1$ indicating that the model explains all the variability and has a perfect fit of the regression line with observed data. R^2 near 0 indicates that the model could not explain the variability at all and has a weaker fit of the regression line with observed data.

5.5.3 Mean Absolute Error

MAE is another commonly used statistical approach to evaluate the performance of the regression model by measuring the difference between the predicted and observed values. MAE is measured as the average of the absolute differences between predicted and observed values (Mkhabela et al., 2011; Pedregosa et al., 2011). MAE gives an estimation of how far the predicted values of the model are from the actual values. It is measured in the same units as the input data, which makes it easier to interpret. It is calculated using the formula shown in Equation 5.

$$\text{MAE} = \frac{1}{\text{number of samples (n)}} \sum_{i=0}^{n-1} |\text{Predicted value} - \text{Observed value}| \qquad \text{(Equation 5)}$$

Unlike R^2 and NSE, there is no specific range of MAE values, and they depend on the range of actual and predicted values. MAE gives equal weightage to all errors, irrespective of their magnitude, and is less sensitive to outliers. A lower value of MAE indicates that the model has performed better and can be used in that particular case.

5.5.4 Root Mean Squared Error

Similar to MAE, RMSE is another commonly used statistical approach to assess the regression model's performance by measuring the difference between the predicted and observed values. RMSE is the square root of the averaged squared difference between predicted and observed values (Mkhabela et al., 2011; Pedregosa et al., 2011). RMSE shows how much the model's predictive values deviate from the observed values. Like MAE, it is also measured in the same units as the input data, which makes it easier to interpret. It is calculated using the formula shown in Equation 6.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} (\text{Predicted value} - \text{Observed value})^2} \qquad \text{(Equation 6)}$$

There is no specific range of RMSE values, which depends on the range of actual and predicted values. Unlike MAE, RMSE gives more weightage to larger errors and is more sensitive to outliers.

In other words, it penalizes large errors more heavily than small errors. A lower value of RMSE indicates a better fit between the predicted and observed values, and that model can be used in that particular application. However, the presence of large errors in the prediction can result in a greater value of RMSE (Despotovic et al., 2016).

5.5.4 Relative Root Mean Squared Error

RRMSE is also a statistical approach used to evaluate the performance of regression models. It measures the difference between the predicted and observed values, taking into account the magnitude of the observed values. RRMSE gives an idea of the size of the error relative to the scale of the observed values. RRMSE is estimated by dividing the RMSE by the mean of observed values (Despotovic et al., 2016). It is calculated using the shown in Equation 7.

$$\text{RRMSE} = \frac{\text{RMSE}}{\text{mean(observed values)}} * 100\% \qquad \text{(Equation 7)}$$

The RRMSE is expressed in percentages; the value typically ranges from 0% to 100%. The lower the value of RRMSE, the more model is accurate and can be used in that particular application. However, the RRMSE values are considered together with other evaluation metrics, including R^2, MAE, RMSE, etc.

5.5.5 Root Mean Squared Logarithmic Error

RMSLE is also a statistical method used to assess the performance of the regression models. RMSLE is similar to RMSE; however, it transforms the data first using a natural logarithm ($\log_e$) before calculating the error between the observed and predicted values. This approach is helpful when dealing with the large range of values in the target variable. RMSLE takes the difference between the natural logarithm of observed and predicted values, followed by adding the squared differences, taking their average, and then taking the square root (Despotovic et al., 2016). It is calculated using the formula shown in equation 8.

$$\text{RMSLE} = \sqrt{\frac{1}{\text{n}}\sum_{\text{i}=0}^{\text{n}-1}(\log_{\text{e}}(1 + \text{observed value}) - \log_{\text{e}}(1 + \text{predicted value}))^2} \qquad \text{(Equation 8)}$$

The value of RMSLE ranges from 0 to +∞. The lower the value of RMSLE, the more efficient the model is. A key difference between RMSE and RMSLE is that RMSLE only considers the relative error between the predicted and observed values and doesn't take into account the scale of error. However, RMSE considers the scale of error, and its value increases with large errors. RMSLE penalizes the model more for underestimating the observed value rather than overestimating it. In other words, it gives more weightage to errors made for smaller values. This is used where the overestimation of the predictive variable is acceptable but not underestimation (Saxena, 2020).

5.6 Important Feature Identification

One of the study's objectives is identifying the appropriate remotely sensed features correlating to the wheat crop yield. The importance of different input features was measured with an RFR model using the mean decrease in impurity method. In RF, an ensemble of multiple decision trees is constructed and used for prediction. Each DT is trained on a random subset of the input data and features. For every feature, the mean decrease in impurity method estimates the feature importance by measuring how much the performance of these decision trees decreases if we remove that feature. It is done by calculating the impurity of the data before and after removing the feature and then taking the difference. The greater the impurity decrease, the more important the feature is for prediction. For the final feature importance score, the average over all the decision trees is computed (Pedregosa et al., 2011). The feature importance score is normalized, so their sum equals 1.

The current study calculates the feature importance score at two different levels. First, at the individual feature level. Secondly, at the group level, grouping different features in categories of vegetation, weather, soil moisture, and crop growth model.

The complete methodology of this study is shown in Figure 9.

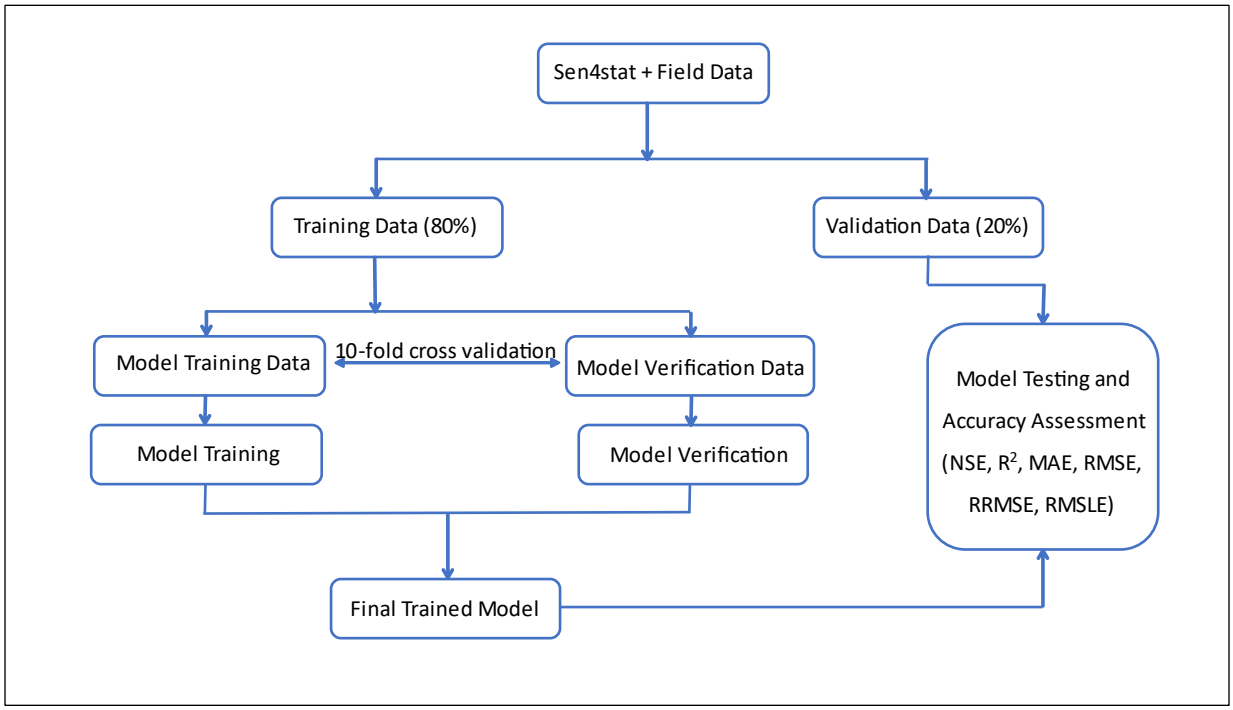

Figure 9: Methodology of the Current Study

Chapter 6: Results

This chapter presents the results obtained by implementing the methods described in Chapter 5. Section 6.1 describes the results obtained from the multivariate regression model's training, verification, and validation phases. Section 6.2 discusses the results of various machine learning models, including DTR, RFR, and SVR, for the training, verification, and validation phases. Section 6.3 reports the results of the deep learning model for the training, verification, and validation phases. Furthermore, section 6.4 compares the performance of the different models using various model evaluation metrics. Finally, Section 6.5 highlights the individual and group importance of the remote sensing features in estimating crop yield at the field level.

6.1 Multivariate Regression Model

6.1.1 Least Absolute Shrinkage and Selection Operator Model

The optimal values for the LASSO model hyperparameters, including alpha (controls the strength of penalty), max_iter (controls the total number of iterations), tol (tolerance for optimization), and selection (method used for predictor variable selection), were obtained using grid search technique. The following grid was used to find the best combination of hyperparameters:

1. alpha = [0.1, 0.5, 1, 5, 10]
2. max_iter = [10, 20, 30, 40, 50, 100, 1000]
3. tol = [0.0001, 0.001, 0.01]
4. selection = [‘cyclic’, ‘random’]

The training data identified alpha = 0.1, max_iter = 10, tol = 0.0001, and selection = ‘cyclic’ as the best values for hyperparameters.

6.1.1.1 Model Training Checking

The stability of the LASSO model with the selected hyperparameters was checked by training and predicting the model on only the training subset of each of the ten folds. The model showed stability in all ten folds, with a standard deviation of 0.009 for both R^2 and NSE.

6.1.1.2 Model Verification

The selected LASSO model was then trained using a 10-fold cross validation approach. The observed and predicted yield values from each fold were combined to estimate the performance of the training phase. The model verification phase showed an R^2, NSE, MAE, RMSE, RRMSE, and RMSLE of 0.40, 0.40, 609.50 kg/ha, 767.26 kg/ha, 0.31, and 0.30, respectively. Figure 10 shows the result of the LASSO model verification phase.

6.1.1.3 Model Validation

In the final stage, the selected model was validated with the validation data (20%), which was not exposed to the model before in any stage. The evaluation of the model on validation data yielded an R^2 value of 0.48 and an NSE value of 0.45. This indicates the model performed well on the unseen validation data, achieving a closer fit between the observed yield and predicted yield values. However, some discrepancies between the observed yield and predicted yield still exist, as indicated by the values of MAE (689.65 kg/ha), RMSE (900.45 kg/ha), RRMSE (0.32), and RMSLE (0.29). These differences may be attributed to the inherent variability in the validation data or potential limitations of the model in capturing all factors influencing crop yield. Figure 11 shows the result of the LASSO model validation.

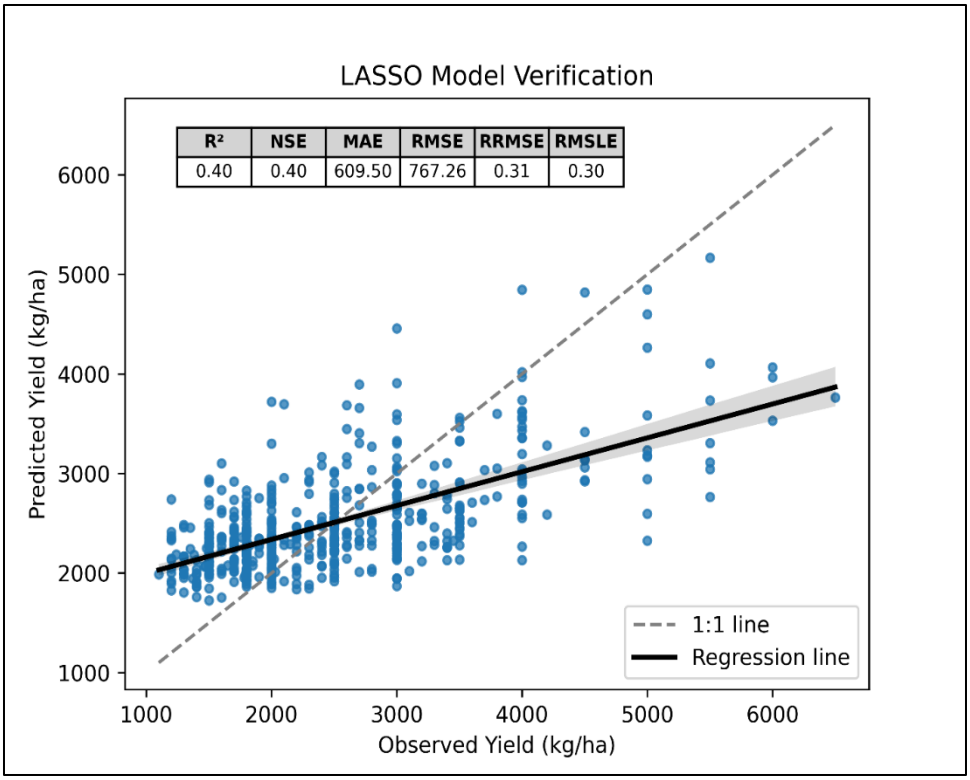

Figure 10:Result of LASSO Model Verification Phase

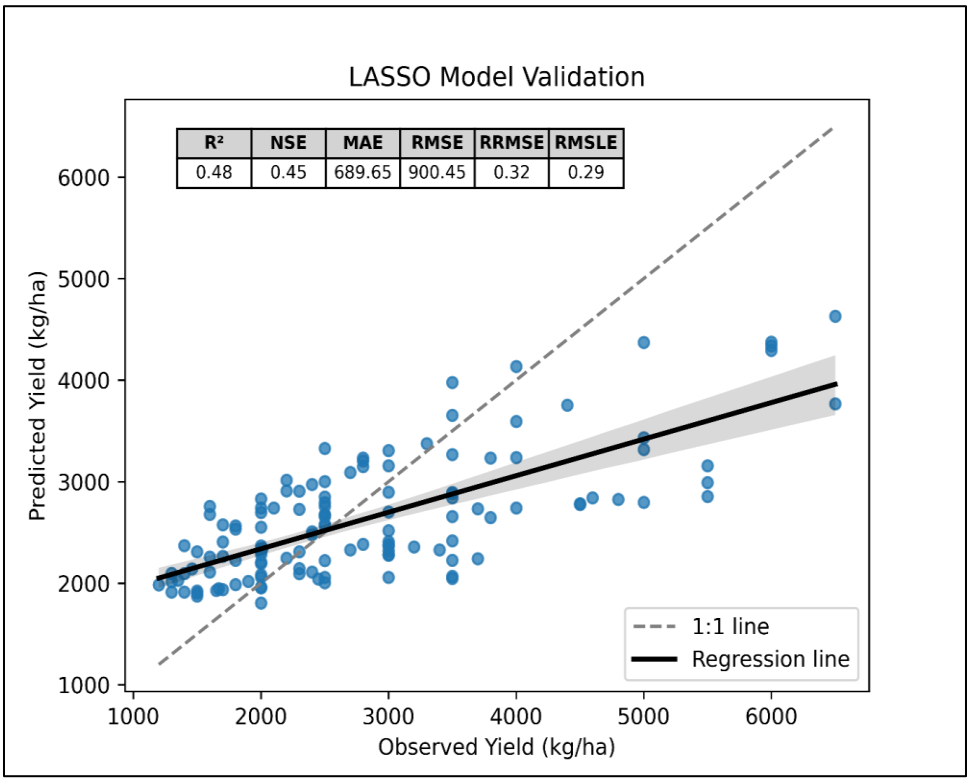

Figure 11: Result of LASSO Model Validation Phase

6.2 Machine Learning Model

6.2.1 Decision Tree Regression Model

The optimal values for DT hyperparameters, including max_depth (maximum depth of the tree) and min_samples_leaf (minimum number of yield points required to be a leaf node), were obtained using the grid search technique. The following grid was used to find the best combination of hyperparameters:

1. max_depth = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
2. min_samples_leaf = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]

Based on the training data, max_depth of 3 and min_samples_split of 8 were observed as the best values for hyperparameters.

6.2.1.1 Model Training Checking

The stability of the DTR model with the selected hyperparameters was checked by training and predicting the model on only the training subset of each of the ten folds. The model showed stability in all ten folds, with a standard deviation of 0.020 for both R^2 and NSE.

6.2.1.2 Model Verification

The selected DTR model was then trained using a 10-fold cross validation approach. The observed and predicted yield values from each fold were combined to estimate the performance of the training phase. The model verification phase showed an R^2, NSE, MAE, RMSE, RRMSE, and RMSLE of 0.33, 0.33, 627.19 kg/ha, 810.42 kg/ha, 0.32, and 0.31, respectively. Figure 12 shows the result of the DTR model verification phase.

6.2.1.3 Model Validation

In the final stage, the selected DTR model was validated using the validation data (20%), which was not exposed to the model before in any stage. The validation results of the DTR model demonstrated promising performance. The model achieved an R^2 and NSE of 0.55. The R^2 and NSE values indicate a better fit of the model to the validation data, suggesting improved predictive

accuracy. The MAE value of 630.76 kg/ha, RMSE of 817.55 kg/ha, RRMSE of 0.29, and RMSLE of 0.29 provide insights into the magnitude of the prediction errors. These metrics indicate the average and proportional differences between the predicted and observed values. Overall, the validation results demonstrate that the DTR model generalizes well to the unseen data, although it faces some challenges in accurately predicting the yield values. Figure 13 shows the result of the DTR model validation.

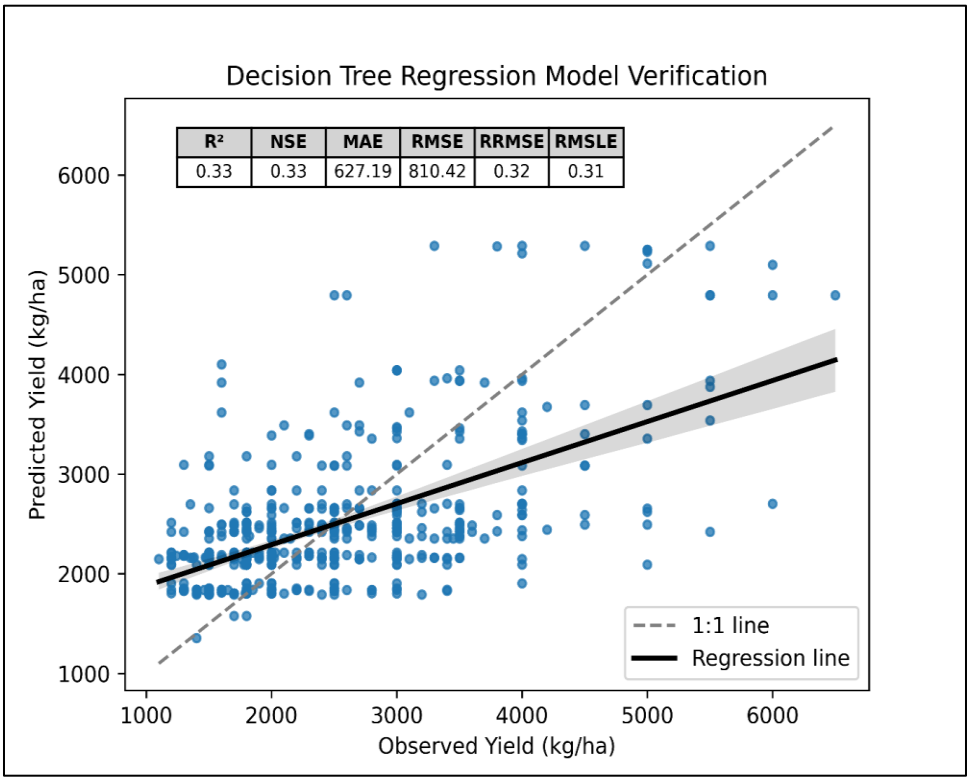

Figure 12: Result of Decision Tree Model Verification Phase

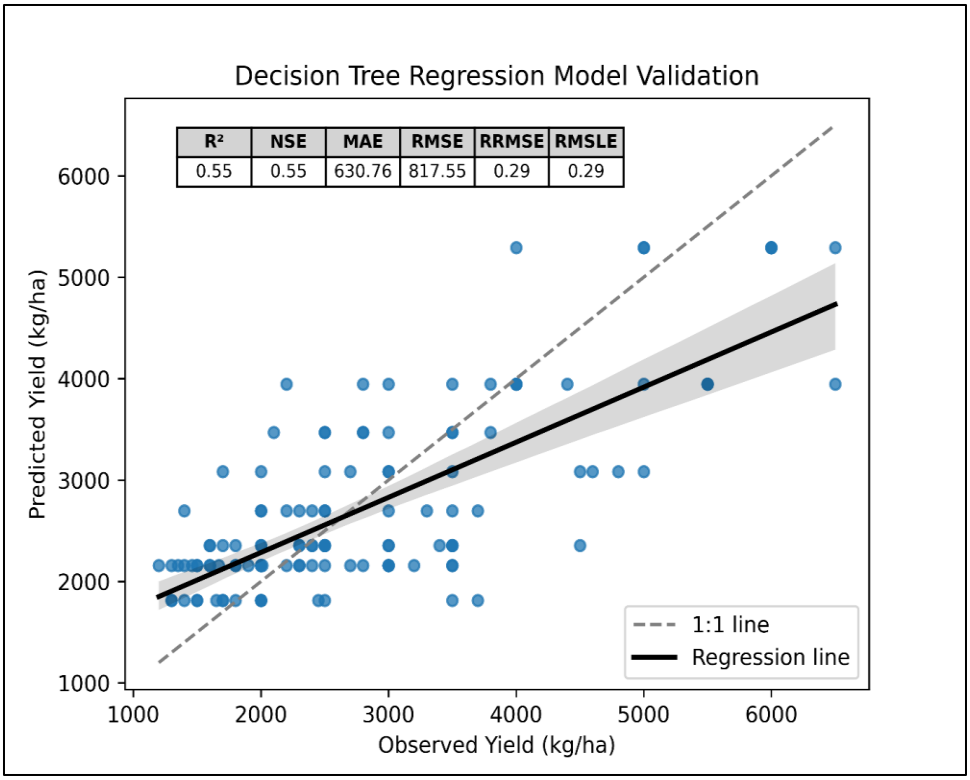

Figure 13: Result of Decision Tree Model Validation Phase

6.2.2 Random Forest Regression Model

The optimal values for RFR hyperparameters, including n_estimators (total number of trees), max_depth (maximum depth of the tree), and min_samples_leaf (minimum number of yield points required to be a leaf node), were obtained using the grid search technique. The following grid was used to find the best combination of hyperparameters:

1. n_estimators = [40, 50, 60, 70, 80, 90, 100. 110, 120, 130, 140]
2. max_depth = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
3. min_samples_leaf = [2, 3, 4, 5, 6, 7, 8, 9, 10]

Based on the training data, n_estimators = 90, max_depth = 10, and min_samples_split = 3 were identified as the best values for hyperparameters.

6.2.2.1 Model Training Checking

Similar to the DTR, the stability of the RFR model with the selected hyperparameters was checked by training and predicting the model on only the training subset of each of the ten folds. The model showed stability in all ten folds, with a standard deviation of 0.003 for both R^2 and NSE.

6.2.2.2 Model Verification

The selected RFR model was then trained using a 10-fold cross validation approach. The observed and predicted yield values from each fold were combined to estimate the performance of the training phase. The model verification phase showed an R^2, NSE, MAE, RMSE, RRMSE, and RMSLE of 0.47, 0.47, 558.77 kg/ha, 719.61 kg/ha, 0.29, and 0.29, respectively. Figure 14 shows the result of the RFR model verification phase.

6.2.2.3 Model Validation

In the final stage, the selected model was validated with validation data (20%), which was not exposed to the model before in any stage. The model achieved an R^2 of 0.64 and NSE of 0.62, indicating a strong fit of the model to the validation data and improved predictive accuracy. The RRMSE value of 0.26 and RMSLE value of 0.25 signify the proportionate and logarithmic differences between the predicted and observed values. These metrics further support the model's ability to capture the patterns and variability present in the validation dataset. The MAE value of 565.62 kg/ha and RMSE value of 744.31 kg/ha quantifies the magnitude of error in prediction. While these values indicate some deviation between the predicted and observed values, they are within an acceptable range given the complexities of crop yield prediction at the field level. Overall, the validation results show that the RFR model performed well in capturing the patterns and variability of the validation dataset, resulting in improved accuracy in predicting crop yield. Figure 15 shows the result of the RFR model validation.

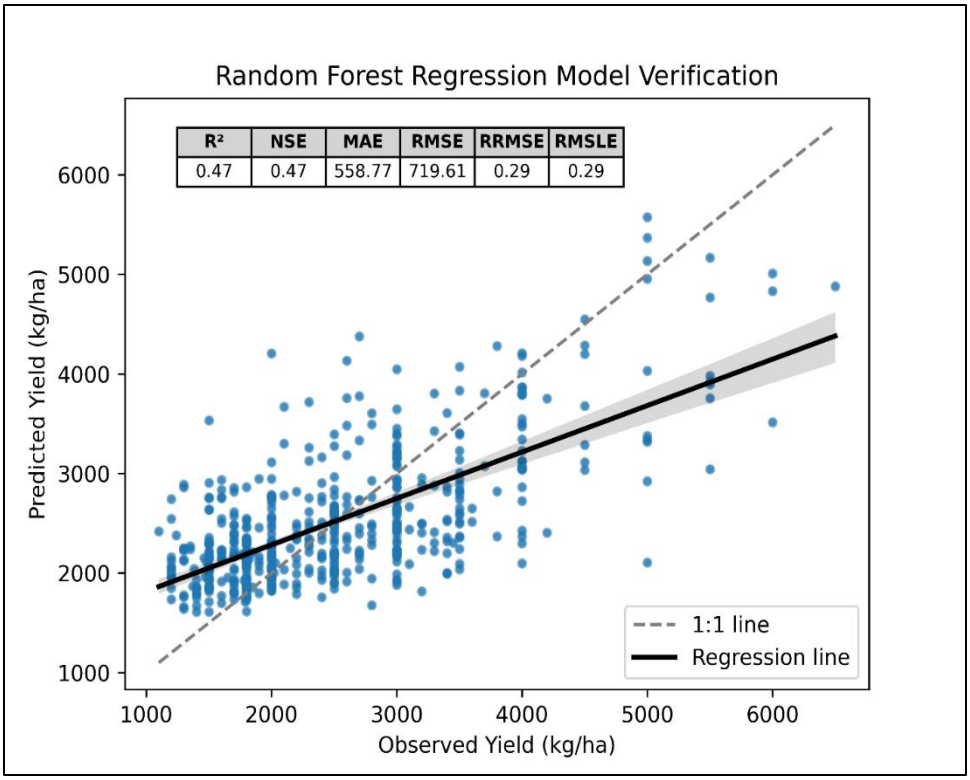

Figure 14: Result of Random Forest Model Verification Phase

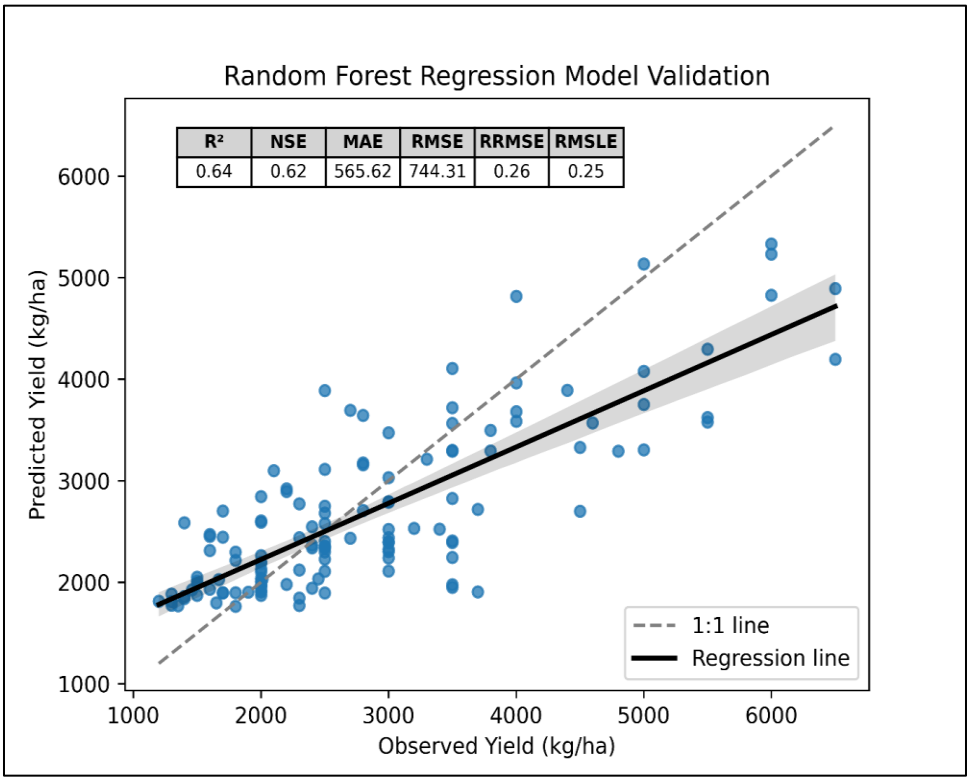

Figure 15: Result of Random Forest Model Validation Phase

6.2.3 Support Vector Regression Model

The optimal values for SVR hyperparameters including kernel (kernel type used in the model to find the hyperplane), C (regularization parameter to have a balance between the model's ability to fit the data and its generalization performance), gamma (kernel coefficient), and epsilon (margin of error or tolerance acceptable) were obtained using the grid search technique. The following grid was used to find the best combination of hyperparameters:

1. kernel = ['rbf']
2. C = [0.1, 1, 10, 100]
3. gamma = ['scale', 'auto', 0.1, 1, 10]
4. epsilon = [0.1, 0.5, 1, 5, 10]

Based on the training data, kernel = 'rbf', C = 100, gamma = 'auto', and epsilon = 0.5 were observed to be the best values for hyperparameters of the SVR model.

6.2.3.1 Model Training Checking

In the first phase, the stability of the SVR model with the selected hyperparameters was checked by training and predicting the model on only the training subset of each of the ten folds. The model showed stability in all ten folds, with a standard deviation of 0.002 for R^2 and 0.003 for NSE.

6.2.3.2 Model Verification

The selected SVR model was trained in the second phase using a 10-fold cross validation approach. The observed and predicted yield values from each fold were combined to estimate the performance of the training phase. The model verification phase showed an R^2, NSE, MAE, RMSE, RRMSE, and RMSLE of 0.36, 0.36, 600.65 kg/ha, 790.82 kg/ha, 0.32, and 0.32, respectively. Figure 16 shows the result of the MLP model verification phase.

6.2.3.3 Model Validation

In the final stage, the selected SVR model was validated with validation data (20%), which was not exposed to the model before in any stage. The validation results of the SVR model demonstrated satisfactory performance, indicating its capability to predict crop yield accurately.

The SVR model achieved an R^2 of 0.48, indicating that the model's predictions can explain 48% of the variance in the observed yield. The NSE value of 0.48 further supports the model's ability to replicate the patterns and variability present in the validation dataset. The RRMSE value of 0.31 and RMSLE value of 0.32 provide insights into the proportionate and logarithmic differences between the predicted and observed values. These metrics indicate that the SVR model successfully captures the patterns and variability in the validation dataset. The MAE of 680.32 kg/ha and RMSE of 876.22 kg/ha quantifies the magnitude of error in prediction. While the model captures the overall trend in the data well, it struggles to predict individual data points accurately, resulting in higher errors. Figure 17 shows the result of the SVR model validation phase.

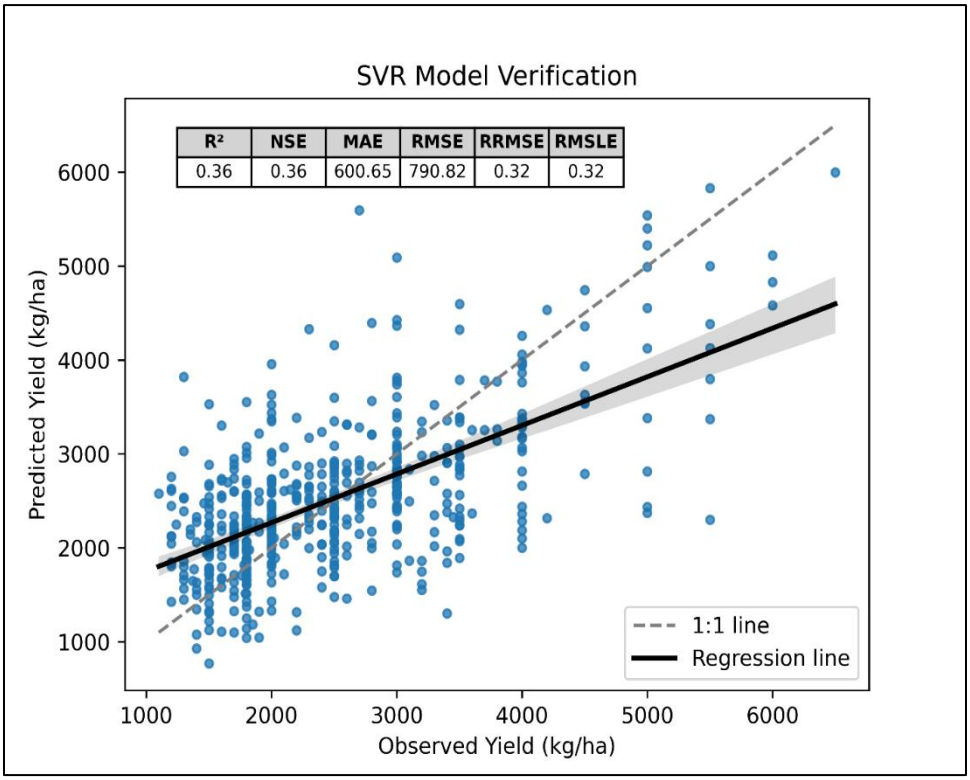

Figure 16: Result of Support Vector Regression Model Verification Phase

:

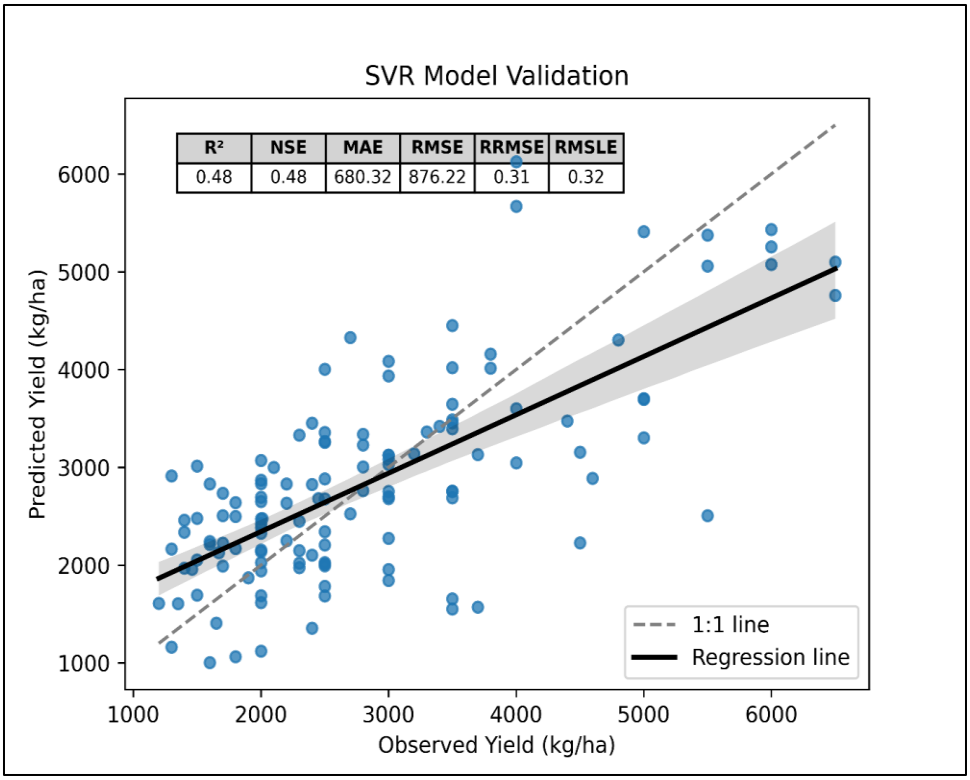

Figure 17: Result of Support Vector Regression Model Validation Phase

6.3 Deep Learning Model

6.3.1 Multilayer Perceptron Model

The optimal values for MLP hyperparameters including num_layers (total number of hidden layers), num_neurons (total number of neurons in each hidden layer), dropout_rate (randomly shut some of the neurons in the layers to avoid any bias in the learning), batch_size (number of samples fed to sample in each iteration) and epochs (number of times the entire dataset is used to train the model) were obtained using the grid search technique.

The following grid was used to find the best combination of hyperparameters:

1. num_layers = [1, 2, 3]
2. num_neurons = [32, 64, 128, 256]
3. dropout_rate = [0, 0.1, 0.2, 0.3, 0.4, 0.5]
4. batch_size = [32, 64, 128, 256]
5. epochs = [50, 100, 200]

Based on the training data, num_layers = 2, num_neurons = 128, dropout_rate = 0.3, batch_size = 128, and epochs = 200 were observed to be the best values for hyperparameters of the multilayer perceptron model. Similarly, the activation (activation function to learn nonlinear relationships) = 'relu', input_shape (number of predictor variables) = 43, optimizer (adjusts the weights to minimize the loss) = 'adam', and loss (loss function) = 'mae' were selected to build the model.

The early stopping criteria was also incorporated to avoid underfitting or overfitting with min_delta (minimum amount of change to be considered as an improvement) of 0.001 and patience (number of epochs to wait before stopping) of 20. The model architecture is shown in Figure 18.

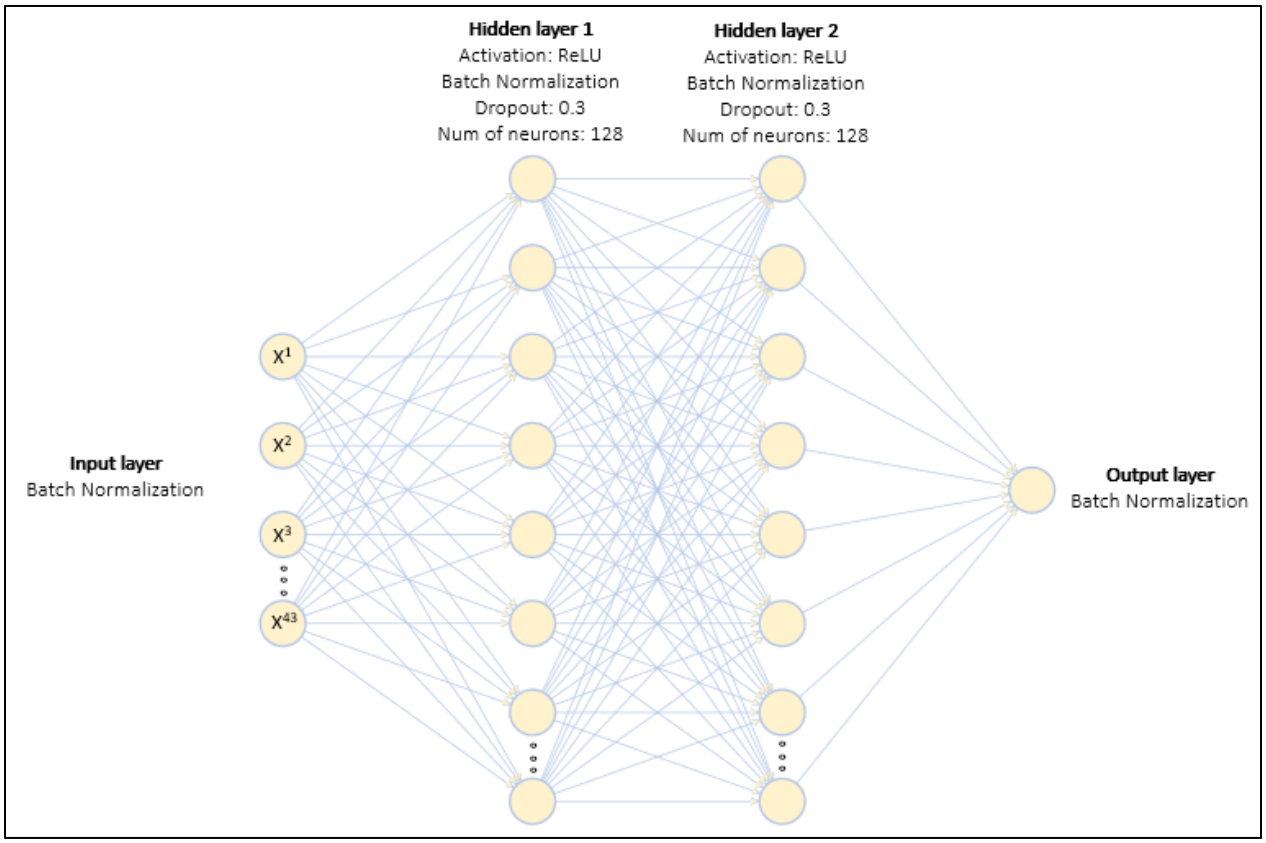

Figure 18: Model Architecture of Multilayer Perceptron Model

6.3.1.1 Model training checking

In the first phase, the stability of the MLP model with the selected hyperparameters was checked by training and predicting the model on only the training subset of each of the ten folds. The model showed stability in all ten folds, with a standard deviation of 0.010 for R^2 and 0.014 for NSE.

6.3.1.2 Model Verification

The selected MLP model was trained in the second phase using a 10-fold cross validation approach. The observed and predicted yield values from each fold were combined to estimate the performance of the training phase. The model verification phase showed an R^2, NSE, MAE, RMSE, RRMSE, and RMSLE of 0.45, 0.44, 540.72 kg/ha, 740.15 kg/ha, 0.30, and 0.43, respectively. Figure 19 shows the result of the MLP model verification phase.

6.3.1.3 Model Validation

In the final stage, the selected MLP model was validated with validation data (20%), which was not exposed to the model before in any stage. The validation results of the MLP model demonstrated promising performance, indicating its capability to predict crop yield at the field level accurately. The MLP model achieved an R^2 of 0.64, indicating that the model's predictions can explain 64% of the variance in the observed yield. This indicates a strong fit between the predicted and observed yield values. The NSE value of 0.64 further supports the model's ability to replicate the patterns and variability present in the validation dataset. The RRMSE value of 0.26 and RMSLE value of 0.25 provide insights into the proportionate and logarithmic differences between the predicted and observed values. These metrics indicate that the MLP model successfully captures the patterns and variability in the validation dataset, as the values are relatively low. The MAE of 552.65 kg/ha and RMSE of 733.53 kg/ha quantifies the magnitude of error in prediction. While these values indicate some deviation between the predicted and observed values, they are within an acceptable range considering the complexities involved in predicting the crop yield at the field level. Overall, the validation results demonstrate that the MLP model performed well in capturing the patterns and variability of the validation dataset, resulting in improved accuracy in predicting crop yield. Figure 20 shows the result of the MLP model validation.

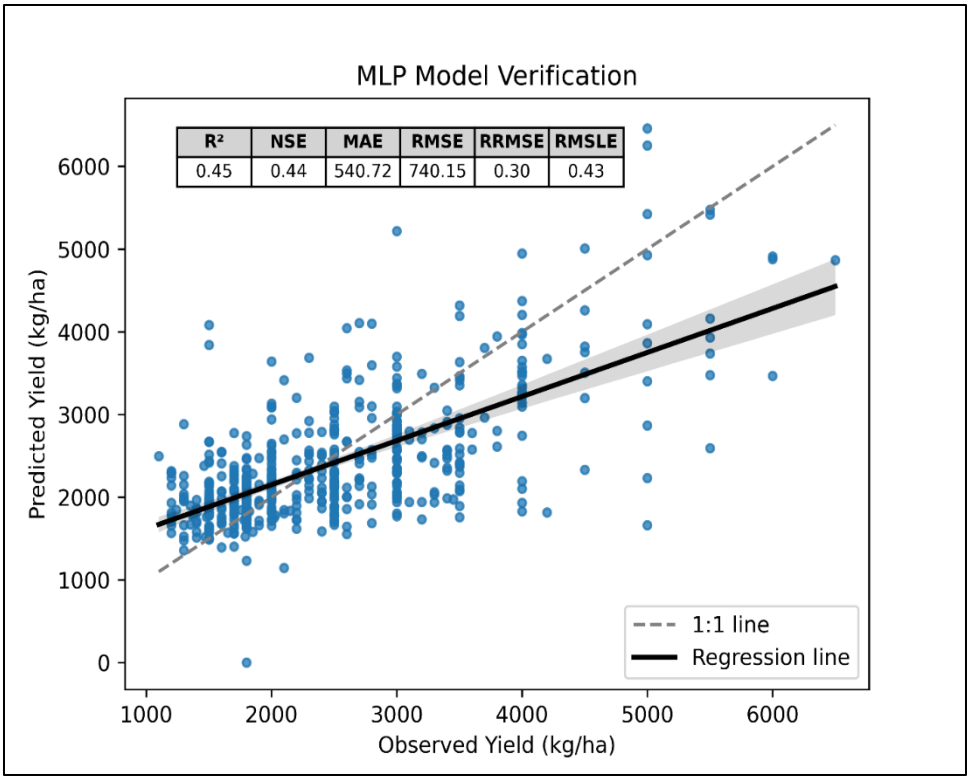

Figure 19: Result of Multilayer Perceptron Model Verification Phase

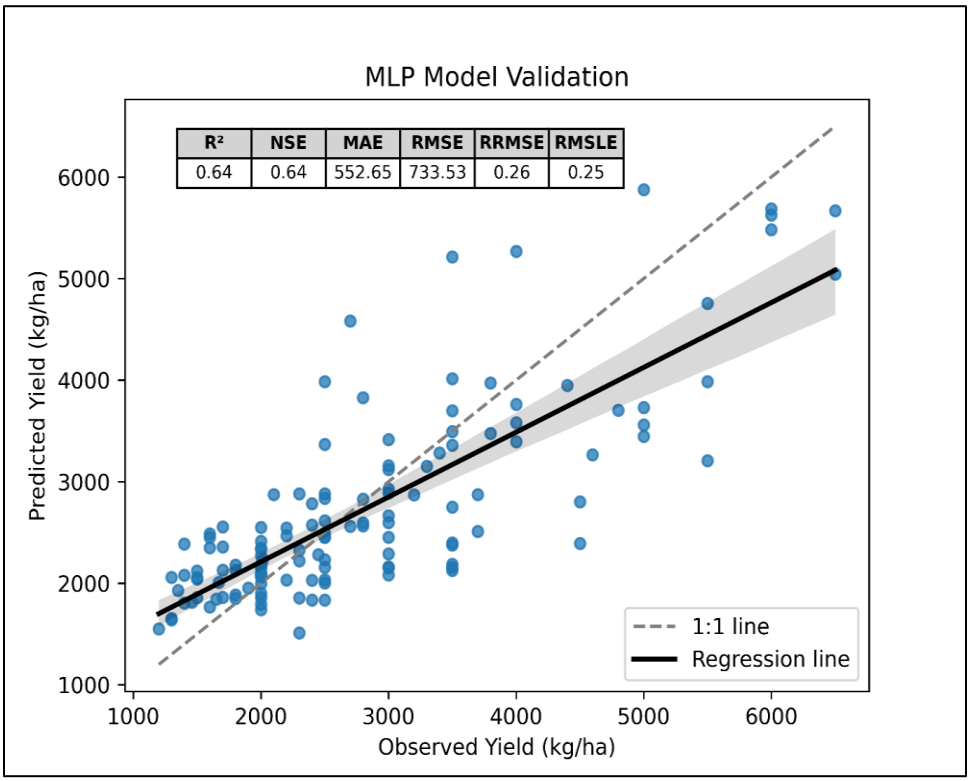

Figure 20: Result of Multilayer Perceptron Model Validation Phase

6.4 Comparison of models

One of the study's objectives was to compare the performance of different models using multiple performance metrics. The performance was compared based on the performance metrics, including R^2, NSE, RRMSE, RMSLE, MAE, and RMSE on the validation data. In a nutshell, the model with a high value of R^2 and NSE and a low value of RRMSE, RMSLE, MAE, and RMSE is considered the best model among others.

The key difference between R^2 and NSE is in their calculation and interpretation. R^2 determines the proportion of the variance in the target variable, which predictor variables can explain, while NSE evaluates the model's ability to replicate the observed variability. Using these two metrics, we assess the accuracy of predictions (NSE) and goodness of fit (R^2). In the current study, the MLP model showed the highest values of R^2 and NSE with values of 0.64 and 0.64, respectively, followed by RFR with R^2 and NSE of 0.64 and 0.62, DTR with both R^2 and NSE of 0.55, SVR with both R^2 and NSE of 0.48 and LASSO with R^2 and NSE of 0.48 and 0.45 respectively as shown in Figure 21.

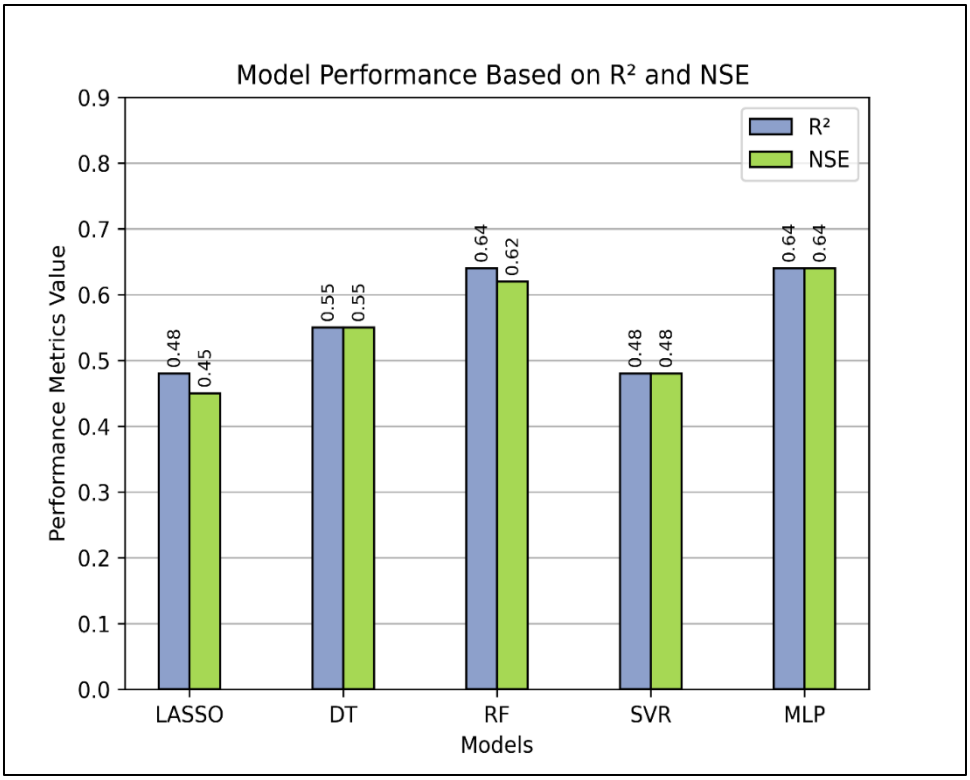

Figure 21: Model Performance Based on R^2 and NSE

Similarly, RRMSE and RMSLE capture different aspects of the model's performance. RRMSE determines the relative magnitude of the prediction error in relation to the range of the target variable, giving a normalized evaluation. RMSLE evaluates the error on a logarithmic scale, which is good for handling skewed data (such as the data in our study) and reducing the impact of large errors. In the current study, MLP and RFR models showed the lowest values of RRMSE and RMSLE with values of 0.26 and 0.25, respectively, followed by DTR with both RRMSE and RMSLE of 0.29, SVR with RRMSE and RMSLE of 0.31 and 0.32, and LASSO with RRMSE and RMSLE of 0.32 and 0.29 respectively.

In terms of MAE and RMSE, MAE assesses the average magnitude of the errors, while RMSE is more sensitive to larger errors hence useful for capturing the impact of outliers. We use both to assess the overall error magnitude and the impact of extreme errors. In this study, the MLP model outperformed the other models with the lowest MAE and RMSE of 552.65 kg/ha and 733.53 kg/ha, followed by the RFR model with MAE and RMSE of 565.62 kg/ha and 744.31 kg/ha, DTR model with MAE and RMSE of 630.76 kg/ha and 817.55, SVR model with MAE and RMSE of 680.32 kg/ha and 876.22 kg/ha and LASSO with MAE and RMSE of 689.65 kg/ha and 900.45 kg/ha respectively as shown in Figure 22.

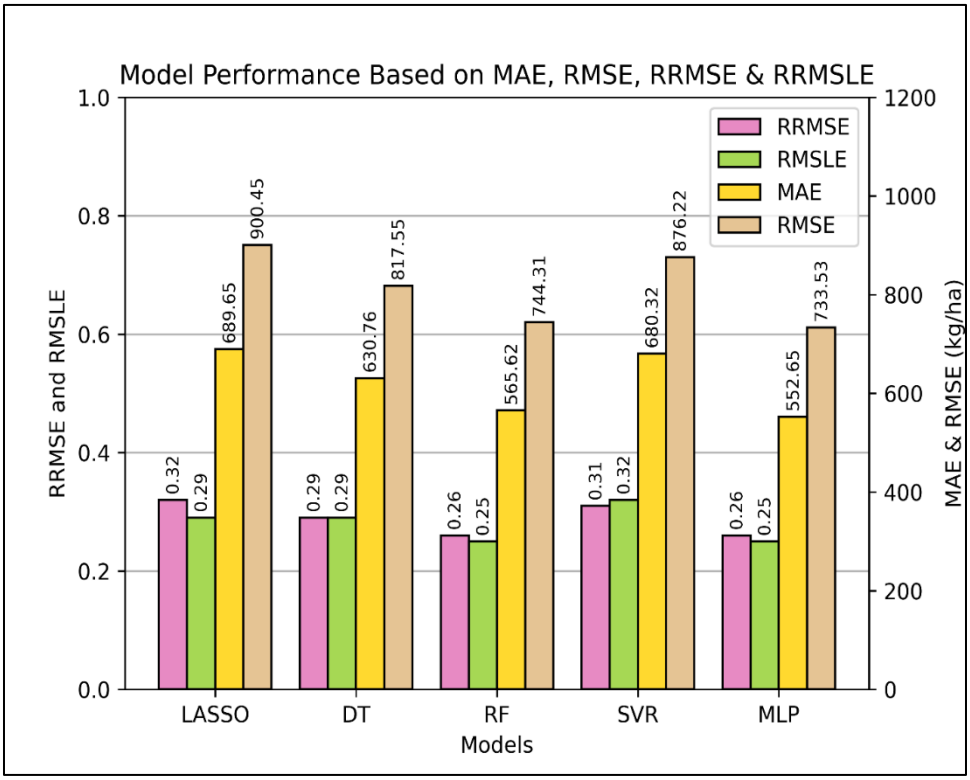

Figure 22: Model Performance Based on MAE, RMSE, RRMSE & RMSLE

6.5 Identification of the Appropriate Remotely Sensed Features

6.5.1 Individual Feature Importance

The RFR model was used to identify the most suitable features computed by the Sen4Stat system. The importance score of all the input features was calculated using the mean decrease in impurity method. The scores were normalized, and the sum of all the feature scores corresponds to 1. Figure 23 shows the feature importance in decreasing order. The three most important features were "sumlaisgint2", "laisgmax", and "sumt252". "sumlaisgint2" is the sum of the LAI interpolated with the Savitzky-Golay filter at the senescence stage. "laisgmax" is the maximum LAI interpolated with Savitzky-Golay filter. "sumt252" is the sum of temperatures higher than 25° at the senescence stage. It is estimated by taking all the temperature readings greater than 25° at the senescence stage, subtracting 25° from each, and then adding all the new values of these temperatures. The most important feature of crop yield prediction of Spain at the field level was "sumlaisgint2" with a score of 0.214, followed by "laisgmax" and "sumt252" with scores of 0.081 and 0.068, respectively. The score of "sumlaisgint2" highlighted that it had contributed 21.4% to the overall predictive power of the random forest model.

On the other hand, the three least important features were "hott2", "coldt1", and "sumt251". "hott2" represents the number of days when the temperature was higher than 35° at the senescence stage. "coldt1" is the number days when the temperature is below 0° at flowering stage. "sumt251" is the sum of temperatures higher than 25° at flowering stage and is estimated in a similar way as we estimated "sumt252," i.e., taking all the temperature readings that were greater than 25° at the flowering stage, subtracting 25° from each, and then adding all the new values of these temperatures. The least important feature on crop yield prediction of Spain at field level was "hott2" with a score of 0.001, followed by "coldt1" and "sumt251" with scores of 0.003 and 0.005, respectively. More than 60% of the predictive power of the random forest model came from the ten features, including "sumlaisgint2", "laisgmax", "sumt252", "laimax", "meant2", "meane1", "safyyield", "sumlaisgint1", "meansw32", and "meansw22" which belonged to the vegetation, weather, crop growth model and soil moisture groups.

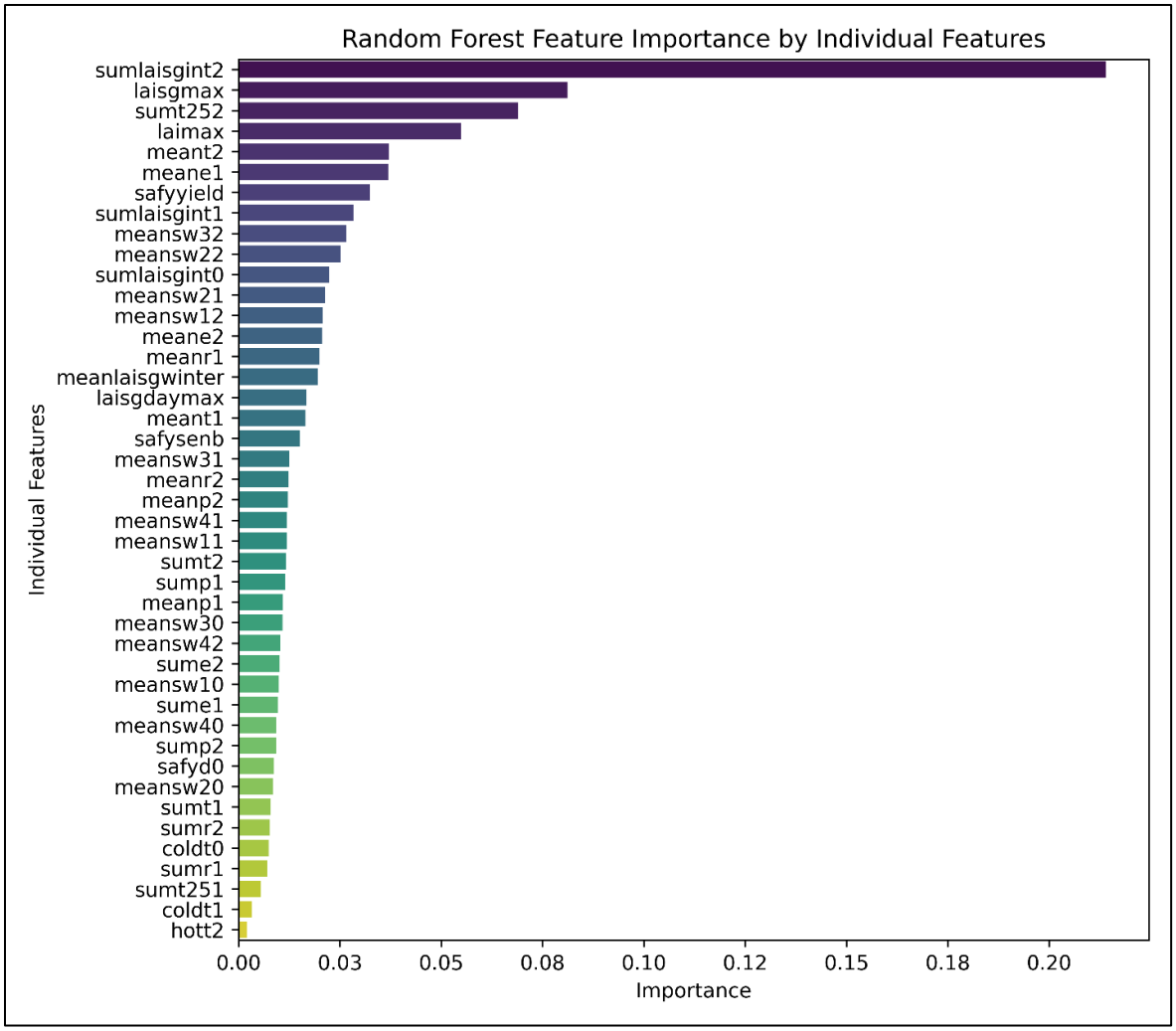

Figure 23: Random Forest Feature Importance by Individual Features

6.5.2 Feature Group Importance

To summarize the results, the group importance of features was calculated by grouping them into four categories, i.e., vegetation, weather, soil moisture, and crop growth model. The group importance score was divided by the total number of features of the respective group, as the distribution of features in each category was not uniform. The weather group had the greatest number of features, i.e., 21, followed by soil moisture with 12 features, vegetation with 7 features, and crop growth model with 3 features. The group with the highest number of features will influence the model prediction more; however, if we balance the importances by dividing the group score by the number of features in the group, we can get the relative importance of groups and highlight the most important ones. Figure 24 shows the relative importance of each group

contributing toward the overall predictive power of the random forest model. The vegetation group has the highest contribution with an importance score of 0.062, followed by the crop growth model with 0.018, weather with 0.015, and soil moisture with 0.014.

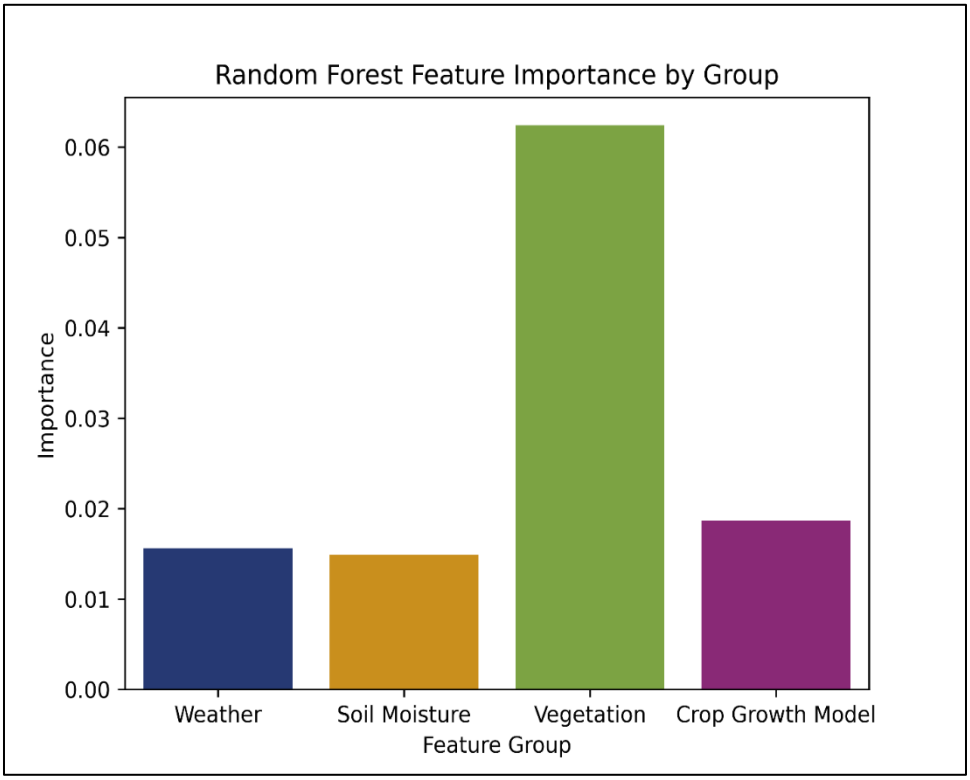

Figure 24: Random Forest Feature Importance by Group

Chapter 7: Discussion

This chapter will comprehensively analyze the results obtained in Chapter 6 and compare them with the previous research findings outlined in Chapter 2. Section 7.1 presents a detailed discussion of the results obtained from the multivariate regression model and does a comparative analysis with previous studies. In section 7.2, the results of various machine learning models are discussed and compared with prior research findings. Furthermore, section 7.3 discusses the results of the deep learning model and provides a comparative analysis with past studies. Finally, section 7.4 discusses the most significant and non-significant Sen4Stat features in estimating crop yield at the field level.

7.1 Multivariate Regression Model

As discussed in the in-situ data preparation section, out of 643 yield values in the range of 1100 kg/ha to 6500 kg/ha, 92.6% of them lies in the range from 1100 kg/ha to 4000 kg/ha, and only 7.4% falls between 4001 kg/ha to 6500 kg/ha. The skewed distribution of yield values has affected the model prediction, and models have not been able to predict with much accuracy, especially for yields in the 4001 kg/ha – 6500 kg/ha range.

The multivariate regression model LASSO over-predicted some values from 1100 kg/ha to 2500 kg/ha and under-predicted some yields ranging from 2501 kg/ha to 4000 kg/ha. However, it somewhat underpredicted all the yield values from 4001 kg/ha to 6500 kg/ha. The error magnitude in the higher range was significantly greater than in, the lower range. In addition to skewed yield distribution during the training phase, another contributing factor for this under and over estimation of yield values is the model's inability to accurately capture the non-linear relationship between the predictor variables and yield. LASSO assumes that there exists a linear relationship between the target and predictor variables; however, studies by Hao et al. (2021) and Zhang et al. (2018) showed that, in reality, the relationship between the environmental variables and crop yield is non-linear and heterogeneous.

LASSO was only able to explain 48% of the variance in the yield. Moreover, MAE and RMSE of 689.65 kg/ha and 900.45 kg/ha were significantly higher but in line with the previous research of

Cai et al. (2019) and Tang et al. (2022). Cai et al. (2019) used the LASSO model at a sub-regional scale, and the model showed an R^2 ranging from 0.4 to 0.66, whereas our LASSO model at the field level showed an R^2 of 0.48. Also, in that study, the LASSO model result was poorer than the SVM, RF, and NN, which supports the findings of our study. A similar result was also seen in the study of Tang et al. (2022), where LASSO showed an R^2 and RMSE ranging from 0.23-0.78 and 548-1023 kg/ha, and it performed poorer than the deep learning method.

7.2 Machine Learning Model

In machine learning models, the SVR showed a slight improvement in performance as compared to LASSO. From 1100 kg/ha to 4000 kg/ha model underestimated and overestimated the yield values almost equally. However, from 4001 kg/ha to 6500 kg/ha, the model mostly underestimated with a few yield points of overestimation. Like LASSO, SVR was also able to explain 48% of the variance in the yield, and the magnitude of error in the higher range was significantly greater than in, the lower range. However, the average prediction error was smaller, and the model showed an MAE and RMSE of 680.32 kg/ha and 876.22 kg/ha, respectively. The result of SVR at the field level was similar to previous studies of Khan et al. (2022), Wang et al. (2022), and Zeleke et al. (2021), in which they got an RMSE of 811 kg/ha, 725.8 kg/ha and 1295 kg/ha respectively at county, provincial and town levels. The results were also aligned with Tian et al. (2021), and Zhu et al. (2022) research which produced an R^2 of 0.41 and 0.43 at a plain and county level, and SVR showed poor performance compared to deep learning methods. At the field level, results were in line with Kayad et al. (2019) research in which the support vector showed an R^2, MAE and RMSE ranging from 0.27 – 0.55, 680 kg/ha – 1200 kg/ha, and 970 kg/ha – 1700 kg/ha respectively. Like our study, the support vector performed better than the linear regression model but was outperformed by the random forest.

On the other hand, the DTR model showed a significant improvement in performance as compared to LASSO and SVR. Similar to SVR, from 1100 kg/ha to 4000 kg/ha model underestimated and overestimated the yield values equally, however from 4001 kg/ha to 6500 kg/ha, the model did mostly the underestimation with a few yield points of overestimation, but the magnitude of error at this range was much smaller than LASSO and SVR. DT was also able to explain 55% of the variance in the yield, 7% more than the LASSO and SVR. The average prediction error was

reduced further with MAE and RMSE of 630.76 kg/ha and 817.55 kg/ha, respectively. The results of the DTR model were in line with the previous studies of Khan et al. (2022), Sharma et al. (2020), and You et al. (2017), in which they got an RMSE of 1231 kg/ha, 725 kg/ha, and 796 kg/ha respectively but at county and state scales. Moreover, the comparison with different models was also similar, as DT showed poor performance compared to RF and deep learning models.

RFR performed best among the three machine learning models used in this study. It was able to explain 64% variance in the yield, 9% more than DTR and 16% more than SVR and LASSO, similar to another study by Marshall et al. (2022) in which RF was able to explain 20% more variability in the yield as compared to LR method. The model showed a similar trend like SVR and DTR models in almost equally underestimating and overestimating the yield values from 1100 kg/ha to 4000 kg/ha. However, it mostly underestimated the yield values from 4001 kg/ha to 6500 kg/ha. The overall magnitude of underestimation and overestimation was smaller than the LASSO, DT, and SVR, which resulted in a better performance than these models. The average prediction error was lower than SVR, DT, and LASSO, with MAE and RMSE of 565.62 kg/ha and 744.31 kg/ha, respectively. The results of the wheat crop at the field level were similar to the study of Marszalek et al. (2022), who reported an R^2 of 0.68, MAE of 596 kg/ha, and RMSE of 799 kg/ha. A similar agreement was found with the research conducted by Hunt et al. (2019) at the field level, who reported an RMSE of 660 kg/ha. Additionally, Fieuzal et al. (2020) observed an R^2 ranging from 0.40 - 0.60 and RMSE ranging from 813 kg/ha to 700 kg/ha at the field level. Kayad et al. (2019) also obtained similar results for the corn crop with an R^2 and MAE ranging from 0.35 – 0.6 and 620 kg/ha – 1100 kg/ha, respectively, at the field level.

The RF model gave better results than the DT model as DT is prone to overfitting and can be sensitive to the specific data used to train them. Moreover, it can also be biased towards a majority class in an imbalanced dataset, which in our case, the range of 1100 kg/ha – 4000 kg/ha having 92.6% of the data, as compared to the range of 4001 kg/ha – 6500 kg/ha with only 7.4% of the total data points. These issues are addressed with the ensemble approach of the RF model, which uses multiple decision trees and averages their results for the final prediction, outvoting the error or poor performance of any induvial tree. This helps to reduce overfitting and improve the generalization performance of the RF model.

The underestimation done by all three machine learning models in the range of 4001 kg/ha to 6500 kg/ha is explainable as there were not enough training samples of this range, and models were not able to truly familiar with them. Also, most machine learning algorithms are designed to reduce the mean squared error or a similar metric like RMSE, which penalizes large errors more than small ones. Due to this, these models are conservative in their predictions and mostly underpredict the values of higher ranges.

Normally, results at country and regional scales are much better compared to field level due to error compensation through averaging. However, our findings show better performance than the current field level studies, and the difference between field level and country/regional level results is relatively small. This could be attributed to a larger sample size for training our models and a higher number of independent samples in the validation dataset, potentially contributing to improved accuracy.

7.3 Deep Learning Model

The deep learning model MLP outperformed all the machine learning and multivariate regression models. Like the RF model, MLP was able to explain 64% variance in the yield, 16% more than SVR and LASSO, and 9% more than DT. The model showed a similar trend to RF, DT, and SVR in almost equally underestimating and overestimating the yield values from 1100 kg/ha to 4000 kg/ha and underestimating mostly from 4001 kg/ha to 6500 kg/ha. However, the magnitude of error in prediction throughout these ranges was smaller than the machine learning and multivariate regression models, which resulted in better performance than these models. The model showed an average error in prediction MAE and RMSE of 552.65 kg/ha and 733.53 kg/ha, respectively. The results obtained were better than the findings of Fernandes et al. (2017), who achieved an R^2 of 0.43 at the municipality level. However, the obtained results at the field level were similar to the findings of Bali & Singla. (2021), who reported an RMSE of 626.13 kg/ha for wheat crop but at the district scale. The results were similar to Zeleke et al. (2021) research which obtained an R^2 of 0.69 and RMSE of 1559 kg/ha for sugarcane yield.

In general, the top performing models of our study (MLP and RF) performed better than the crop growth models used in various studies at the field level. For instance, Gilardelli et al. (2019) used

the WARM rice model at the sub-field level and achieved MAE values of 660 kg/ha and 820 kg/ha with and without the assimilation of LAI. Similarly, Silvestro et al. (2017) compared AquaCrop and SAFY models for wheat crop at the field level and got an RMSE of 1090 kg/ha for SAFY and 1120 kg/ha for AquaCrop. The Manivasagam et al. (2021) study of the SAFY model yielded an R^2 of 0.45 and RMSE of 690 kg/ha for wheat crop at the field scale. Chahbi Bellakanji et al. (2018) obtained an RMSE of 795 kg/ha by the SAFY model for the wheat crop at the field level, while Gaso et al. (2019) achieved an RMSE of 1532 kg/ha with SAFY for winter wheat yield estimation on field level.

One of the reasons our machine and deep learning models perform better is that they can understand complex interactions between vegetation, climate, and crop yield better than the simple growth models. Furthermore, incorporating an extensive set of 43 features encompassing vegetation, weather, and soil moisture (in addition to the three SAFY features) contributed to the overall improved performance compared to previous studies.

7.4 Remote Sensing Feature Importance

As shown in the results, the most important feature of RF model prediction was the sum of the LAI interpolated with the Savitzky-Golay filter at the senescence stage (sumlaisgint2), which showed an importance score of 0.214. This score means it contributed 21.4% to the overall predictive power of the RF model. This illustrates that the LAI at the senescence stage, which is from the maximum LAI point to when the LAI reaches 10% of the maximum, is an important indicator of the crop's growth and productivity during the growing season, much more than the emergence and flowering stages.

The second most important feature was the maximum LAI interpolated with the Savitzky-Golay filter (laisgmax), contributing 8.1% to the overall predictive power. It is the start of the senescence stage and doesn’t cover till the end of the growing cycle; hence its importance is not the same as that of “sumlaisgint2”. It is important to note that both top features are computed by the Savitzky-Golay filter, which shows that reducing noise from data and interpolating for the missing timeline is important.

The third most important feature was the sum of temperatures higher than 25° at the senescence stage (sumt252), contributing 6.8% to the overall prediction. This feature represents the accumulated heat stress experienced by the wheat crop during its late growth stages. Higher temperatures during growth stages can lead to heat stress and reduce grain filling. This shows that temperature is an important factor that affects the growth at the senescence stage and yield.

Maximum LAI observed (laimax) also contributed 5.4% in yield prediction. Three out top four features corresponding to LAI measurement show the significance of LAI in the yield prediction, as higher LAI generally implies higher productivity due to greater photosynthesis. The importance of LAI features in yield prediction could also be explained by two other reasons. Firstly, unlike the coarse grid of meteorological data, the LAI-derived features are computed from pixels at the field level. Therefore, the features capture the peculiarities of each field. Secondly, the great maturity of Sentinel-2 LAI retrieved for wheat in Sen4Stat combined with an efficient interpolation method. At the same time, Spain's cloud coverage is limited and provides a smooth and rather accurate LAI profile to compute the features.

Moreover, three out of the top four features represent the senescence stage, which indicates that this stage is much more important for yield prediction than the flowering and emergence stages. A comparison of the importance of the emergence, flowering, and senescence stages is shown in Figure 25, which indicates that the relative importance of the senescence stage is around 2.5 times more than the flowering and emergence stages and the sum of the importance of flowering and emergence stages is still less than the senescence importance.

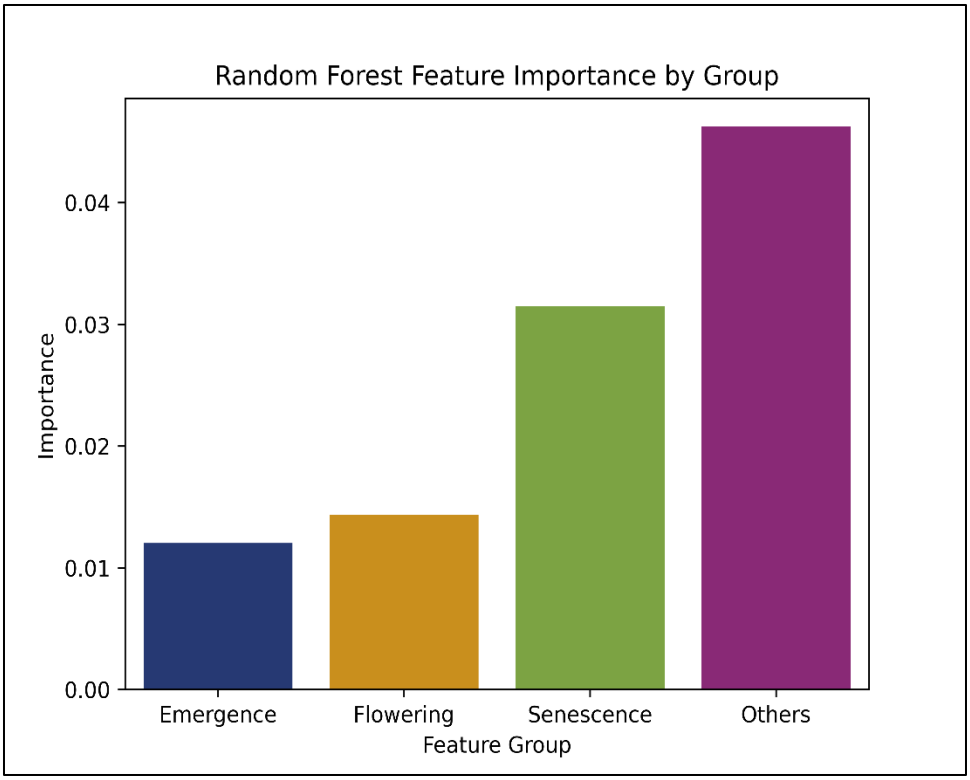

Figure 25: Random Forest Feature Importance by Phenological Stages

The reason "other" category shows the maximum importance in comparison to phenological stages features is there are four features ("safyyield", "laimax", "laisgmax", "laisgdaymax") that don't correspond to any specific growth stage. However, three out of four belong to the general vegetation category or to be precise, LAI, which we discussed, is very important for yield prediction.

Similarly, the number of days where the temperature is higher than 35° at the senescence stage (hott2), the number of days where the temperature is below 0° at the flowering stage (coldt1), and the sum of temperatures higher than 25° at flowering stage (sumt251) were the three least important features in contribution towards yield prediction as they only contributed 0.2%, 0.32%, and 0.55% respectively. There were very few days and only for very few fields when the temperature was higher than 35° at the senescence stage or higher than 25° at the flowering stage. Hence, for most of the fields, the value was 0 for these two features, so the model didn't learn anything from it as it did not provide any variability in the data. For coldt1, there were some days when the temperature was below 0°, but these days were approximately the same for all the fields.

Hence, again, the model didn’t learn any pattern from it as it did not provide any variability in the data.

As discussed in the methodology, these feature importance scores are computed based on how much the model's accuracy changes when the values of a particular feature are randomly shuffled or removed. The feature is considered important when its shuffling or removal significantly decreases the model's accuracy. In our case, there is a lot of variability in the values of the top three features (sumlaisgint2, laisgmax, sumt252), so their shuffling or removal affects the model performance a lot, and hence their score is significantly higher. However, if that feature has little effect on the model's accuracy, its importance score will be low. In our case, there is not much variability in the values of those three features (hott2, coldt1, sumt251), so their shuffling or removal didn’t affect the model performance much, and hence their score is significantly lower.

The results of group importance by different categories of features were also interesting and in line with the previous studies. The vegetation group showed the highest importance and had three times more importance than the weather, soil moisture, and crop growth model group. Moreover, the sum of the importance of weather, soil moisture, and crop growth model was still less than the vegetation group's importance.

Chapter 8: Conclusion

Crop yield estimation plays a significant role in ensuring any country's food security and sustainable agricultural practices. The accurate crop yield estimation can help properly plan and manage important agricultural resources, including land, water, and fertilizers, which leads to improved crop productivity and profitability. Furthermore, crop yield estimation can help predict the effects of climate change and extreme weather events on agriculture and hence help mitigate them. With the accurate prediction of crop yields, policymakers can make informed decisions regarding food distribution and trade, ensuring that the required amount of food reaches everyone. Thus, the study of crop yield estimation is important for achieving food security and sustainable agriculture goals.

Currently, most national statistics offices responsible for collecting crop yield data rely on a small set of measurements and then aggregate it on a regional or national level. They also use some uncertain data sources, like eye estimation methods, farmer’s surveys, etc., which are inaccurate ways of estimating the yield. Once they use this kind of uncertain data, aggregate those few samples and consider them a representation of the whole region or nation, the error increases exponentially. That’s where our research comes in; using satellite earth observation data and artificial intelligence to better estimate the crop yield at field level, preferably on all the fields of a region or a country, and then aggregate them. This will result in an accurate assessment of the crop yield of a region or nation, which will be helpful for all the stakeholders in the field of agriculture and food security.

After conducting a study on wheat yield estimation in Spain at the field level using artificial intelligence, the results show that all the machine and deep models, including DTR, RFR, SVR, and MLP, have performed better than traditional crop growth models like SAFY and AquaCrop and multivariate regression models. The study has shown that these machine and deep learning models have the potential to accurately predict crop yields at the field level, which can be beneficial for farmers, policymakers, and other stakeholders in the agriculture industry.

Among these models, the MLP and RFR have shown better results in terms of accuracy, as they have higher R^2 and NSE values and lower MAE, RMSE, RRMSE, and RMSLE values. This can

be attributed to the MLP model's ability to learn the complex non-linear relationships between the remote sensing features and crop yield and the ensemble nature of the RFR model, which reduces the risk of overfitting. On the other hand, DTR, SVR, and LASSO have also performed well but are relatively less accurate compared to the multilayer perceptron and random forest regression.

It is also important to note that the success of all these models mainly depends on the quality of the input data used in the training and validation phases. Therefore, careful selection of these input features and appropriate data preprocessing is crucial to obtain accurate results. Additionally, the model calibration and validation should always be carried out on different subsets of data to ensure the model's robustness and generalizability.

The issue of the inconsistent number of samples in different ranges can be addressed by collecting more data in the range of 4001-6500 kg/ha. Alternatively, a data augmentation technique can be used to increase the amount of data in this range artificially. This will further improve the model's ability to predict yields across all ranges accurately.

The results of important feature identification are also important and promising as currently, Sen4Stat computes 43 different remote sensing features, which requires high computational resources and time. This can be reduced by removing some of the features which are not important at all. Also, it gives the opportunity to study and add more features especially related to the senescence stage and vegetation, as it’s the most important in predicting the crop yield.

The transferability of the MLP model to another year will also be interesting, but it will mainly depend on whether the underlying patterns and relationships between the remote sensing features and crop yield remain consistent over time. If there are changes in the climate, soil, management practices, etc., then the model may need to be recalibrated to account for these changes.

Overall, the results of this study suggest that artificial intelligence has great potential in crop yield estimation, even at the field level, and can be used as an alternative to traditional approaches. Further research can be conducted to explore the potential of other machine and deep learning models and to investigate the impact of different input features on the model performance.

References

Abdul-Jabbar, T. S., Ziboon, A. T., & Albayati, M. M. (2023). Crop yield estimation using different remote sensing data: Literature review. *IOP Conference Series: Earth and Environmental Science*, *1129*(1), 012004. https://doi.org/10.1088/1755-1315/1129/1/012004

Aghighi, H., Azadbakht, M., Ashourloo, D., Shahrabi, H., & Radiom, S. (2018). Machine Learning Regression Techniques for the Silage Maize Yield Prediction Using Time-Series Images of Landsat 8 OLI. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *PP*, 1–15. https://doi.org/10.1109/JSTARS.2018.2823361

Alarcón-Segura, V., Grass, I., Breustedt, G., Rohlfs, M., & Tscharntke, T. (2022). Strip intercropping of wheat and oilseed rape enhances biodiversity and biological pest control in a conventionally managed farm scenario. *Journal of Applied Ecology*, *59*(6), 1513–1523. https://doi.org/10.1111/1365-2664.14161

Aravind, K., Vashisth, A., Krishanan, P., & Das, B. (2022). Wheat yield prediction based on weather parameters using multiple linear, neural network and penalised regression models. *Journal of Agrometeorology, 24*. https://doi.org/10.54386/jam.v24i1.1002

Bali, N., & Singla, A. (2021). Deep Learning Based Wheat Crop Yield Prediction Model in Punjab Region of North India. *Applied Artificial Intelligence*, *35*(15), 1304–1328. https://doi.org/10.1080/08839514.2021.1976091

Barriguinha, A., Jardim, B., de Castro Neto, M., & Gil, A. (2022). Using NDVI, climate data and machine learning to estimate yield in the Douro wine region. *International Journal of Applied Earth Observation and Geoinformation*, *114*, 103069. https://doi.org/10.1016/j.jag.2022.103069

Blasch, G., Spengler, D., Itzerott, S., & Wessolek, G. (2015). Organic Matter Modeling at the Landscape Scale Based on Multitemporal Soil Pattern Analysis Using RapidEye Data. *Remote Sensing*, *7*(9), Article 9. https://doi.org/10.3390/rs70911125

Cai, Y., Guan, K., Lobell, D., Potgieter, A., Wang, S., Peng, J., Xu, T., Asseng, S., Zhang, Y., You, L., & Peng, B. (2019). Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agricultural and Forest Meteorology*, *274*, 144–159. https://doi.org/10.1016/j.agrformet.2019.03.010

Chahbi Bellakanji, A., Zribi, M., Lili-Chabaane, Z., & Mougenot, B. (2018). Forecasting of Cereal Yields in a Semi-arid Area Using the Simple Algorithm for Yield Estimation (SAFY) Agro-Meteorological Model Combined with Optical SPOT/HRV Images. *Sensors*, *18*(7), Article 7. https://doi.org/10.3390/s18072138

Chen, H., Wu, W., & Liu, H.-B. (2016). Assessing the relative importance of climate variables to rice yield variation using support vector machines. *Theoretical and Applied Climatology*, *126*(1), 105–111. https://doi.org/10.1007/s00704-015-1559-y

Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, *151*, 61–69. https://doi.org/10.1016/j.compag.2018.05.012

Despotovic, M., Nedic, V., Despotovic, D., & Cvetanovic, S. (2016). Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation. *Renewable and Sustainable Energy Reviews*, *56*, 246–260. https://doi.org/10.1016/j.rser.2015.11.058

Dhillon, M. S., Dahms, T., Kübert-Flock, C., Liepa, A., Rummler, T., Arnault, J., Steffan-Dewenter, I., & Ullmann, T. (2023a). Impact of STARFM on Crop Yield Predictions:

Fusing MODIS with Landsat 5, 7, and 8 NDVIs in Bavaria Germany. *Remote Sensing*, *15*(6), Article 6. https://doi.org/10.3390/rs15061651

Dhillon, M. S., Kübert-Flock, C., Dahms, T., Rummler, T., Arnault, J., Steffan-Dewenter, I., & Ullmann, T. (2023b). Evaluation of MODIS, Landsat 8 and Sentinel-2 Data for Accurate Crop Yield Predictions: A Case Study Using STARFM NDVI in Bavaria, Germany. *Remote Sensing*, *15*(7), Article 7. https://doi.org/10.3390/rs15071830

Diskin, P. (1999). *Agricultural productivity indicators measurement guide*. Food Security and Nutrition Monitoring (IMPACT) Project. https://scholar.google.com/scholar_lookup?title=Agricultural+productivity+indicators+measurement+guide&author=Diskin%2C+Patrick.&publication_year=1999

Doraiswamy, P., Akhmedov, B., Beard, L., Stern, A., & Mueller, R. (2010). *Operational prediction of crop yields using MODIS data and products*.

Dumanski, J., & Onofrei, C. (1989). Techniques of crop yield assessment for agricultural land evaluation. *Soil Use and Management*, *5*(1), 9–15. https://doi.org/10.1111/j.1475-2743.1989.tb00754.x

Encuesta sobre Superficies y Rendimientos de Cultivos. (2019). Ministerio de Agricultura, Pesca y Alimentación.

FAO 2009. (n.d.). Retrieved April 6, 2023, from https://www.fao.org/fileadmin/templates/wsfs/docs/Issues_papers/HLEF2050_Global_Agriculture.pdf

Fermont, A., & Benson, T. (2011). Estimating yield of food crops grown by smallholder farmers: A review in the Uganda context. *International Food Policy Research Institute Discussion Paper*, *01097*.

Fernandes, J., Ebecken, N., & Esquerdo, J. (2017). Sugarcane yield prediction in Brazil using NDVI time series and neural networks ensemble. *International Journal of Remote Sensing*, *38*, 4631–4644. https://doi.org/10.1080/01431161.2017.1325531

Fieuzal, R., Bustillo, V., Collado, D., & Dedieu, G. (2020). Combined Use of Multi-Temporal Landsat-8 and Sentinel-2 Images for Wheat Yield Estimates at the Intra-Plot Spatial Scale. *Agronomy*, *10*(3), Article 3. https://doi.org/10.3390/agronomy10030327

Forkuor, G., Hounkpatin, O. K. L., Welp, G., & Thiel, M. (2017). High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models. *PLOS ONE*, *12*(1), e0170478. https://doi.org/10.1371/journal.pone.0170478

Fritz, S., See, L., Bayas, J. C. L., Waldner, F., Jacques, D., Becker-Reshef, I., Whitcraft, A., Baruth, B., Bonifacio, R., Crutchfield, J., Rembold, F., Rojas, O., Schucknecht, A., Velde, M. V. der, Verdin, J., Wu, B., Yan, N., You, L., Gilliams, S., … McCallum, I. (2019). A comparison of global agricultural monitoring systems and current gaps. *Agricultural Systems*, *168*, 258–272. https://doi.org/10.1016/j.agsy.2018.05.010

Gaso, D. V., Berger, A. G., & Ciganda, V. S. (2019). Predicting wheat grain yield and spatial variability at field scale using a simple regression or a crop model in conjunction with Landsat images. *Computers and Electronics in Agriculture*, *159*, 75–83. https://doi.org/10.1016/j.compag.2019.02.026

Geipel, J., Link, J., & Claupein, W. (2014). Combined Spectral and Spatial Modeling of Corn Yield Based on Aerial Images and Crop Surface Models Acquired with an Unmanned Aircraft System. *Remote Sensing*, *6*(11), Article 11. https://doi.org/10.3390/rs61110335

Gilardelli, C., Stella, T., Confalonieri, R., Ranghetti, L., Campos-Taberner, M., García-Haro, F. J., & Boschetti, M. (2019). Downscaling rice yield simulation at sub-field scale using remotely sensed LAI data. *European Journal of Agronomy, 103*, 108–116. https://doi.org/10.1016/j.eja.2018.12.003

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., & Toulmin, C. (2010). Food security: The challenge of feeding 9 billion people. *Science (New York, N.Y.)*, *327*(5967), 812–818. https://doi.org/10.1126/science.1185383

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Gozdowski, D., Samborski, S., & Dobers, E. (2010). Evaluation of methods for the detection of spatial outliers in the yield data of winter wheat. *Colloquium Biometricum*, *40*. https://bw.sggw.pl/info/article/WULSc264a6a72c8a46c387ad5d7f79a295b4/

Hao, X., Zhang, X., Ye, Z., Jiang, L., Qiu, X., Tian, Y., Zhu, Y., & Cao, W. (2021). Machine learning approaches can reduce environmental data requirements for regional yield potential simulation. *European Journal of Agronomy*, *129*, 126335. https://doi.org/10.1016/j.eja.2021.126335

Hunt, M. L., Blackburn, G. A., Carrasco, L., Redhead, J. W., & Rowland, C. S. (2019). High resolution wheat yield mapping using Sentinel-2. *Remote Sensing of Environment*, *233*, 111410. https://doi.org/10.1016/j.rse.2019.111410

J.a, B., L.p, P., M.d.k, O., & G.l, T. (1994). Soybean yield and pest management as influenced by nematodes, herbicides, and defoliating insects. *Agronomy Journal*.

https://scholar.google.com/scholar_lookup?title=Soybean+yield+and+pest+management+as+influenced+by+nematodes%2C+herbicides%2C+and+defoliating+insects.&author=Browde+J.A.&publication_year=1994

Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim, K.-M., Gerber, J. S., Reddy, V. R., & Kim, S.-H. (2016). Random Forests for Global and Regional Crop Yield Predictions. *PLOS ONE*, *11*(6), e0156571. https://doi.org/10.1371/journal.pone.0156571

Johnson, D. M. (2014). An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sensing of Environment*, *141*, 116–128. https://doi.org/10.1016/j.rse.2013.10.027

Johnson, M., Hsieh, W., Cannon, A., Davidson, A., & Bédard, F. (2016). Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. *Agricultural and Forest Meteorology*, *218–219*, 74–84. https://doi.org/10.1016/j.agrformet.2015.11.003

JRC MARS Bulletin April. (2019). European Commission. https://agri4cast.jrc.ec.europa.eu/bulletinsarchive

JRC MARS Bulletin August. (2019). European Commission. https://agri4cast.jrc.ec.europa.eu/bulletinsarchive

JRC MARS Bulletin February. (2019). European Commission. https://agri4cast.jrc.ec.europa.eu/bulletinsarchive

JRC MARS Bulletin January. (2019). European Commission. https://agri4cast.jrc.ec.europa.eu/bulletinsarchive

JRC MARS Bulletin July. (2019). European Commission. https://agri4cast.jrc.ec.europa.eu/bulletinsarchive

JRC MARS Bulletin June. (2019). European Commission. https://agri4cast.jrc.ec.europa.eu/bulletinsarchive

JRC MARS Bulletin March. (2019). European Commission. https://agri4cast.jrc.ec.europa.eu/bulletinsarchive

JRC MARS Bulletin May. (2019). European Commission. https://agri4cast.jrc.ec.europa.eu/bulletinsarchive

JRC MARS Bulletin November. (2019). European Commission. https://agri4cast.jrc.ec.europa.eu/bulletinsarchive

JRC MARS Bulletin September. (2019). European Commission. https://agri4cast.jrc.ec.europa.eu/bulletinsarchive

Kaggle. (n.d.). *Intro to Deep Learning*. Retrieved April 16, 2023, from https://www.kaggle.com/learn/intro-to-deep-learning

Kayad, A., Sozzi, M., Gatto, S., Marinello, F., & Pirotti, F. (2019). Monitoring Within-Field Variability of Corn Yield using Sentinel-2 and Machine Learning Techniques. *Remote Sensing*, *11*(23), Article 23. https://doi.org/10.3390/rs11232873

Khan, S. N., Li, D., & Maimaitijiang, M. (2022). A Geographically Weighted Random Forest Approach to Predict Corn Yield in the US Corn Belt. *Remote Sensing*, *14*(12), Article 12. https://doi.org/10.3390/rs14122843

Khanal, S., Fulton, J., Klopfenstein, A., Douridas, N., & Shearer, S. (2018). Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of

soil properties and corn yield. *Computers and Electronics in Agriculture*, *153*, 213–225. https://doi.org/10.1016/j.compag.2018.07.016

Krause, P., Boyle, D. P., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, *5*, 89–97. https://doi.org/10.5194/adgeo-5-89-2005

Kuwata, K., & Shibasaki, R. (2015). Estimating crop yields with deep learning and remotely sensed data. *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 858–861. https://doi.org/10.1109/IGARSS.2015.7325900

Lambert, M.-J., Blaes, X., Traoré, P. S., & Defourny, P. (2017). Estimate yield at parcel level from S2 time serie in sub-Saharan smallholder farming systems. *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, 1–7. https://doi.org/10.1109/Multi-Temp.2017.8035204

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), Article 7553. https://doi.org/10.1038/nature14539

Lobell, D. B., Thau, D., Seifert, C., Engle, E., & Little, B. (2015). A scalable satellite-based crop yield mapper. *Remote Sensing of Environment*, *164*, 324–333. https://doi.org/10.1016/j.rse.2015.04.021

Mahdi, M. D., Mrittika, N. J., Shams, M., Chowdhury, L., & Siddique, S. (2020). A Deep Gaussian Process for Forecasting Crop Yield and Time Series Analysis of Precipitation Based in Munshiganj, Bangladesh. *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 1331–1334. https://doi.org/10.1109/IGARSS39084.2020.9323423

Manivasagam, V. S., Sadeh, Y., Kaplan, G., Bonfil, D. J., & Rozenstein, O. (2021). Studying the Feasibility of Assimilating Sentinel-2 and PlanetScope Imagery into the SAFY Crop

Model to Predict Within-Field Wheat Yield. *Remote Sensing*, *13*(12), Article 12. https://doi.org/10.3390/rs13122395

Marshall, M., Belgiu, M., Boschetti, M., Pepe, M., Stein, A., & Nelson, A. (2022). Field-level crop yield estimation with PRISMA and Sentinel-2. *ISPRS Journal of Photogrammetry and Remote Sensing*, *187*, 191–210. https://doi.org/10.1016/j.isprsjprs.2022.03.008

Marszalek, M., Körner, M., & Schmidhalter, U. (2022). Prediction of multi-year winter wheat yields at the field level with satellite and climatological data. *Computers and Electronics in Agriculture*, *194*, 106777. https://doi.org/10.1016/j.compag.2022.106777

Mehdaoui, R., & Anane, M. (2020). Exploitation of the red-edge bands of Sentinel 2 to improve the estimation of durum wheat yield in Grombalia region (Northeastern Tunisia). *International Journal of Remote Sensing*, *41*, 8986–9008. https://doi.org/10.1080/01431161.2020.1797217

Mkhabela, M. S., Bullock, P., Raj, S., Wang, S., & Yang, Y. (2011). Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. *Agricultural and Forest Meteorology*, *151*(3), 385–393. https://doi.org/10.1016/j.agrformet.2010.11.012

Mulder, V. L., de Bruin, S., Schaepman, M. E., & Mayr, T. R. (2011). The use of remote sensing in soil and terrain mapping—A review. *Geoderma*, *162*(1), 1–19. https://doi.org/10.1016/j.geoderma.2010.12.018

Nuske, S., Wilshusen, K., Achar, S., Yoder, L., Narasimhan, S., & Singh, S. (2014). Automated Visual Yield Estimation in Vineyards. *Journal of Field Robotics*, *31*(5), 837–860. https://doi.org/10.1002/rob.21541

Padilla, F. L. M., Maas, S. J., González-Dugo, M., Mansilla, F., Rajan, N., Gavilán, P., & Dominguez, J. (2012). Monitoring regional wheat yield in Southern Spain using the

GRAMI model and satellite imagery. *Field Crops Research*, *130*, 145–154. https://doi.org/10.1016/j.fcr.2012.02.025

Pang, A., Chang, M. W. L., & Chen, Y. (2022). Evaluation of Random Forests (RF) for Regional and Local-Scale Wheat Yield Prediction in Southeast Australia. *Sensors*, *22*(3), Article 3. https://doi.org/10.3390/s22030717

Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R. L., & Mouazen, A. M. (2016). Wheat yield prediction using machine learning and advanced sensing techniques. *Computers and Electronics in Agriculture*, *121*, 57–65. https://doi.org/10.1016/j.compag.2015.11.018

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.

Peng, B., Guan, K., Pan, M., & Li, Y. (2018). Benefits of Seasonal Climate Prediction and Satellite Data for Forecasting U.S. Maize Yield. *Geophysical Research Letters*, *45*(18), 9662–9671. https://doi.org/10.1029/2018GL079291

Peng, Y., Xiong, X., Adhikari, K., Knadel, M., Grunwald, S., & Greve, M. H. (2015). Modeling Soil Organic Carbon at Regional Scale by Combining Multi-Spectral Images with Laboratory Spectra. *PLOS ONE*, *10*(11), e0142295. https://doi.org/10.1371/journal.pone.0142295

Peng, Y., Zhu, T., Yucui, L., Dai, C., Fang, S., Gong, Y., Wu, X., Zhu, R., & Liu, K. (2019). Remote prediction of yield based on LAI estimation in oilseed rape under different planting methods and nitrogen fertilizer applications. *Agricultural and Forest Meteorology*, *271*, 116–125. https://doi.org/10.1016/j.agrformet.2019.02.032

Prasad, A., Chai, L., Singh, R., & Kafatos, M. (2006). Crop yield estimation model for Iowa using remote sensing and surface parameters. *International Journal of Applied Earth Observation and Geoinformation*, *8*, 26–33. https://doi.org/10.1016/j.jag.2005.06.002

Prasad, P. V. V., Raí A. Schwalbert, Telmo Amado, Geomar Corassa, Luan Pierre Pott, P.V.Vara Prasad, & Ignacio A. Ciampitti. (2020). Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. *Agricultural and Forest Meteorology*, *284*, 107886-. https://doi.org/10.1016/j.agrformet.2019.107886

Pratama, M. T., Kim, S., Ozawa, S., Ohkawa, T., Chona, Y., Tsuji, H., & Murakami, N. (2020). Deep Learning-based Object Detection for Crop Monitoring in Soybean Fields. *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–7. https://doi.org/10.1109/IJCNN48605.2020.9207400

R Documentation. (n.d.). *Nash-Sutcliffe Efficiency*. Retrieved April 13, 2023, from https://search.r-project.org/CRAN/refmans/hydroGOF/html/NSE.html

Regulation (EU) No 1306/2013 of the European Parliament and of the Council on the financing, management and monitoring of the common agricultural policy and repealing Council Regulations (EEC), 347 OJ L (2013). http://data.europa.eu/eli/reg/2013/1306/oj/eng

Roell, Y. E., Beucher, A., Møller, P. G., Greve, M. B., & Greve, M. H. (2020). Comparing a Random Forest Based Prediction of Winter Wheat Yield to Historical Yield Potential. *Agronomy*, *10*(3), Article 3. https://doi.org/10.3390/agronomy10030395

Rojas, O. (2007). Operational maize yield model development and validation based on remote sensing and agro-meteorological data in Kenya. *International Journal of Remote Sensing*, *28*(17), 3775–3793. https://doi.org/10.1080/01431160601075608

Ruiter, H. R. D., Macdiarmid, J. I., Matthews, R. B., Kastner, T., Lynd, L. R., & Smith, P. (2017). Total global agricultural land footprint associated with UK food supply 1986–2011. *Global Environmental Change*, *43*, 72–81. https://doi.org/10.1016/j.gloenvcha.2017.01.007

Saad El Imanni, H., El Harti, A., & El Iysaouy, L. (2022). Wheat Yield Estimation Using Remote Sensing Indices Derived from Sentinel-2 Time Series and Google Earth Engine in a Highly Fragmented and Heterogeneous Agricultural Region. *Agronomy*, *12*(11), Article 11. https://doi.org/10.3390/agronomy12112853

Sapkota, T. B., Jat, M. L., Jat, R. K., Kapoor, P., & Stirling, C. (2016). Yield Estimation of Food and Non-food Crops in Smallholder Production Systems. In T. S. Rosenstock, M. C. Rufino, K. Butterbach-Bahl, L. Wollenberg, & M. Richards (Eds.), *Methods for Measuring Greenhouse Gas Balances and Evaluating Mitigation Options in Smallholder Agriculture* (pp. 163–174). Springer International Publishing. https://doi.org/10.1007/978-3-319-29794-1_8

Saxena, S. (2020, February 13). What's the Difference Between RMSE and RMSLE? *Analytics Vidhya*. https://medium.com/analytics-vidhya/root-mean-square-log-error-rmse-vs-rmlse-935c6cc1802a

Segarra, J., González-Torralba, J., Aranjuelo, Í., Araus, J. L., & Kefauver, S. C. (2020). Estimating Wheat Grain Yield Using Sentinel-2 Imagery and Exploring Topographic Features and Rainfall Effects on Wheat Performance in Navarre, Spain. *Remote Sensing*, *12*(14), Article 14. https://doi.org/10.3390/rs12142278

Sen4Stat. (2020). *ESA Sen4Stat Sentinels for Agricultural Statistics*.

Sentinel, H. (n.d.). *Sentinel-2 L2A*. Retrieved June 5, 2023, from https://docs.sentinel-hub.com/api/latest/data/sentinel-2-l2a/

Shammi, S. A., & Meng, Q. (2020). Use time series NDVI and EVI to develop dynamic crop growth metrics for yield modeling. *Ecological Indicators, 121*. https://doi.org/10.1016/j.ecolind.2020.107124

Sharma, S., Rai, S., & Krishnan, N. C. (2020). *Wheat Crop Yield Prediction Using Deep LSTM Model* (arXiv:2011.01498). arXiv. https://doi.org/10.48550/arXiv.2011.01498

Shastry, A., Ha, & Sanjay. (2017). *Prediction of Crop Yield Using Regression Techniques*. https://www.semanticscholar.org/paper/Prediction-of-Crop-Yield-Using-Regression-Shastry-Ha/5209efb2db7aea2c54ee1cc81d21aee51818b49f

Shi, Y., Thomasson, J. A., Murray, S. C., Pugh, N. A., Rooney, W. L., Shafian, S., Rajan, N., Rouze, G., Morgan, C. L. S., Neely, H. L., Rana, A., Bagavathiannan, M. V., Henrickson, J., Bowden, E., Valasek, J., Olsenholler, J., Bishop, M. P., Sheridan, R., Putman, E. B., … Yang, C. (2016). Unmanned Aerial Vehicles for High-Throughput Phenotyping and Agronomic Research. *PLOS ONE*, *11*(7), e0159781. https://doi.org/10.1371/journal.pone.0159781

Silvestro, P. C., Pignatti, S., Pascucci, S., Yang, H., Li, Z., Yang, G., Huang, W., & Casa, R. (2017). Estimating Wheat Yield in China at the Field and District Scale from the Assimilation of Satellite Data into the Aquacrop and Simple Algorithm for Yield (SAFY) Models. *Remote Sensing*, *9*(5), Article 5. https://doi.org/10.3390/rs9050509

Simbahan, G. C., Dobermann, A., & Ping, J. L. (2004). Screening Yield Monitor Data Improves Grain Yield Maps. *Agronomy Journal*, *96*(4), 1091–1102. https://doi.org/10.2134/agronj2004.1091

Stas, M., Orshoven, J., Dong, Q., Heremans, S., & Zhang, B. (2016). *A comparison of machine learning algorithms for regional wheat yield prediction using NDVI time series of SPOT-VGT*. 1–5. https://doi.org/10.1109/Agro-Geoinformatics.2016.7577625

Stevens, A., Nocita, M., Tóth, G., Montanarella, L., & Wesemael, B. van. (2013). Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. *PLOS ONE, 8*(6), e66409. https://doi.org/10.1371/journal.pone.0066409

Su, L., Tao, W., Sun, Y., Shan, Y., & Wang, Q. (2022). Mathematical Models of Leaf Area Index and Yield for Grapevines Grown in the Turpan Area, Xinjiang, China. *Agronomy*, *12*(5), Article 5. https://doi.org/10.3390/agronomy12050988

Sudduth, K. A., & Drummond, S. T. (2007). Yield Editor: Software for Removing Errors from Crop Yield Maps. *Agronomy Journal*, *99*(6), 1471–1482. https://doi.org/10.2134/agronj2006.0326

Sun, J., Di, L., Sun, Z., Shen, Y., & Lai, Z. (2019). County-Level Soybean Yield Prediction Using Deep CNN-LSTM Model. *Sensors*, *19*(20), Article 20. https://doi.org/10.3390/s19204363

Tang, X., Liu, H., Feng, D., Zhang, W., Chang, J., Li, L., & Yang, L. (2022). Prediction of field winter wheat yield using fewer parameters at middle growth stage by linear regression and the BP neural network method. *European Journal of Agronomy*, *141*, 126621. https://doi.org/10.1016/j.eja.2022.126621

Tian, H., Wang, P., Tansey, K., Zhang, J., Zhang, S., & Li, H. (2021). An LSTM neural network for improving wheat yield estimates by integrating remote sensing data and meteorological data in the Guanzhong Plain, PR China. *Agricultural and Forest Meteorology*, *310*, 108629. https://doi.org/10.1016/j.agrformet.2021.108629

U.S Department of Agriculture. (2023, April 8). *Europe—Crop Calendars*. IPAD International Production Assessment Division. https://ipad.fas.usda.gov/rssiws/al/crop_calendar/europe.aspx

Vannoppen, A., & Gobin, A. (2022). Estimating Yield from NDVI, Weather Data, and Soil Water Depletion for Sugar Beet and Potato in Northern Belgium. *Water*, *14*(8), Article 8. https://doi.org/10.3390/w14081188

Wan, L., Li, S., Chen, Y., He, Z., & Shi, Y. (2022). Application of Deep Learning in Land Use Classification for Soil Erosion Using Remote Sensing. *Frontiers in Earth Science*, *10*. https://doi.org/10.3389/feart.2022.849531

Wang, J., Si, H., Gao, Z., & Shi, L. (2022). Winter Wheat Yield Prediction Using an LSTM Model from MODIS LAI Products. *Agriculture*, *12*(10), Article 10. https://doi.org/10.3390/agriculture12101707

Wang, L., Zhou, X., Zhu, X., Dong, Z., & Guo, W. (2016). Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *The Crop Journal*, *4*(3), 212–219. https://doi.org/10.1016/j.cj.2016.01.008

World Bank. (2019). *World Bank Climate Change Knowledge Portal*. Climate Change Knowledge Portal. https://climateknowledgeportal.worldbank.org/

Xie, Y., Wang, P., Sun, H., Zhang, S., & Li, L. (2017). Assimilation of Leaf Area Index and Surface Soil Moisture With the CERES-Wheat Model for Winter Wheat Yield Estimation Using a Particle Filter Algorithm. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(4), 1303–1316. https://doi.org/10.1109/JSTARS.2016.2628809

Yang, C., Westbrook, J. K., Suh, C. P.-C., Martin, D. E., Hoffmann, W. C., Lan, Y., Fritz, B. K., & Goolsby, J. A. (2014). An Airborne Multispectral Imaging System Based on Two Consumer-Grade Cameras for Agricultural Remote Sensing. *Remote Sensing*, *6*(6), Article 6. https://doi.org/10.3390/rs6065257

Yang, Q., Shi, L., Han, J., Zha, Y., & Zhu, P. (2019). Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. *Field Crops Research*, *235*, 142–153. https://doi.org/10.1016/j.fcr.2019.02.022

Yao, R. J., Yang, J. S., Wu, D. H., Xie, W. P., Gao, P., & Wang, X. P. (2016). Characterizing Spatial–Temporal Changes of Soil and Crop Parameters for Precision Management in a Coastal Rainfed Agroecosystem. *Agronomy Journal*, *108*(6), 2462–2477. https://doi.org/10.2134/agronj2016.01.0004

You, J., Li, X., Low, M., Lobell, D., & Ermon, S. (2017). Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data. *Proceedings of the AAAI Conference on Artificial Intelligence*, *31*(1), Article 1. https://doi.org/10.1609/aaai.v31i1.11172

Zeleke, G., Tadesse, T., & Awoke, B. (2021). Combined Use of Landsat 8 and Sentinel 2A Imagery for Improved Sugarcane Yield Estimation in Wonji-Shoa, Ethiopia. *Journal of the Indian Society of Remote Sensing*, *50*. https://doi.org/10.1007/s12524-021-01466-8

Zhang, J., Tian, H., Wang, P., Tansey, K., Zhang, S., & Li, H. (2022). Improving wheat yield estimates using data augmentation models and remotely sensed biophysical indices within deep neural networks in the Guanzhong Plain, PR China. *Computers and Electronics in Agriculture*, *192*, 106616. https://doi.org/10.1016/j.compag.2021.106616

Zhang, P.-P., Zhou, X.-X., Wang, Z.-X., Mao, W., Li, W.-X., Yun, F., Guo, W.-S., & Tan, C.-W. (2020). Using HJ-CCD image and PLS algorithm to estimate the yield of field-grown

winter wheat. *Scientific Reports*, *10*(1), Article 1. https://doi.org/10.1038/s41598-020-62125-5

Zhang, X., Xu, H., Jiang, L., Zhao, J., Zuo, W., Qiu, X., Tian, Y., Cao, W., & Zhu, Y. (2018). Selection of Appropriate Spatial Resolution for the Meteorological Data for Regional Winter Wheat Potential Productivity Simulation in China Based on WheatGrow Model. *Agronomy*, *8*(10), Article 10. https://doi.org/10.3390/agronomy8100198

Zhao, H., Duan, S., Liu, J., Sun, L., & Reymondin, L. (2021). Evaluation of Five Deep Learning Models for Crop Type Mapping Using Sentinel-2 Time Series Images with Missing Information. *Remote Sensing*, *13*(14), Article 14. https://doi.org/10.3390/rs13142790

Zhu, Y., Wu, S., Qin, M., Fu, Z., Gao, Y., Wang, Y., & Du, Z. (2022). A deep learning crop model for adaptive yield estimation in large areas. *International Journal of Applied Earth Observation and Geoinformation*, *110*, 102828. https://doi.org/10.1016/j.jag.2022.102828

Appendix

Sr. No	Feature Name	Definition
1	coldt0	Number of days where temperature is below 0° at emergence
2	coldt1	Number of days where temperature is below 0° at flowering
3	hott2	Number of days where temperature is higher than 35° at senescence
4	Meane1	Mean of potential evapotranspiration at flowering
5	Meane2	Mean of potential evapotranspiration senescence
6	Meanp1	Mean of precipitation at flowering
7	Meanp2	Mean of precipitation at senescence
8	Meanr1	Mean of solar radiation at flowering
9	Meanr2	Mean of solar radiation at senescence
10	Meant1	Mean of temperature at flowering
11	Meant2	Mean of temperature at senescence
12	Meansw10	Mean ofvolumetric Soil water in the soil layer 1 (0-7cm) at emergence
13	Meansw11	Mean of volumetric Soil water in the soil layer 1 (0-7cm) at flowering
14	Meansw12	Mean of volumetric Soil water in the soil layer 1 (0-7cm) at senescence
15	Meansw20	Mean of volumetric Soil water in the soil layer 2 (7-28cm) at emergence
16	Meansw21	Mean of volumetric Soil water in the soil layer 2 (7-28cm) at flowering
17	Meansw22	Mean of volumetric sol water in the soil layer 2 (7-28cm) at senescence
18	Meansw30	Mean of volumetric sol water in the soil layer 3 (28-100cm) at emergence
19	Meansw31	Mean of volumetric scil water in the soil layer 3 (28-100cm) at flowering
20	Meansw32	Mean of volumetric scil water in the Soil layer 3 (28-100cm) at senescence
21	Meansw40	Mean of volumetric soil water in the soil layer 4 (100-289cm) at emergence
22	Meansw41	Mean of volumetric sol water in the Soil layer 4 (100-289cm) at flowering
23	Meansw42	Mean of volumetric soil water in the soil layer 4 (100-289cm) at senescence
24	Sumlaisgint0	Sum of the LAI interpolated with Savitzky-Golay at emergence
25	Sumlaisgint1	Sum of the LAI interpolated with Savitzky-Golay at flowering
26	Sumlaisgint2	Sum of the LAI interpolated with Savitzky-Golay at senescence
27	laimax	Maximum LAI observed
28	laisgmax	Maximum LAI interpolated with Savitzky-Golay

29	laisgdaymax	Day of year when LAI interpolated with Savitzky-Golay reach the max
30	Meanlaisgwinter	Mean of the LAI interpolated with Savitzky-Golay of the field in the year emergence
31	Sume1	Sum of potential evapotranspiration at flowering
32	Sume2	Sum of potential evapotranspiration at senescence
33	Sump1	Sum of precipitation at flowering
34	Sump2	Sum of precipitation at senescence
35	Sumr1	Sum of solar radiation at flowering
36	Sumr2	Sum of solar radiation at senescence
37	Sumt1	Sum of temperature at flowering
38	Sumt2	Sum of temperature at senescence
39	Sumt251	Sum of temperatures higher than >25° at flowering
40	Sumt252	Sum of temperatures higher than >25° at senescence
41	Safyd0	Sum of emergence parameters of safy
42	Safyyield	Yield estimate by safy
43	Safysenb	Senescence parameter of safy

Developing Wheat Crop Yield Estimation Method for Spain from Remotely Sensed Metrics Using Artificial Intelligence

Abdullah Toqeer

Agricultural production plays a crucial role in the global environment, society, and economy, especially the strategic crops directly impacting food security. An accurate crop yield estimation is important for farmers and policymakers to make informed decisions about agricultural practices, production, trade, and food security. However, the current methods used by national statistics offices are based on traditional approaches and uncertain data sources, leading to inaccurate yield estimates. To address this issue, this research aims to provide a state of the art wheat crop yield estimation method at the field level, utilizing remotely sensed features derived from satellite earth observation data and leveraging artificial intelligence techniques. The study is conducted on 849 wheat fields located in Castilla y Leon, Castilla-La-Mancha, and Comunidad Valenciana regions of Spain for 2019.

Various multivariate, machine learning, and deep learning models, including LASSO, DTR, RFR, SVR, and MLP, are used to predict crop yield at the field level. MLP outperforms the other models with an R^2, MAE, and RMSE of 0.64, 552.65 kg/ha, and 733.53 kg/ha, respectively. The RFR model followed closely, achieving an R^2, MAE, and RMSE of 0.64, 565.62 kg/ha, and 744.31 kg/ha, respectively. The DTR model achieved an R^2, MAE, and RMSE of 0.55, 630.76 kg/ha, and 817.55 kg/ha, the SVR model obtained 0.48, 680.32 kg/ha, and 876.22 kg/ha, and LASSO model achieved 0.48, 689.65 kg/ha, and 900.45 kg/ha, respectively.

Additionally, the feature importance score to identify the significant contributors to yield estimation revealed that vegetation features corresponding to LAI are the most significant. Moreover, among features recorded at different growing stages, the senescence stage emerged as the most important factor in predicting yield at the field level. Overall, the results demonstrate the potential of artificial intelligence in accurate crop yield estimation, even at the field level, and can be used as an alternative to traditional methods. Further research can be conducted to explore the potential of other machine and deep learning models and to investigate the impact of different input features on model performance.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Faculté des bioingénieurs
Croix du Sud, 2 bte L7.05.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/agro**