earth

MDPI

Article

A Python Framework for Crop Yield Estimation Using Sentinel-2 Satellite Data

Konstantinos Ntouros [1,2], Konstantinos Papatheodorou [1,*], Georgios Gkologkinas [2,3] and Vasileios Drimzakas-Papadopoulos [2]

[1] Department of Surveying and Geoinformatics Engineering, International Hellenic University, Terma Magnesias Street, 621 24 Serres, Greece; gepdouros@ihu.gr
[2] NubiGroup Geoservices & Research Private Company, Ymittou 27 Street, 544 53 Thessaloniki, Greece; aid24004@uom.edu.gr (G.G.); vasidrim@gmail.com (V.D.-P.)
[3] Department of Applied Informatics, University of Macedonia, 156 Egnatia Street, 546 36 Thessaloniki, Greece
* Correspondence: conpap@ihu.gr

Abstract: Remote sensing technologies are essential for monitoring crop development and improving agricultural management. This study investigates the automation of Sentinel-2 satellite data processing to enhance wheat growth monitoring and provide actionable insights for smallholder farmers. The objectives include (i) analyzing vegetation indices across phenological stages to refine crop growth monitoring and (ii) developing a cost-effective user-friendly web application for automated Sentinel-2 data processing. The methodology introduces the "Area Under the Curve" (AUC) of vegetation indices as an independent variable for yield forecasting. Among the indices examined (NDVI, EVI, GNDVI, LAI, and a newly developed RE-PAP), GNDVI and LAI emerged as the most reliable predictors of wheat yield. The findings highlight the importance of the Tillering to the Grain Filling stage in predictive modeling. The developed web application, integrating Python with Google Earth Engine, enables real-time automated crop monitoring, optimizing resource allocation, and supporting precision agriculture. While the approach demonstrates strong predictive capabilities, further research is needed to improve its generalizability. Expanding the dataset across diverse regions and incorporating machine learning and Natural Language Processing (NLP) could enhance automation, usability, and predictive accuracy.

Keywords: crop yield estimation; automated Sentinel-2 data processing; Python framework; Google Earth Engine

check for updates

Academic Editor: Charles Jones

Received: 31 December 2024
Revised: 24 February 2025
Accepted: 26 February 2025
Published: 6 March 2025

Citation: Ntouros, K.; Papatheodorou, K.; Gkologkinas, G.; Drimzakas-Papadopoulos, V. A Python Framework for Crop Yield Estimation Using Sentinel-2 Satellite Data. *Earth* **2025**, *6*, 15. https://doi.org/10.3390/earth6010015

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Satellite data have become an indispensable tool in crop yield monitoring, offering a range of spatial, temporal, and spectral resolutions that enhance the precision and accuracy of agriculture-related assessments. The integration of high-resolution satellite imagery, such as that from Sentinel-2, with artificial intelligence (AI) techniques has significantly advanced crop yield estimation. AI models, including machine learning and deep learning, utilize vegetation indices derived from Sentinel-2 data to predict yields for various crops with high accuracy, although there is a need for standardized methodologies to ensure consistent results across studies [1]. Similarly, the use of satellite-derived gross primary production (GPP) data, combined with dimension reduction techniques and deep learning models, has shown promise in forecasting corn yield and price changes, particularly in regions like Malawi, where predictive capabilities were notably enhanced [2]. The comparison of hyperspectral narrowbands from PRISMA and multispectral broadbands from

Sentinel-2 has revealed that these data sources can complement each other in predicting crop biomass and yield, with PRISMA's narrowbands offering slightly better performance in some cases [3]. Moreover, the fusion of Sentinel-1 and Sentinel-2 data with environmental factors has been effective in constructing dense NDVI time series, crucial for continuous crop monitoring despite challenges like cloudy weather conditions [4]. The combination of PlanetScope and Sentinel-2 imagery with environmental data has also improved wheat yield estimation, demonstrating the benefits of integrating multiple data sources for enhanced accuracy [5]. However, challenges remain, such as the need for more robust methods to ensure the reliability of satellite-derived crop information and the integration of additional biophysical variables to fully explain yield variability [6,7]. Overall, the synergy between satellite data and advanced analytical techniques holds great potential for improving agricultural management and food security, though further research is needed to address existing limitations and optimize methodologies [8,9]. Furthermore, the integration of advanced machine learning techniques with satellite data not only enhances yield predictions, but also facilitates the real-time monitoring of crop health and stress factors. For instance, recent studies have demonstrated that combining Sentinel-1's radar backscatter data with optical imagery from Sentinel-2 can significantly improve the estimation of crop biomass under varying weather conditions, providing a more resilient approach to agricultural assessments [10]. Additionally, incorporating environmental variables such as soil moisture and temperature into these models has been used to capture the complexities of yield variability caused by climatic fluctuations, ultimately leading to more accurate forecasting methods [11]. As researchers continue to refine these methodologies, there is an opportunity to develop comprehensive frameworks that leverage both remote sensing technologies and on-the-ground data collection, creating a holistic view of agricultural productivity that is essential for addressing food security challenges globally.

Vegetation indices (VIs) play a crucial role in crop monitoring and yield prediction, leveraging remote sensing technologies to provide valuable insights into crop health and productivity. Various studies have demonstrated the effectiveness of different VIs in monitoring diverse crops such as potatoes, wheat, cotton, sugarcane, and corn [12–18]. For instance, the Normalized Difference Vegetation Index (NDVI) is frequently used across multiple studies for its ability to predict yield and monitor crop growth stages effectively. In potato cultivation, the NDVI, along with the red-edge chlorophyll index and optimized soil-adjusted vegetation index, is particularly useful during the tuber initiation stage for yield prediction and growth monitoring [15]. Similarly, in wheat, NDVI combined with machine learning models like Random Forests has shown promise in predicting yield with high accuracy, especially when used at key growth stages such as jointing and Flowering [19]. For cotton, the integration of VIs with texture features derived from UAV-based RGB images has enhanced yield prediction accuracy, highlighting the potential of combining different data types for improved monitoring [17]. In sugarcane, RGB-based VIs such as GLI and MGRVI, when used with machine learning models, have provided accurate growth predictions, demonstrating the utility of low-cost UAV platforms in crop management [18]. The use of Sentinel-2 data and VIs like GNDVI has been effective in assessing the within-field variability of corn yield, with machine learning techniques further enhancing prediction accuracy [20]. Despite the advantages, challenges remain, such as the need for selecting appropriate VIs and growth stages for specific crops, and the difficulty in differentiating crops with similar spectral signatures [21]. Overall, the integration of VIs with advanced remote sensing technologies and machine learning models offers a robust framework for dynamic crop monitoring and yield estimation, aiding precision agriculture practices across various crop types [15,17–21].

Crop growth models like WOFOST (World Food Studies) [22] are pivotal in simulating crop development and predicting yields, which are essential for precision agriculture and food security. The WOFOST model, a process-based model, simulates the mechanistic processes of crop growth and yield production, and has been enhanced through various integrations and adaptations to improve its accuracy and applicability across different crops and environmental conditions. For instance, the integration of remote sensing data, such as UAV-derived vegetation indices, with the WOFOST model using the ensemble Kalman filter algorithm has significantly improved the accuracy of maize yield predictions, achieving a relative error of less than 5% when assimilating historical meteorological data [23]. Similarly, the coupling of WOFOST with other models, such as the WHEATGROW [24] and FROSTOL [25] models, has enhanced its ability to simulate winter wheat growth under low-temperature stress, improving simulation accuracy by 8.03% in frost years [26]. Sensitivity analyses using methods like EFAST [27] have further refined the model's parameters, enhancing its adaptability and accuracy in specific regions, such as the Yellow River Irrigation Area in China [28]. Additionally, the integration of deep learning techniques with WOFOST has shown promise in improving yield prediction accuracy at various growth stages, particularly during the reproductive phase of crops like corn [29]. The assimilation of high-resolution remote sensing data, such as Sentinel-2, into WOFOST has also been explored, demonstrating improved yield estimation accuracy at finer spatial resolutions [30,31]. Moreover, the development of the WOFOST-N model, which incorporates nitrogen dynamics, has expanded the model's applicability to fruit trees, such as the Korla Fragrant pear, by accurately simulating growth and nitrogen use efficiency [32]. Despite these advancements, challenges remain, such as accurately simulating agricultural stresses like lodging, which can affect yield estimates [31]. Overall, the continuous development and integration of WOFOST with other models and technologies underscore its critical role in advancing crop yield prediction and supporting agricultural decision making [33–35]. While these models offer high accuracy, they often require detailed field data, extensive parameter calibration, and computational resources, limiting their accessibility for smallholder farmers.

Developing a web application for the automated processing of Sentinel-2 data to facilitate systematic crop monitoring and to provide support to smallholder farmers necessitates the integration of advanced technologies and methodologies, as outlined in the referenced studies. The integration of machine learning and artificial intelligence can significantly enhance the ability of smallholder farmers to manage crops by providing expert advice and automatic disease detection, achieving high accuracy in identifying plant diseases and recommending fertilizers [36].

The use of Sentinel-2 data facilitates the precise mapping of cropping systems over multiple years, which is essential for understanding the spatial and temporal distribution of crops in smallholder regions [37]. This workflow, implemented on Google Earth Engine, effectively classifies complex cropping patterns and provides valuable insights into crop type changes. A cloud-based GIS system leveraging Sentinel-2 data to visualize crop type maps and calculate vegetation indices, offering a scalable solution for agricultural monitoring, has been developed [38]. The ADDPro pipeline further enhances efficiency by automating satellite data downloading and processing on AWS, delivering near-real-time vegetation health information while reducing data transfer costs through efficient compression techniques [39]. Open-source remote sensing tools present a cost-effective decision support system for precision agriculture, particularly benefiting resource-constrained settings [40]. Automated workflows for processing high-resolution satellite images have proven invaluable for tracking crop development and storing critical spectro-temporal data for smallholder agriculture [41].

The "Sentinel-2 for Agriculture" project advances global agricultural monitoring by developing open-source systems to generate relevant agricultural products, thereby enhancing market transparency and improving food production data [42]. Additionally, integrating multiscale and multitemporal systems, and combining satellite and UAV data, enables improved plant management and fosters better environmental health [43]. Finally, creating a real-time cropland mask using Sentinel-2 data is crucial for timely agricultural monitoring across diverse agro-ecological contexts, ensuring responsive and effective decision making [44].

Collectively, these studies highlight the potential of a web application powered by Sentinel-2 data and advanced processing techniques to deliver actionable insights to smallholder farmers and experts involved in the field, improving crop monitoring and promoting sustainable agricultural practices.

This study proposes an alternative methodology that automates Sentinel-2 data processing for crop monitoring through a cost-effective user-friendly web application. Unlike traditional crop growth models, this approach focuses on real-time vegetation index analysis and predictive insights tailored to smallholder farmers. By leveraging open-source tools such as Python and Google Earth Engine in order to eliminate the need for proprietary software and complex data preprocessing, it makes advanced crop monitoring more accessible and scalable.

The aim of this research was to investigate the potential for automating the processing of Sentinel-2 satellite data to enhance crop growth monitoring and provide valuable agricultural insights for smallholder farmers. The specific objectives and final outputs of the study include the following:

i. Analyzing vegetation indices across different phenological stages to generate more detailed and actionable crop growth monitoring insights.
ii. Developing a cost-effective user-centric web application for automated Sentinel-2 data processing designed to support agronomists, farmers, and other agricultural professionals by enabling systematic crop growth monitoring.

2. Materials and Methods

2.1. Operational Crop Fields Under Investigation

The operational crop fields under investigation are situated on the outskirts of the town of Ptolemaida, approximately 100 km west of Thessaloniki, within the Western Macedonia regional unit of Greece, Figure 1.

The present research focuses on nine operational wheat crop fields, Figure 2, five (5) of these fields (marked in red) were sown with durum wheat (Triticum durum) and are located to the north and south of the town of Ptolemaida, while the remaining four (4) fields (marked in yellow) were sown with soft wheat (Triticum aestivum) and are situated to the west and south of Ptolemaida.

Since this research was conducted in operational fields rather than controlled experimental conditions, key information such as crop installation dates and farmer interventions was obtained through farmer interview and ancillary data sources. However, the interviewed farmer did not maintain a detailed farm calendar. Additional data, including publicly available soil maps and meteorological records, were used to complement the information. The available information is summarized as follows:

- Crop installation: The installation of crops across different fields took place over a period of 3 to 9 days.
- Wheat cultivars:
    - Durum wheat: Svevo

- Soft wheat: Giorgione

- Farmer interventions: Agronomic practices related to fertilization, weed control, and disease management were consistent across all fields under investigation. Specifically, this includes the following:
- Fertilization:
  - Base fertilization: A winter application of fertilizer (N = 20, P = 23, K = 0) at a rate of 25 kg per stremma (1000 m^2) in December.
  - Spring fertilization: Two separate applications (N = 40, P = 0, K = 0) in early March and early April, with 15 kg per stremma (1000 m^2) per application.
- Weed control: Application of Avoxa Maxx 12/2.88 OD (Syngenta) herbicide between 15 April and 20 April.
- Fungal disease prevention: Application of Amistar Gold 12.5/12.5 SC (Syngenta) fungicide in early May.



Figure 1. Location of study fields (CRS: WGS84, EPSG: 4326).

Weather data, including temperature and precipitation during the growing season, were obtained from a nearby meteorological station located approximately 5 km from the study fields. The station, Ardassa (Coordinates: Latitude: 40.50 N, Longtitude: 21.60 E, Elevation: 625 m, CRS: WGS84; EPSG: 4326), is operated by the National Observatory

of Athens/Institute of Environmental Research and Sustainable Development and the data are public available (https://www.meteo.gr/, accessed on 24 February 2025). Table 1 presents the monthly average temperatures (mean, maximum, and minimum) and total precipitation recorded during the study period.

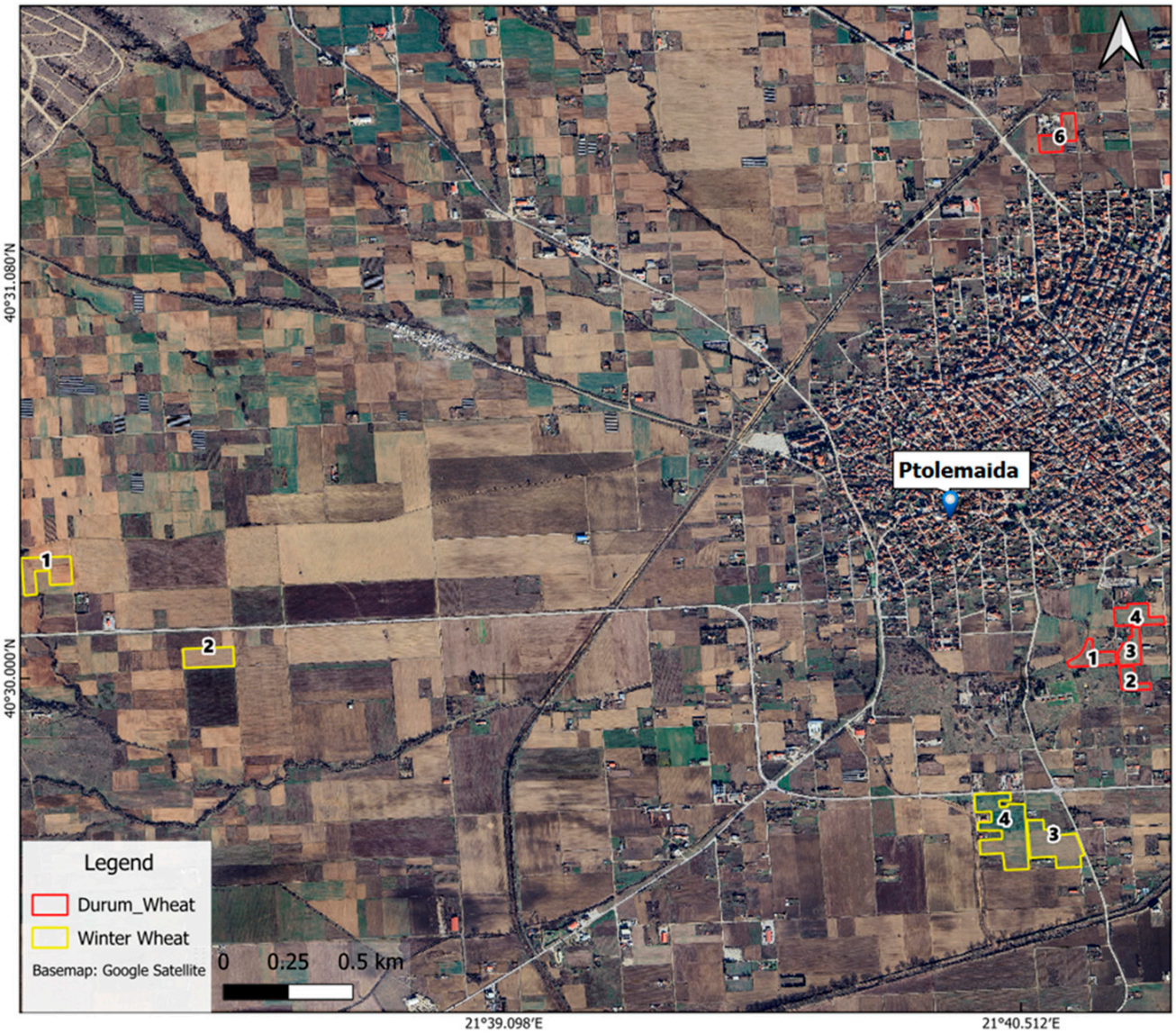


Figure 2. Location of the operational wheat crop field under investigation (CRS: WGS84, EPSG: 4326).

Table 1. Monthly weather conditions during the growing season.

Month	Mean Maximum Temperature (°C)	Mean Minimum Temperature (°C)	Mean Temperature (°C)	Total Precipitation (mm)
December 2023	11.7	−0.9	3.6	72.8
January 2024	8.8	−0.8	3.6	36.0
February 2024	13.8	2.4	7.6	25.4
March 2024	15.8	3.5	9.3	45.2
April 2024	21.7	6.7	14.3	38.6
May 2024	22.4	10.9	16.4	27.4
June 2024	32.4	16.1	24.4	12.4

Since specific soil data were not available from the farmer for each field, soil characteristics were derived from a publicly available soil map (https://iris.gov.gr/SoilServices/, accessed on 24 February 2025). Table 2 summarizes the soil properties for the investigated fields.

Table 2. Soil characteristics for each field under investigation.

Field ID	Soil Texture			Soil Type	Parent Material	Hydromorphy
	Depth 0–25 cm	Depth 25–75 cm	Depth 75–150 cm			
Soft Wheat						
FID 1 and FID 2	Moderately Fine	Medium to Moderately Fine	Moderately Fine to Fine	Luvisols	Alluvial Terraces	Well-Drained Soils
FID 3 and FID 4	Fine	Moderately Fine to Fine	Moderately Fine to Fine	Vertisols	Alluvial Deposits	Well-Drained Soils
Durum Wheat						
FID 1, FID 2, FID 3, and FID 4	Fine	Moderately Fine to Fine	Moderately Fine to Fine	Vertisols	Alluvial Deposits	Well-Drained Soils
FID 6	Moderately Fine	Coarse	Moderately Fine to Fine	Cambisols	Alluvial Terraces	Well-Drained Soils

2.2. Satellite Data

The publicly accessible Sentinel-2 MSI (Multispectral Imager) Surface Reflectance (SR), previously denoted as Bottom of Atmosphere (BOA) reflectance, datasets were selected for this investigation owing to their superior spatial resolution, regular temporal coverage, and convenient accessibility. The atmospheric bands, which encompass Coastal Aerosol, Water Vapor, and SWIR-Cirrus, were omitted from the analysis and are consequently not represented in Table 3. All datasets were procured at no financial cost from the Copernicus Dataspace (formerly known as the Copernicus Open Access Hub) [45].

Table 3. Sentinel-2 MSI bands.

Sentinel-2 Bands	Spectral Region	Central Wavelength (μm)	Resolution (m)
Band 2	Blue	0.490	10
Band 3	Green	0.560	10
Band 4	Red	0.665	10
Band 5	Red Edge 1	0.705	20
Band 6	Red Edge 2	0.740	20
Band 7	Red Edge 3	0.783	20
Band 8	NIR	0.842	10
Band 8A	Vegetation Red Edge	0.865	20
Band 11	SWIR 1	1.610	20
Band 12	SWIR 2	2.190	20

The acquisition dates of the images used in this study are presented in Table 4, along with the corresponding phenological stages. The phenological stages are defined based on the BBCH (Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie) scale. This scale categorizes the plant's developmental cycle into 10 distinct principal stages [46].

Table 4. Wheat phenological stages and the corresponding imagery acquisition date.

Phenological Stage	Days per Stage	BBCH Scale	Imagery Acquisition Date
Emergence	7–14 days	BBCH 10–19	13 January 2024
Tillering	25–40 days	BBCH 20–29	26 January 2024 15 February 2024 17 February 2024 22 February 2024
Stem Elongation	20–30 days	BBCH 30–39	3 March 2024 21 March 2024
Booting	20–30 days	BBCH 40–49	28 March 2024
Heading	7–14 days	BBCH 51–59	15 April 2024
Flowering	5–10 days	BBCH 61–69	7 May 2024
Grain Filling	10–20 days	BBCH 71–77	22 May 2024 27 May 2024
Maturity and Harvest	10–20 days	BBCH 83–99	9 June 2024
Post Harvest			24 June 2024

2.3. Yield Data

Since this research was conducted on operational crop fields, actual yield data for each wheat field were provided by the farmer. The data were collected using yield mapping equipment, specifically, the FarmTRX™ system mounted on combine machinery. The data were presented in vector point format (ESRI shapefile) and underwent pre-processing to remove any inaccurate or erroneous value entries, including extreme high and low yield values.

The growing season for this dataset extends from December 2023 to June 2024. Descriptive statistics for each wheat field are summarized in Table 5, while the spatial distribution, and interpolated yield value derived from the Inverse Distance Weighted method (IDW), of yield in each wheat crop field is illustrated in Figure 3.

Table 5. Descriptive statistics of the measured yield production per operational field and crop.

Field ID	Mean (Kg/stremma *)	Standard Deviation	Minimum (Kg/stremma *)	Maximum (Kg/stremma *)	Range (Kg/stremma *)
Soft Wheat					
FID 1	373.54	127.00	100.28	737.88	637.60
FID 2	598.44	94.25	351.06	841.57	490.51
FID 3	635.82	86.61	350.37	898.26	547.89
FID 4	605.32	115.08	150.12	649.91	499.79
Durum Wheat					
FID 1	425.59	105.98	150.12	649.91	499.79
FID 2	290.09	73.16	119.88	534.89	415.01
FID 3	258.73	82.42	61.04	542.13	481.09
FID 4	367.79	86.24	200.47	645.8	445.33
FID 6	339.34	63.99	200.6	561.36	360.76

* 1 stremma = 1000 m^2.

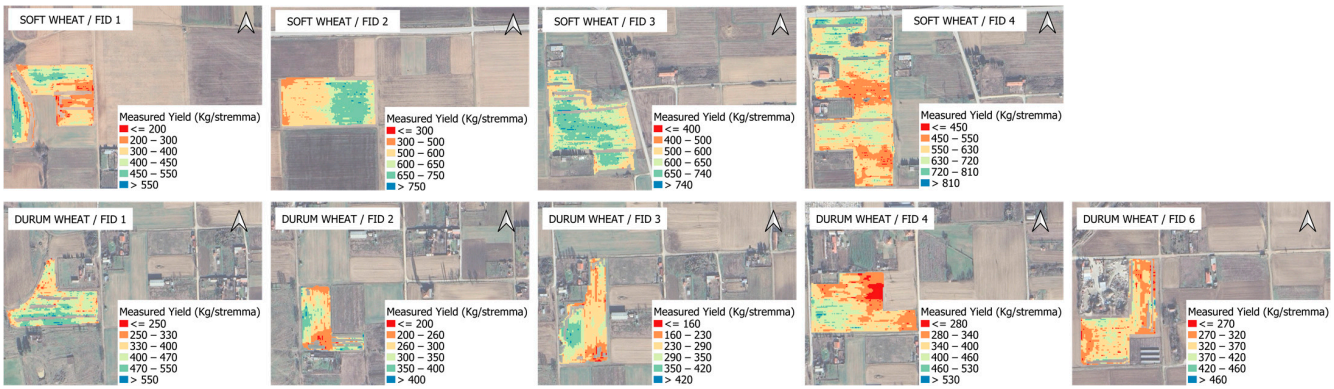


Figure 3. Spatial distribution of measured yield per field: soft wheat (**Top**) and durum wheat (**Bottom**).

2.4. Software

This study exclusively employed noncommercial and open-source software and programming frameworks. ESA SNAP software (version 10.0.0) [47] was utilized for Sentinel-2 MSI data preprocessing, while Quantum GIS (version 3.34.5) [48] facilitated the manipulation of vector data, including yield data.

To enhance workflow efficiency and automation, Python (version 3.11.1) [49] was employed within a Jupyter Notebook (version 7.0.6) environment [50] for data analysis, including regression analysis.

The development of the web application utilized Google Earth Engine (GEE) [51] for data access and analysis, implemented using the Streamlit (version 1.24.0) [52] open-source application framework.

2.5. Methods

2.5.1. Satellite Data Pre-Processing

To ensure a consistent spatial resolution of 10 m across all bands, spatial resampling of the satellite data was performed using the S2 Resampling Processor tool in the open-source ESA SNAP software. Subsequently, cloud masking was applied to each image using the 'Land/Sea Mask' tool by selecting the 'Use Vector as Mask' option and incorporating the 'Opaque_clouds' and 'Cirrus_clouds' masks provided by ESA. Finally, a subset of each dataset was created for the study area, ensuring that the area of interest, including the fields, remained free of cloud cover.

2.5.2. Satellite Data Processing

All vegetation indices (NDVI, GNDVI, EVI, LAI, RE-PAP) used in this study were calculated in the ESA SNAP software using the 'Band Maths' tool. Following the computation and extraction of the new data via the 'Bands Extractor' processor, the results were imported into QGIS. In QGIS, the mean index values for each date were extracted at the field level for each field using the Zonal Statistics tool.

2.5.3. Yield Estimation

Yield estimation was conducted using regression analysis to assess the relationship between the 'Area Under the Curve' (AUC) of vegetation index plots (mean index value) and the corresponding mean yield values across all fields. The AUC was calculated over the entire growing season as well as within a specific time period (3 March 2024 to 27 May 2024), corresponding to key phenological stages from Tillering to Stem Elongation and continuing through Grain Filling. The AUC, $\int_{t_1}^{t_2} f(t) \cdot t$, was computed using the trapezoidal rule,

implemented in Python. The analysis was conducted independently for durum and soft wheat crop fields.

For this study, five (5) distinct vegetation indices were examined, as summarized in Table 6. These include the Normalized Difference Vegetation Index (NDVI) [53], Green Normalized Difference Vegetation Index (GNDVI) [54], Enhanced Vegetation Index (EVI) [55], and the Leaf Area Index (LAI) [56], which was derived from EVI [57]. Additionally, a novel vegetation index, RE-PAP, was developed utilizing the Red Edge (RE) bands of Sentinel-2. This new index is based on findings from the literature that highlight the value of the RE region in improving the performance of vegetation indices (VIs) [58]. The RE region is particularly effective in capturing variations in chlorophyll content. Healthy plants, which have higher chlorophyll levels, exhibit a red-edge shift toward longer wavelengths. Conversely, unhealthy plants with lower chlorophyll content display a red-edge shift toward shorter wavelengths [59]. The formulation of this index is grounded in the following rationale. The subtraction term (RE3 − RE1) effectively captures the physiological state of vegetation by isolating red-edge shifts, which serve as key indicators of chlorophyll content. The division by RE2 (B6) accounts for variations in light penetration and scattering influenced by the canopy structure, thereby functioning as a reference band to normalize reflectance variations and ensure that the index output accurately reflects structural and physiological changes in vegetation. Finally, the multiplication by the ratio (NIR/RED) enhances the sensitivity of the index to vegetation vigor by leveraging the pronounced contrast between the high near-infrared (NIR) reflectance and low red reflectance observed in healthy plants.

Table 6. Sentinel-2 vegetation indices under investigation.

Index	Formula	
NDVI	$\frac{NIR-RED}{NIR+RED}$	(1)
GNDVI	$\frac{NIR-GREEN}{NIR+GREEN}$	(2)
EVI	$2.5\frac{NIR-RED}{(NIR+6RED-7.5BLUE)+1}$	(3)
LAI	$3.618 * EVI - 0.118$	(4)
RE-PAP	$\frac{Vegetation\ RE3(B7)-Vegetation\ RE1(B5)}{Vegetation\ RE2(B6)} * \frac{NIR\ (B8)}{RED(B4)}$	(5)

2.5.4. Web Application Development

The web application was implemented using the Python programming language (version 3.11.1) along with Python libraries (geemap, folium, fastkml, shapely, jsonlib, os-sys, pandas, numpy, plotly, Python-IO, tempfile, zipfile, osgeo) in combination with Google Earth Engine. To further enhance user interaction and accessibility, the Streamlit framework (version 1.24.0) was exploited. The graphical user interface (GUI) was designed to be user-friendly and accessible to users with limited expertise in remote sensing. The interface adopts a two-panel layout: the left panel allows users to select and configure task parameters, while the right panel displays the results of these tasks in an interactive map format, complemented by relevant informational notifications, Figure 4. This design emphasizes usability and fosters efficient intuitive interaction with the application.

Table 7 provides a summary of the parameterization of the available tasks as presented in the left panel of the graphical user interface (GUI).

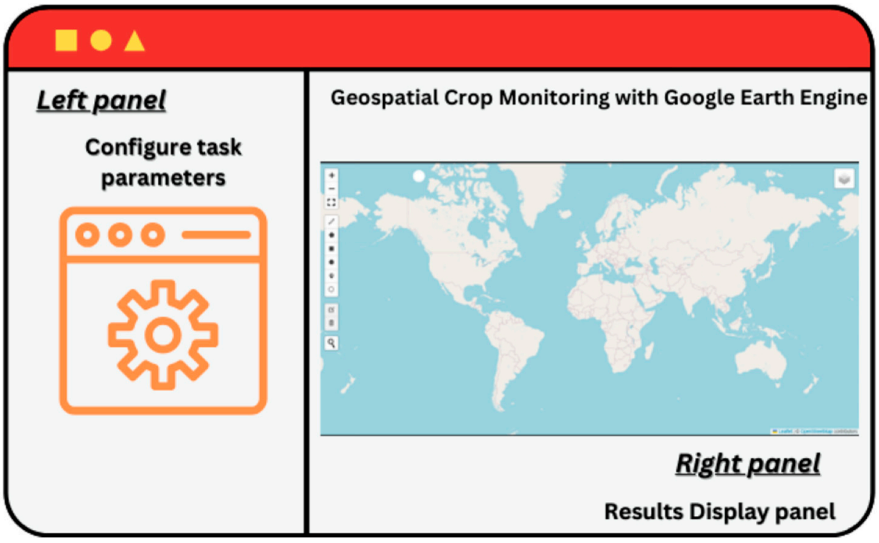


Figure 4. User interface of the app.

Table 7. Configuration and parameterization of tasks within the application.

Configure Task Parameters	Type
Search available Sentinel 2 imagery	
Start date/end date	Date picker
Cloud cover	Slider
Shadow Threshold	Slider—optional
Apply cloud mask	Check box—optional
Dates upload	Upload txt file with desired dates for acquisition—optional
Upload Region of Interest (ROI)	Upload GeoJSON or KML file
Calculate vegetation index	
Select the available images	List box
Select index to calculate	Drop—down list
Display image	Button—optional
Generate histogram	Button—optional
Export images	Button—optional
Index statistics per parcel	
Upload parcels	Upload GeoJSON or KML file
Calculate mean index value for each parcel	Button
Download values to CSV	Button—optional
Plot results	Button—optional
Dates subset for AUC calculation	
Subset start date/end date	Drop—down list/optional
Plot subset results	Button—optional
Yield prediction	
Yield prediction	Button
Download yield prediction results to CSV	Button—optional
Update parcel information (on the app)	Button—optional
Download updated parcels to GeoJSON or KML	Button—optional
Shadow/Cloud mask check	
Show shadow mask layer	Check box—optional
Plot histogram (Original/Masked)	Button—optional
Download metadata to TXT	Button—optional

Appendix A presents a comprehensive summary of the application components, including the corresponding Streamlit functions, in the form of Streamlit pseudo-code.

3. Results

3.1. Yield Estimation

In this study, we investigated the relationship between the Area Under the Curve (AUC) of mean values of Vis (NDVI; GNDVI; EVI; LAI; and RE-PAP) during the growing season as well as within a specific time period (3 March 2024 to 27 May 2024), corresponding to key phenological stages from Tillering to Stem Elongation and continuing through Grain Filling, and the measured yield at the field scale. AUC values were computed using the trapezoidal rule, complemented by regression analysis conducted within a Python environment. The predicted yield (Ypred) is expressed in Kg/stremma, where stremma is an area measurement unit corresponding to 1000 square meters, which is officially used in Greece.

The results are presented in the following tables: Table 8 for durum wheat and Table 9 for soft wheat. The regression plots are illustrated in the subsequent figures, Figures 5 and 6, corresponding to durum wheat and soft wheat, respectively.

Table 8. Regression analysis results: durum wheat.

Vegetation Index	Time Period	Reg. Equation	R^2	MAPE (%)	RMSE
EVI	Growing season	Ypred = 59.852 × AUC	0.97	16.8	59.36
EVI	Subset of growing season (from Tillering to Stem Elongation through Grain Filling)	Ypred = 85.849 × AUC	0.98	15.7	54.26
GNDVI	Growing season	Ypred = 127.889 × AUC	0.98	13.4	50.85
GNDVI	Subset of growing season (from Tillering to Stem Elongation through Grain Filling)	Ypred = 243.181 × AUC	0.99	11.9	41.35
LAI	Growing season	Ypred = 17.876 × AUC	0.97	16.9	60.03
LAI	Subset of growing season (from Tillering to Stem Elongation through Grain Filling)	Ypred = 24.962 × AUC	0.97	15.7	54.61
NDVI	Growing season	Ypred = 60.643 × AUC	0.98	14.5	53.28
NDVI	Subset of growing season (from Tillering to Stem Elongation through Grain Filling)	Ypred = 91.409 × AUC	0.98	13.5	47.46
RE-PAP	Growing season	Ypred = 10.105 × AUC	0.97	15.1	59.78
RE-PAP	Subset of growing season (from Tillering to Stem Elongation through Grain Filling)	Ypred = 12.798 × AUC	0.97	13.4	59.34

Table 9. Regression analysis results: soft wheat.

Vegetation Index	Time Period	Reg. Equation	R^2	MAPE (%)	RMSE
EVI	Growing season	Ypred = 105.827 × AUC	0.98	14.3	79.91
EVI	Subset of growing season (from Tillering to Stem Elongation through Grain Filling)	Ypred = 160.186 × AUC	0.99	10.8	61.54

Table 9. *Cont.*

Vegetation Index	Time Period	Reg. Equation	R^2	MAPE (%)	RMSE
GNDVI	Growing season	Ypred = 116.035 × AUC	0.97	18.8	102.22
GNDVI	Subset of growing season (from Tillering to Stem Elongation through Grain Filling)	Ypred = 169.039 × AUC	0.98	14.7	81.17
LAI	Growing season	Ypred = 31.814 × AUC	0.98	13.9	77.83
LAI	Subset of growing season (from Tillering to Stem Elongation through Grain Filling)	Ypred = 46.913 × AUC	0.99	10.3	59.13
NDVI	Growing season	Ypred = 98.754 × AUC	0.98	14.3	79.24
NDVI	Subset of growing season (from Tillering to Stem Elongation through Grain Filling)	Ypred = 154.527 × AUC	0.99	10.9	62.16
RE-PAP	Growing season	Ypred = 17.233 × AUC	0.98	11.1	78.51
RE-PAP	Subset of growing season (from Tillering to Stem Elongation through Grain Filling)	Ypred = 21.972 × AUC	0.98	11.8	86.21

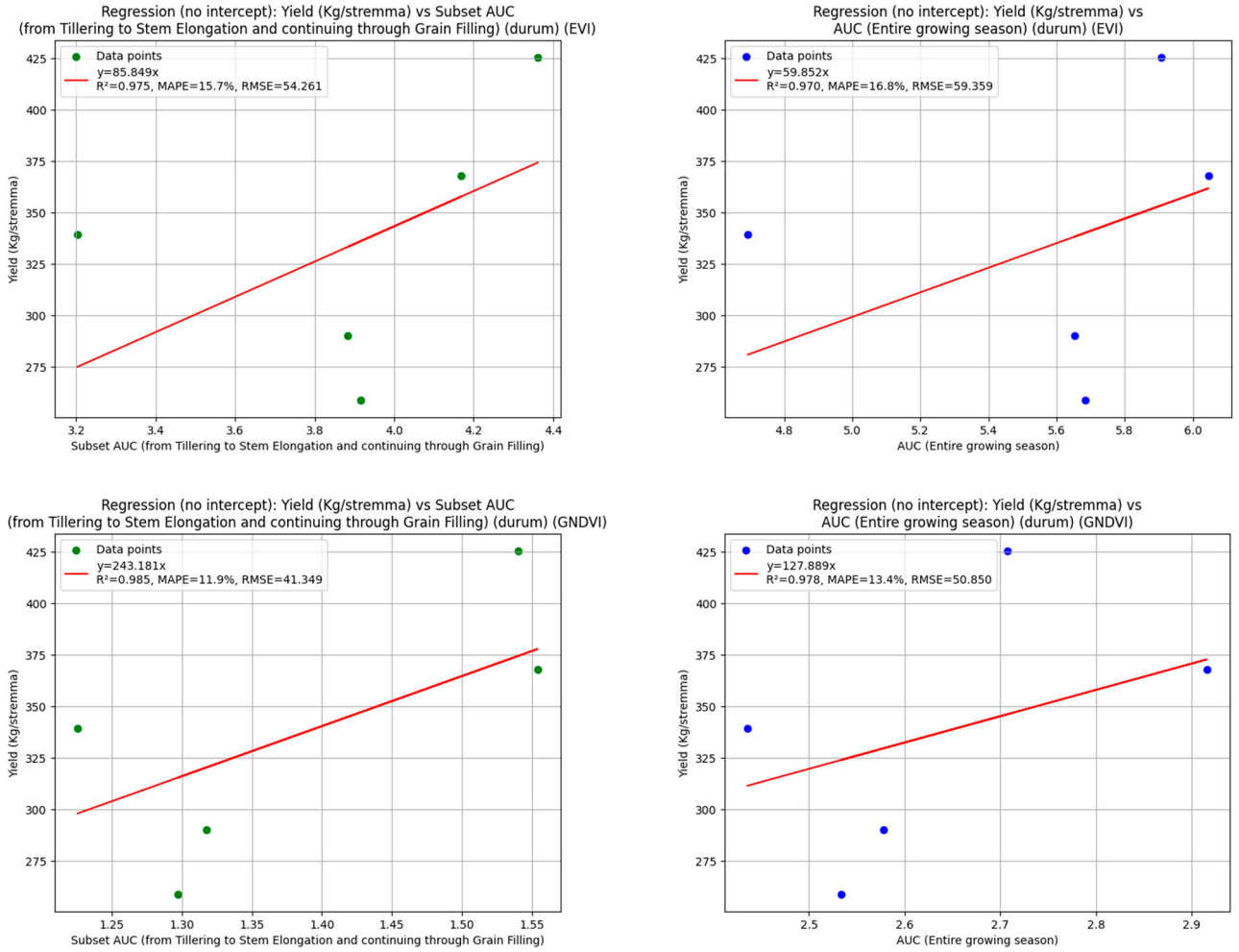


Figure 5. *Cont.*

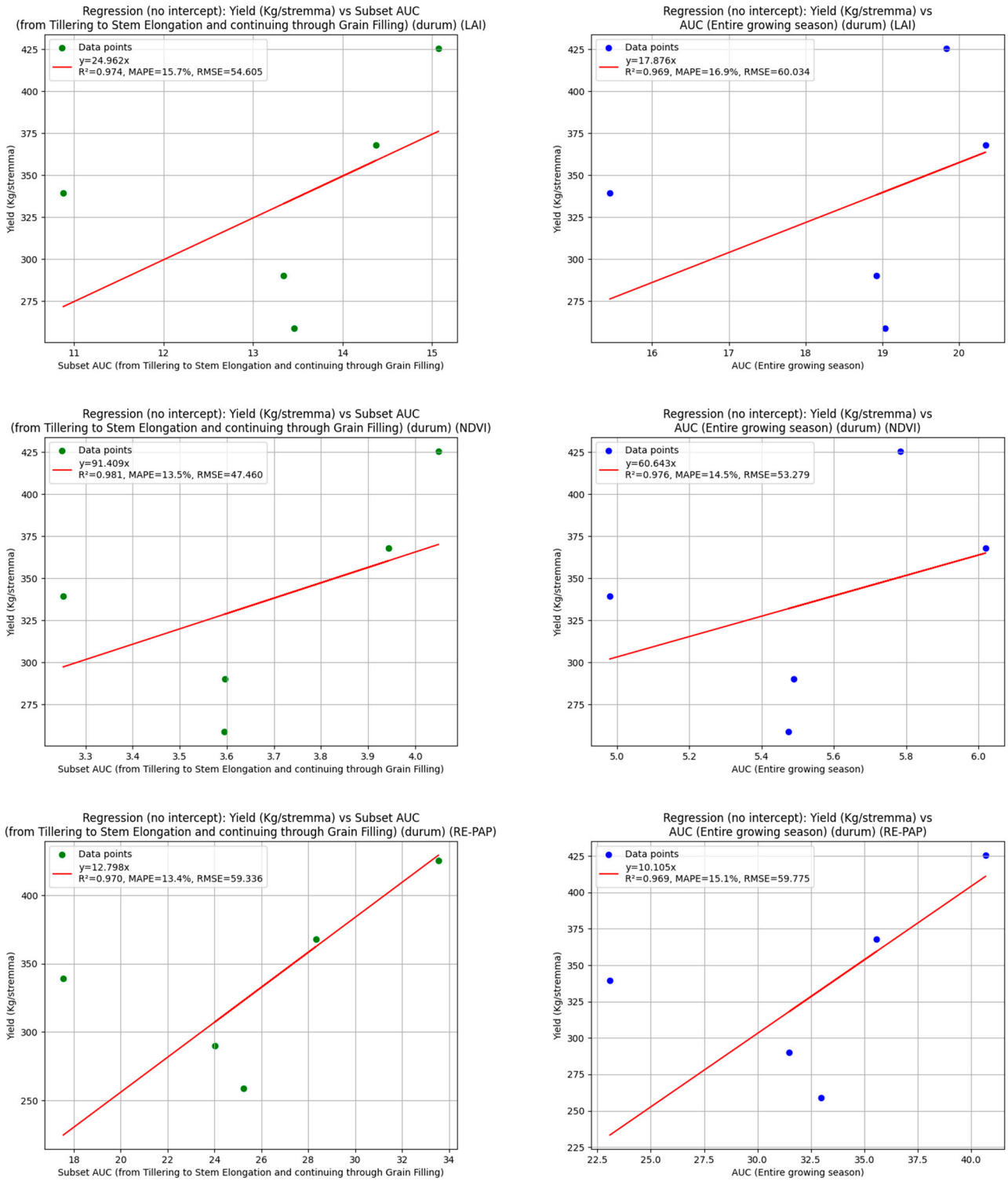
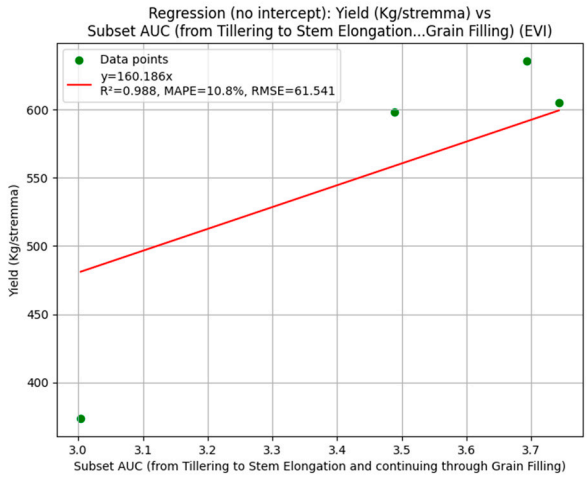
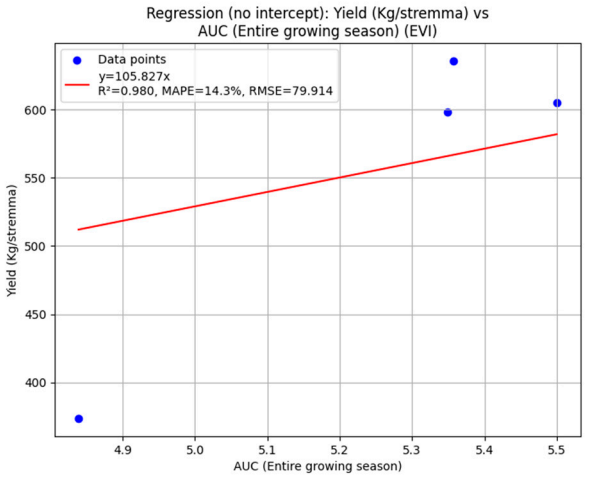
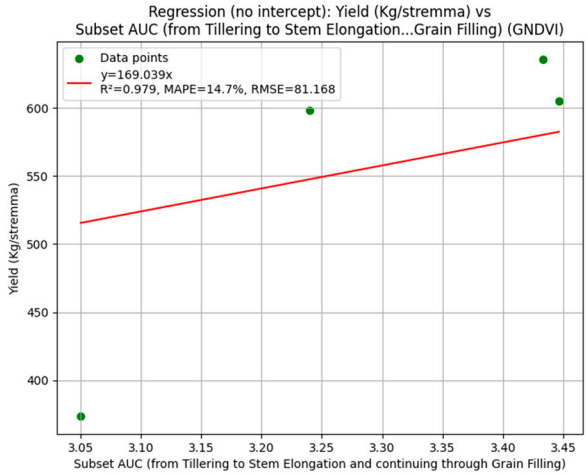
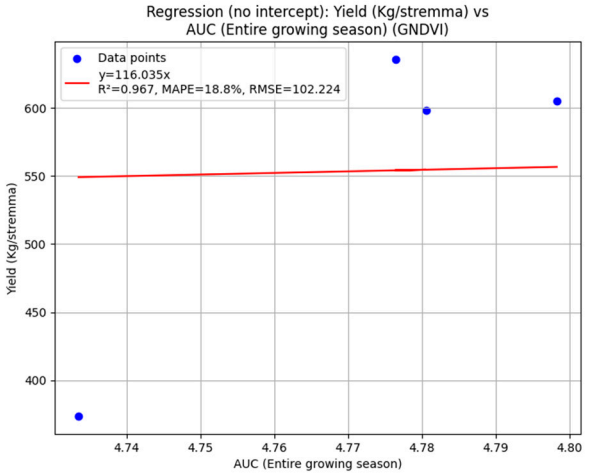


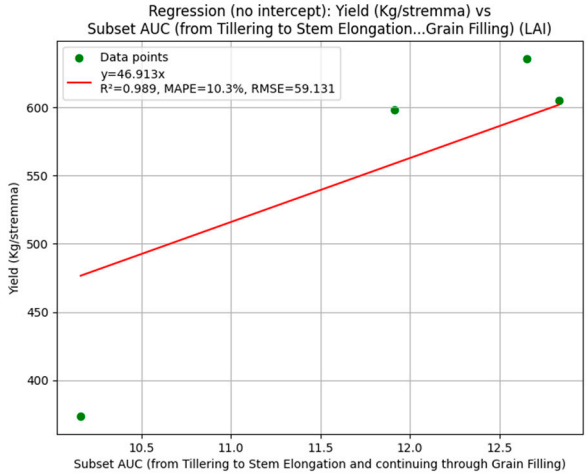
Figure 5. Regression analysis plots for durum wheat: subset period (**left**) and entire growing season (**right**).











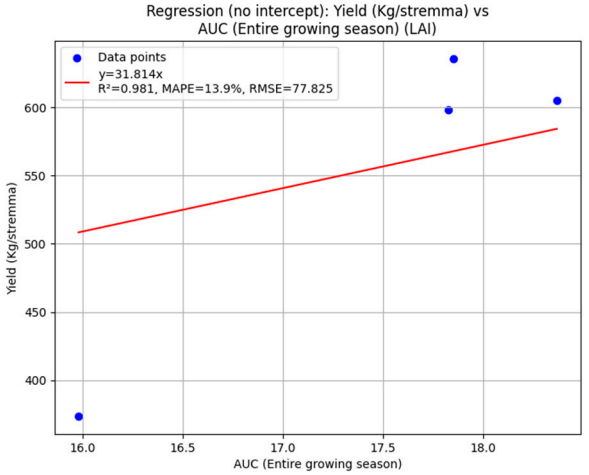


Figure 6. *Cont.*

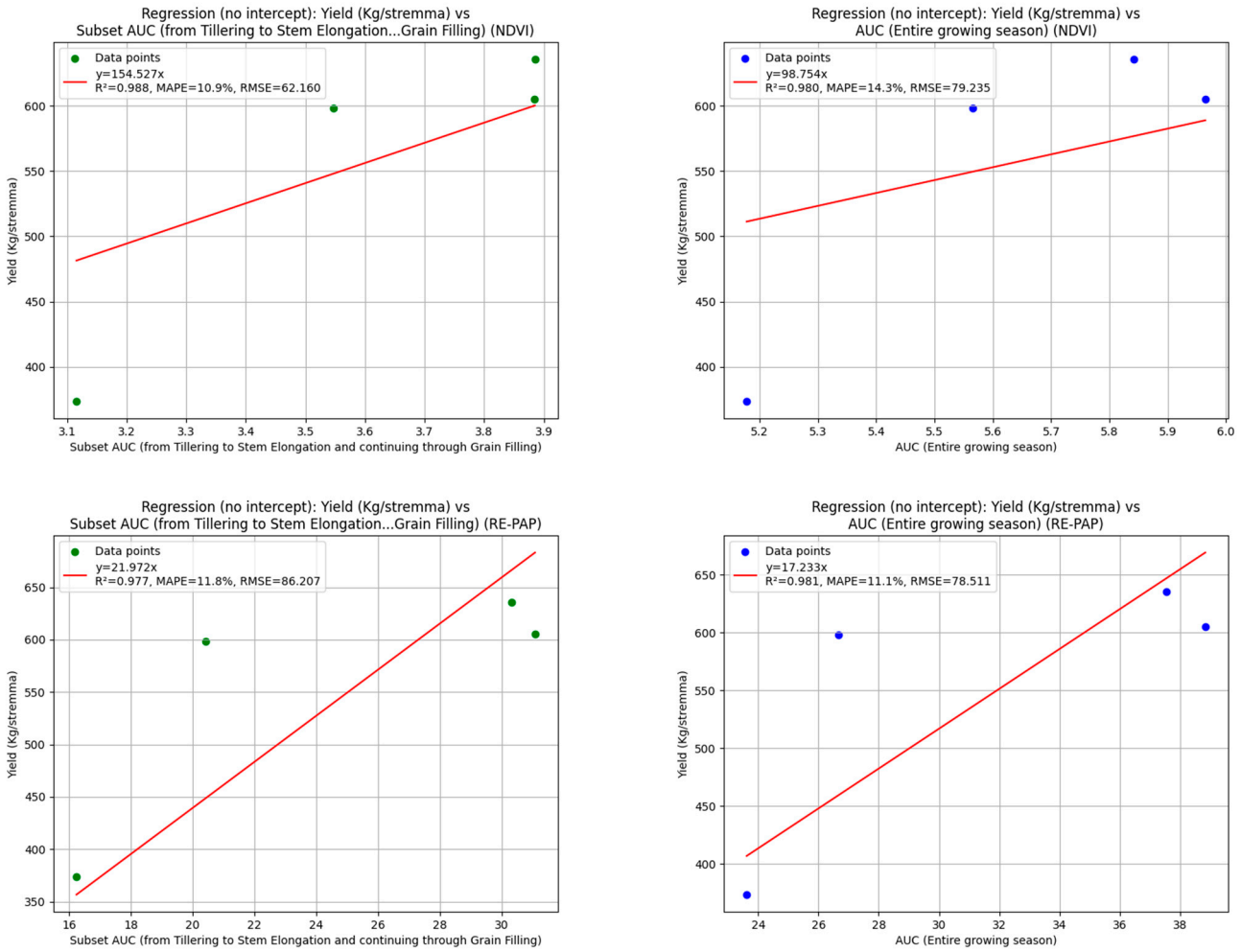


Figure 6. Regression analysis plots for soft wheat: subset period (**left**) and entire growing season (**right**).

The regression analysis conducted for both durum and soft wheat yields demonstrates exceptionally high Coefficient of Determination (R^2) values, indicative of a strong linear relationship between yield and the AUC of vegetation indices. However, these high R^2 values may partly reflect overfitting, which can be attributed to several factors. Firstly, the relatively small dataset examined in this study may limit the generalizability of the results and amplify the appearance of a strong fit. Secondly, the characteristics of the data itself, including the variability and distribution of the AUC values, most likely contribute to the observed trends. Lastly, the use of a zero-intercept model was deliberately chosen due to its physical relevance as a scenario where AUC = 0 logically corresponds to yield = 0. This approach ensures that the regression equation is rooted in agronomic reality.

By modeling yield as a function of AUC rather than directly using vegetation index values, this study aims to establish a predictive framework grounded in an independent variable with clear temporal and physiological significance. While the results highlight the potential of the models, further validation with larger datasets and additional variables is necessary to confirm their predictive reliability and broader applicability.

For durum wheat, the Root Mean Square Error (RMSE) ranges from 40 to 60 kg per stremma (1 stremma = 1000 m^2), with the best performance observed for the GNDVI during the subset of the growing season (Tillering to Stem Elongation through Grain Filling), achieving a Mean Absolute Percentage Error (MAPE) of 11.9% and an RMSE of 41.35. The second-best performance was recorded for the NDVI during the subset period,

followed by the GNDVI for the entire growing season. These findings emphasize the strong predictive capacity of GNDVI, particularly when focused on critical phenological stages, and the reliability of NDVI during the same subset period for yield estimation.

In the case of soft wheat, the Root Mean Square Error (RMSE) spans from 59 to 102 kg per stremma (kg per 1000 m^2). The most accurate predictions were achieved using the LAI during the subset of the growing season (Tillering to Stem Elongation through Grain Filling), which yielded a Mean Absolute Percentage Error (MAPE) of 10.3% and an RMSE of 59.13. The EVI during the same subset period ranked as the second-best predictor, reflecting its close relationship with LAI, as the latter is derived from EVI. Following this, the NDVI for the subset period also demonstrated strong predictive performance. These results highlight the robustness of LAI and related indices when focused on critical growth stages for soft wheat yield estimation.

Across both wheat types, models based on the subset of the growing season consistently outperform those utilizing the full growing season in terms of predictive accuracy. This underscores the significance of focusing on critical phenological stages, where crop growth and yield formation are most responsive to environmental and management factors. Furthermore, the models for durum wheat demonstrate relatively lower prediction errors, indicating a more consistent and robust relationship between AUC values and yield. These findings highlight the potential for subset period models to enhance yield prediction accuracy while emphasizing the distinct physiological and agronomic characteristics influencing each wheat type.

By examining the best-performing index, GNDVI (durum wheat), in relation to phenological stages, as illustrated in Figure 7, it becomes evident that fields with larger yields, such as Field 1 (425.59 kg/stremma) and Field 4 (367.79 kg/stremma), when compared to lower-yielding fields such as Field 2 (290 kg/stremma) and Field 3 (258.73 kg/stremma), exhibit greater GNDVI values during the Stem Elongation to Booting stages. This finding suggests that these phenological stages are particularly critical for yield prediction. Moreover, Field 1 shows notably higher GNDVI values during the Booting and Heading stages than Field 4, which may highlight the additional, albeit less pronounced, importance of these stages in supporting yield potential. Field 6, with a yield of 339.34 kg/stremma, performs slightly better than Fields 2 and 3. This advantage persists despite Field 6 having lower GNDVI values during the Stem Elongation to Booting stages. The increased GNDVI values observed in Field 6 from the end of Booting to Flowering likely contribute to its performance, as Fields 2 and 3 display a slight decline in GNDVI during the same period, particularly at the Heading stage. This trend underscores the potential impact of later phenological stages on yield outcomes when earlier growth stages are less optimal.

The evaluation of the best-performing index, LAI (soft wheat), in relation to phenological stages, as depicted in Figure 8, reveals that fields with higher yields, such as Field 3 (636 kg/stremma), Field 4 (605 kg/stremma), and Field 2 (598 kg/stremma), tend to maintain elevated LAI values during the critical period spanning from Stem Elongation to Flowering. In contrast, the lower-yielding Field 1 (374 kg/stremma) shows comparatively reduced LAI values during these stages, highlighting their central role in influencing productivity. Additionally, variations in yield among the top-performing fields appear to be influenced by the LAI dynamics from Booting to Grain Filling, indicating that maintaining robust LAI values during these growth phases is crucial for maximizing yield potential.

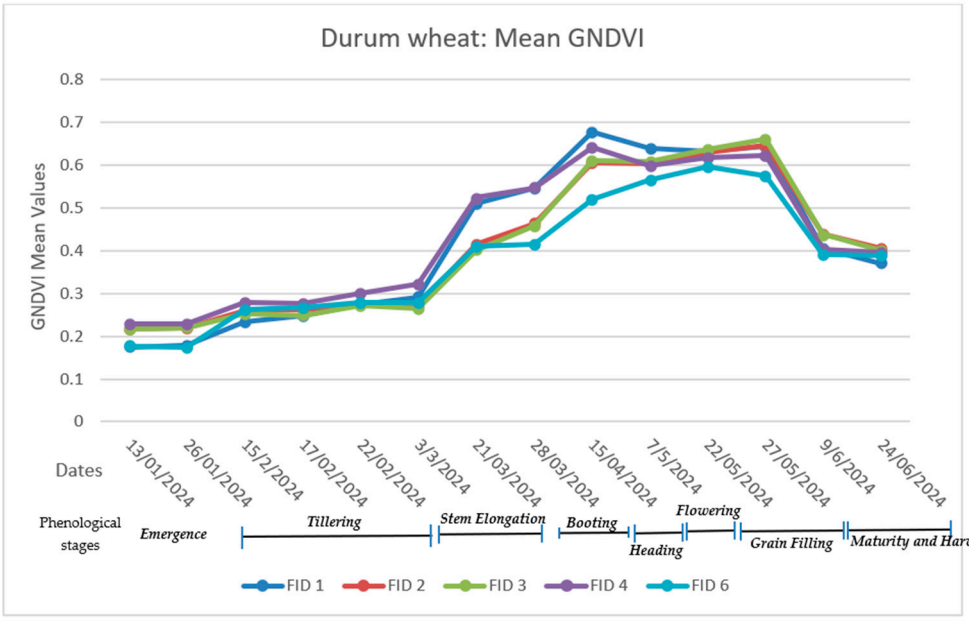


Figure 7. Growing season plot of GNDVI.

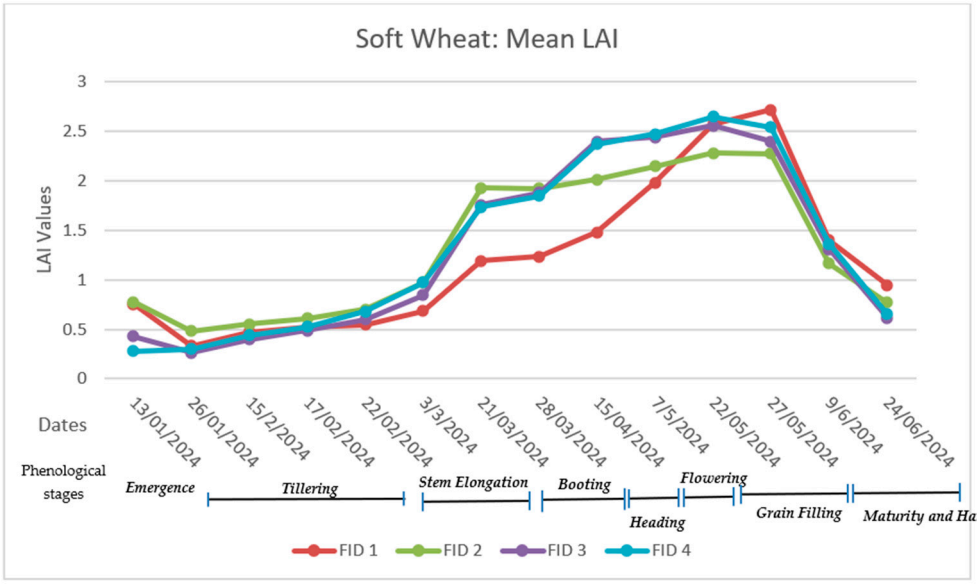


Figure 8. Growing season plot of LAI.

3.2. Functional Overview and User Interaction Design for Web Application Development

The web application features a carefully designed graphical user interface (GUI) that prioritizes accessibility and efficient navigation. The interface is systematically divided into two primary sections: the left panel facilitates task selection and configuration, while the right panel presents the outcomes of executed tasks through an interactive map, complemented by contextual informational notifications. This structured layout is intentionally developed to enhance usability and support effective user interaction, ensuring a coherent and productive experience.

The application comprises four (4) primary functionalities:

1. Search Sentilel-2 images: As shown in Figure 9, users can search for satellite imagery (Sentinel-2 L2A, Surface Reflectance) based on specific parameters, including area of interest (AOI; this can be uploaded in either GeoJSON or KML format), cloud cover percentage, and date ranges. Additionally, users have the option to upload their dates

of interest in a text format, with dates separated by commas. Moreover, users can optionally apply a cloud mask and a shadow threshold to remove cloud-covered areas and pixels affected by cloud shadows.

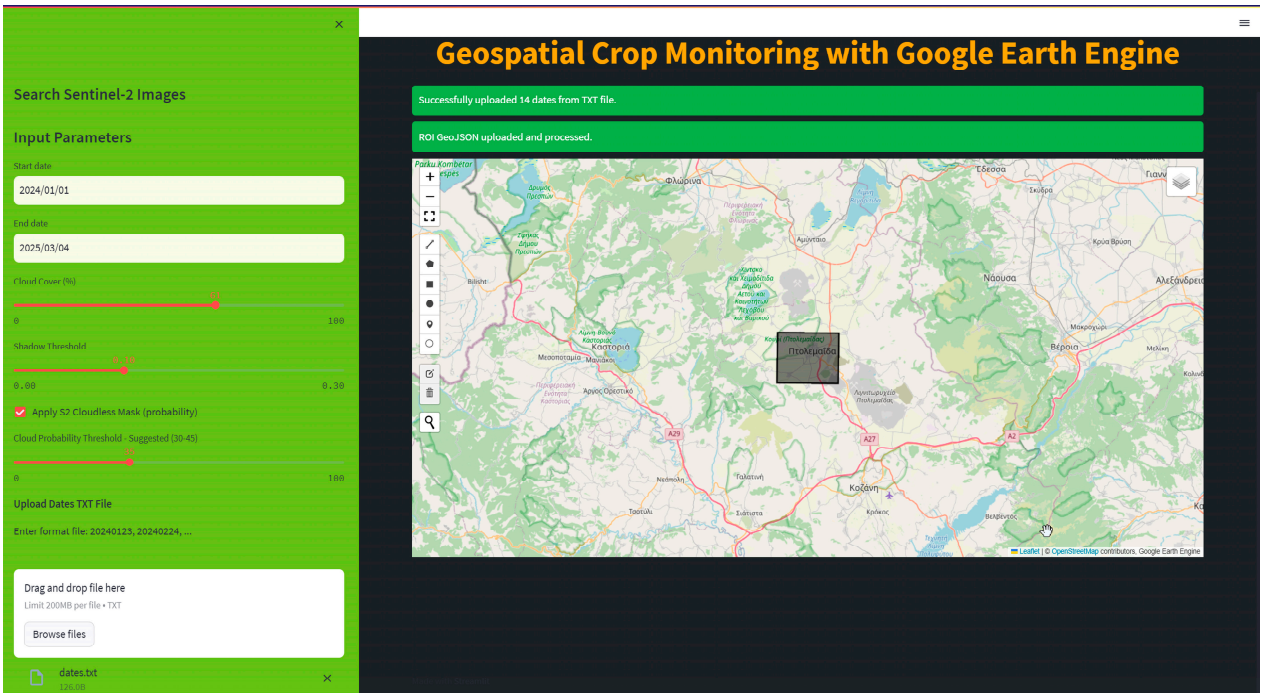


Figure 9. Search for available Sentinel-2 images.

2. Vegetation index calculator: This feature enables users to select and compute various spectral indices from the chosen Sentinel-2 images, Figure 10. The resulting vegetation index maps are displayed interactively and can be exported. Additionally, users can generate and visualize histograms of the computed indices.

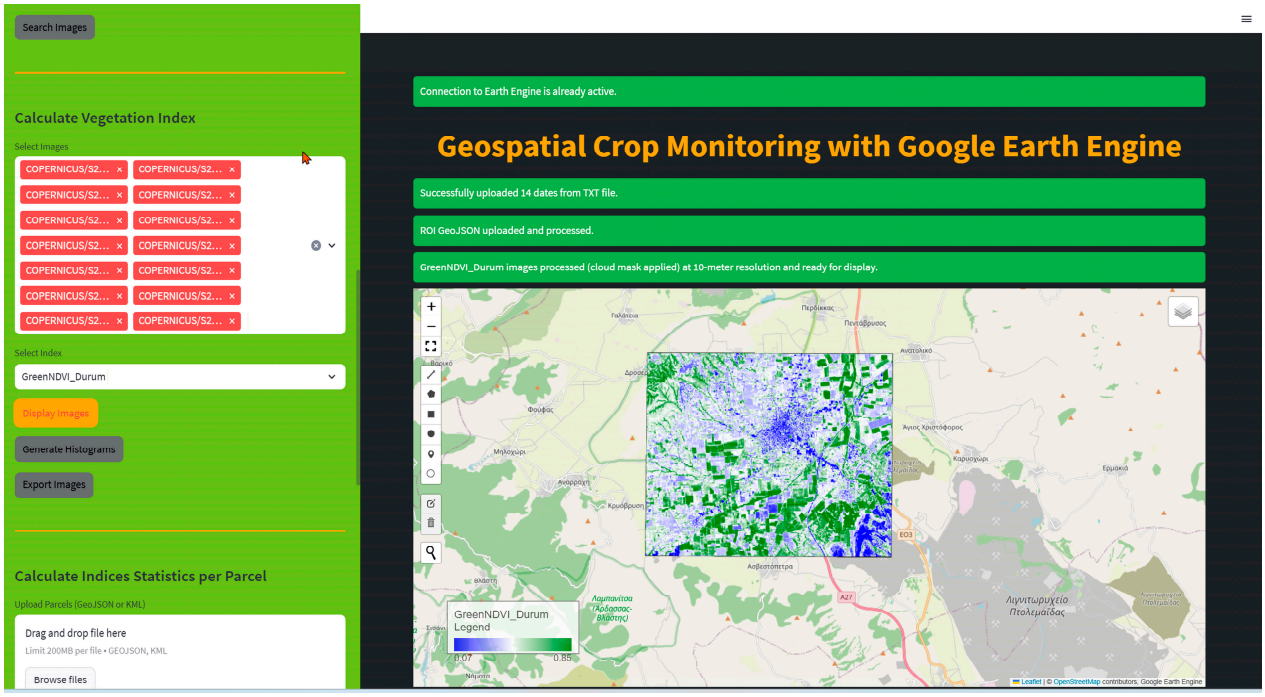


Figure 10. Vegetation index calculator.

3. VIs statistics calculator: This feature allows users to upload their fields and calculate statistics for previously computed vegetation indices for their areas of interest, Figures 11 and 12. Users can also specify a time period to subset the VI statistics and generate corresponding plots.

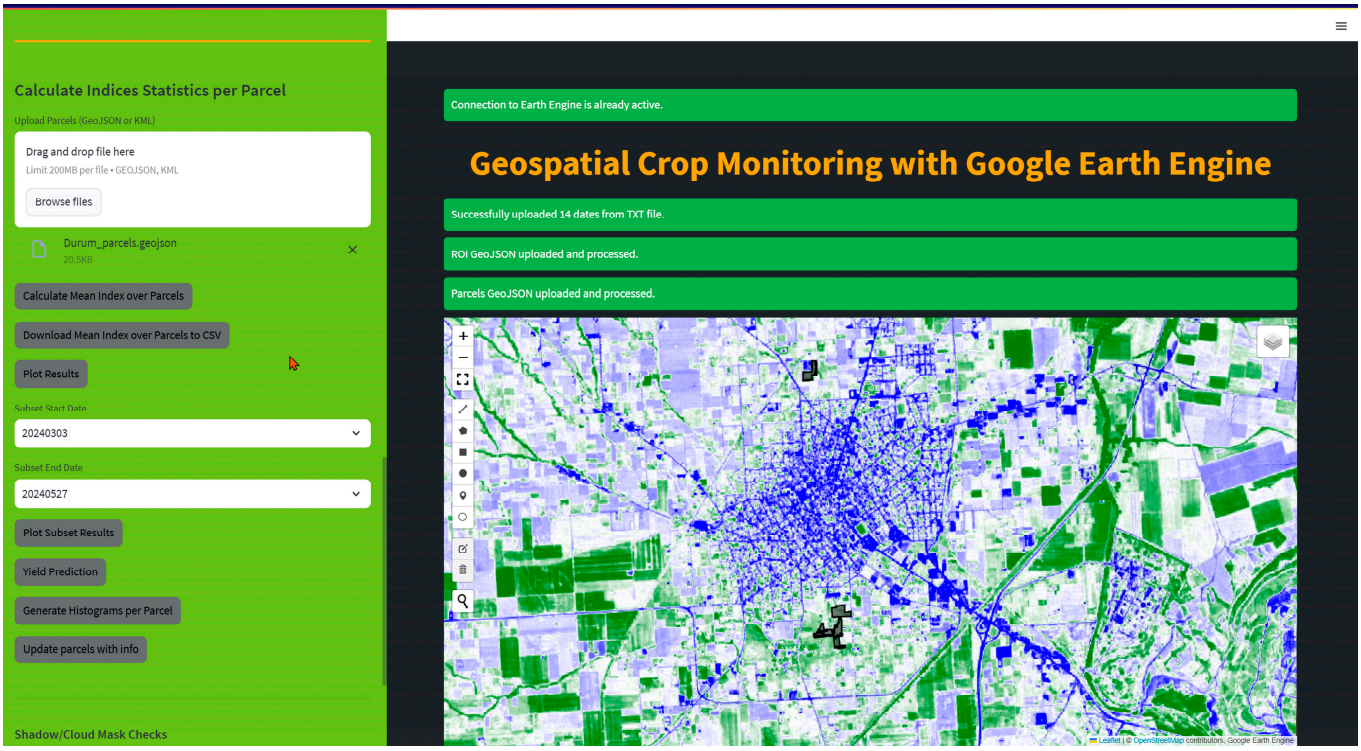


Figure 11. VI statistics per field—uploaded fields are shown on the map.

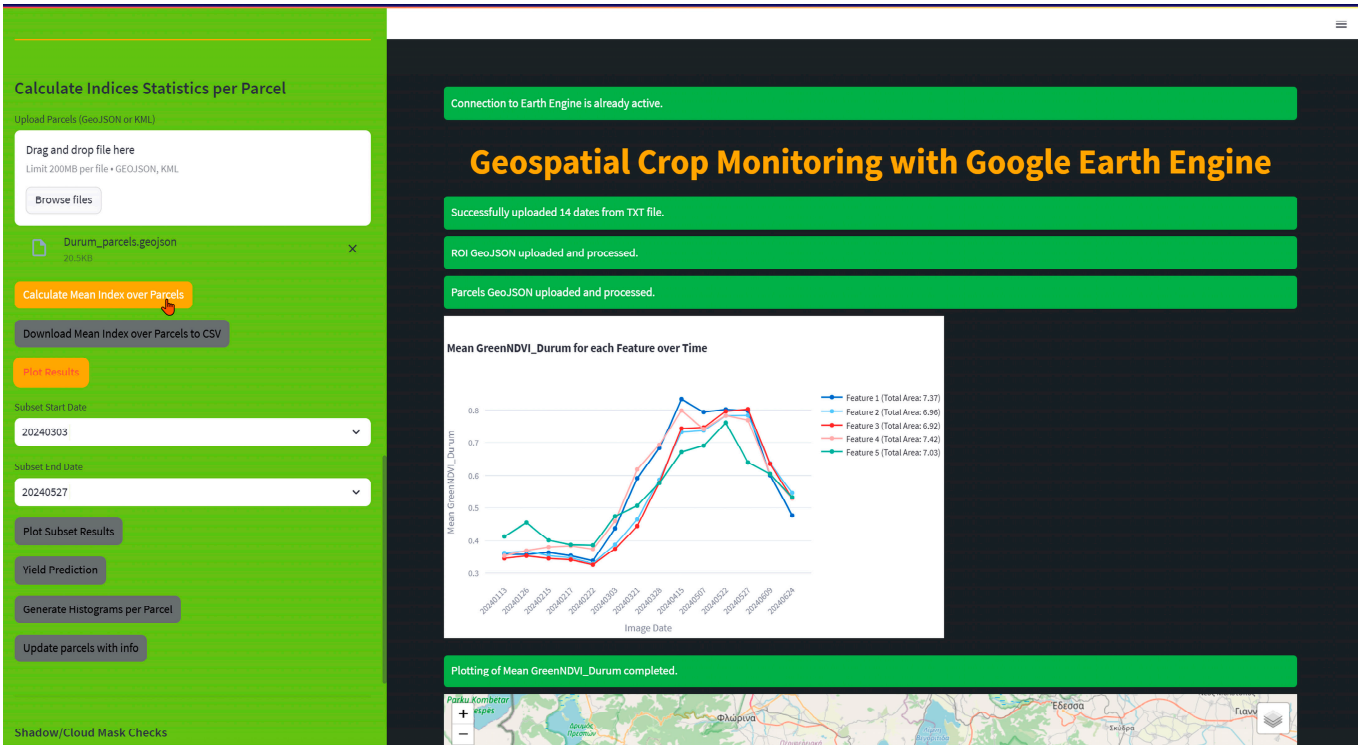


Figure 12. VI statistics per field: displays the plotted results of vegetation index statistics for each field.

4. Yield prediction: This feature leverages the selected vegetation index to predict crop yield for the entire growing season or a specified subset of the season, Figure 13. Additionally, it offers the option to update the fields with the generated yield predictions, Figure 14.

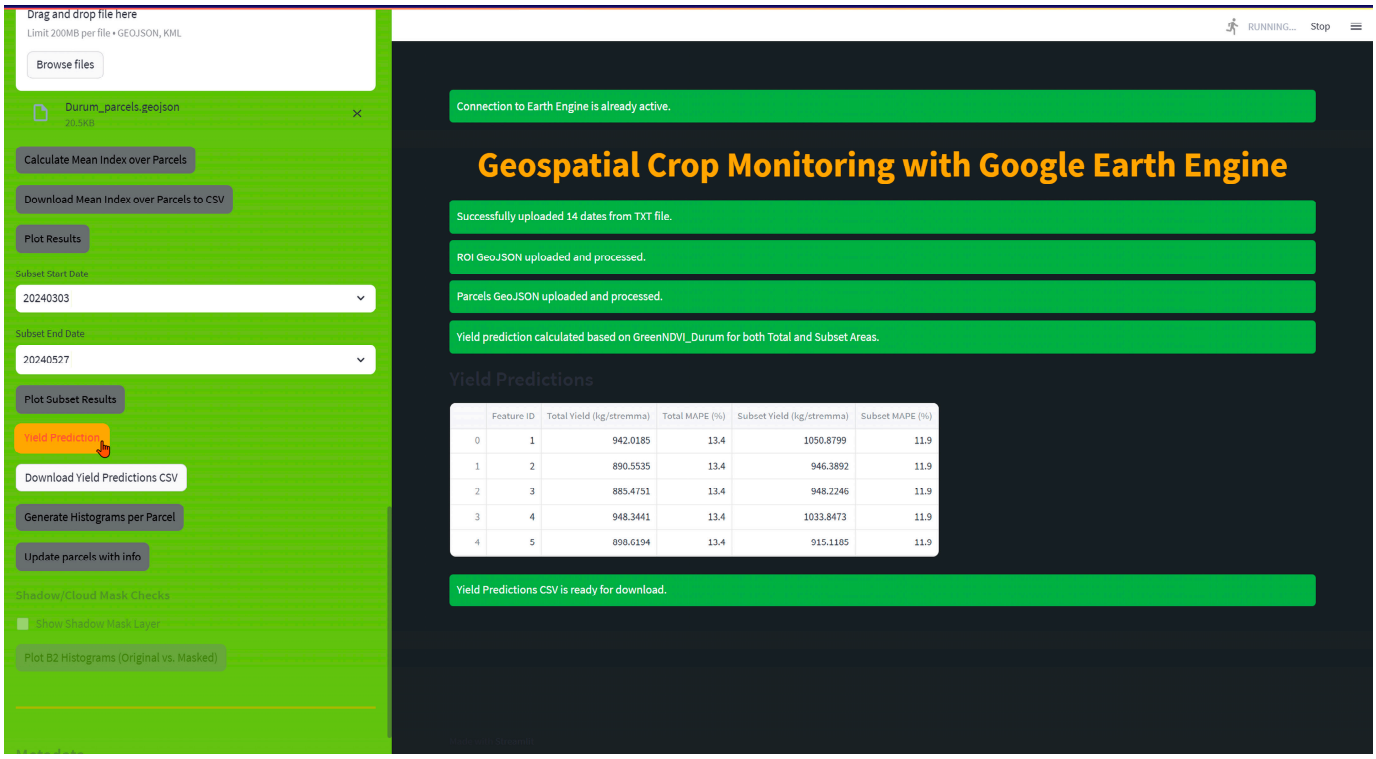


Figure 13. Yield prediction.

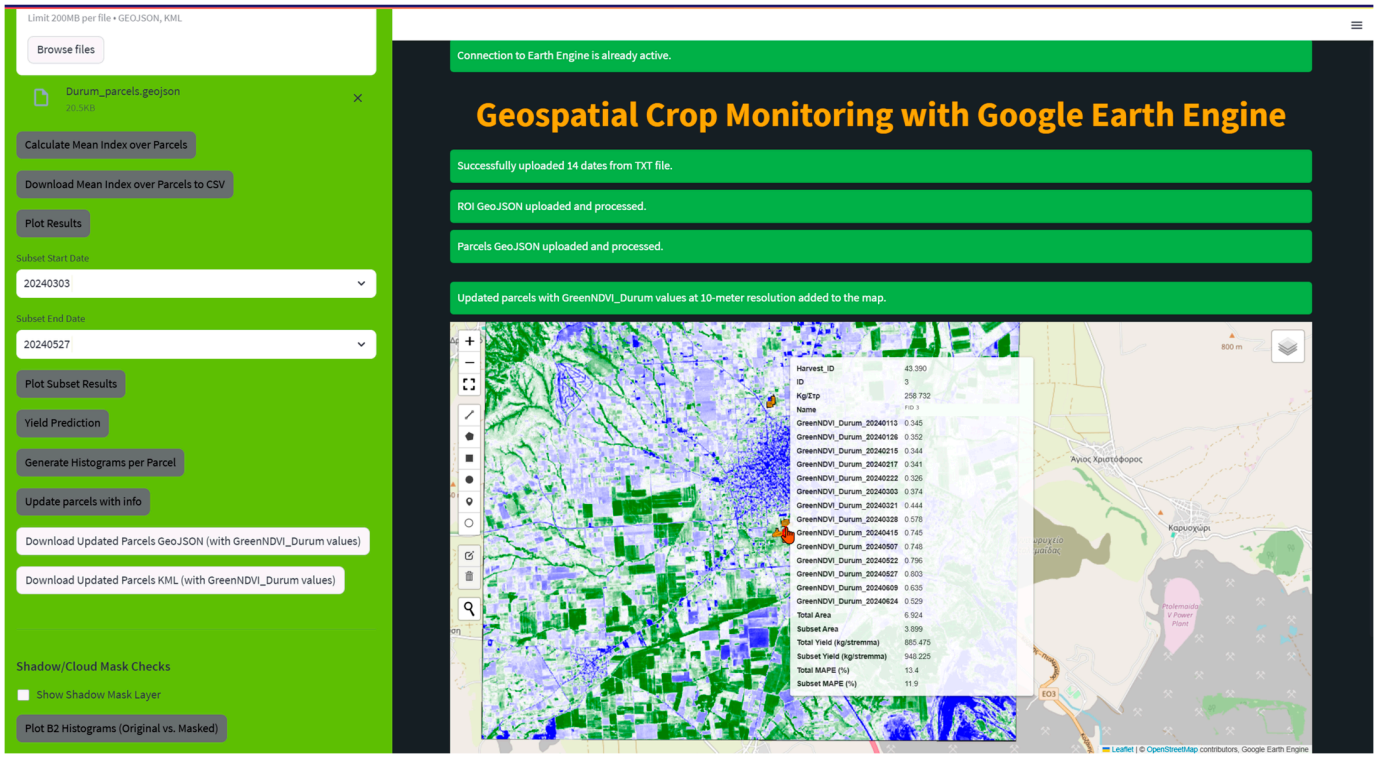


Figure 14. Yield prediction—updated information per field.

4. Discussion

4.1. Yield Estimation

Since the 1970s, the employment of satellite remote sensing has been crucial in the assessment of key agricultural crops by scrutinizing agro-climatic conditions, crop health, and yield predictions, with certain systems also evaluating food security to deliver early warnings of possible food deficits. The integration of free access to medium to high-resolution satellite data with cloud-computing platforms like Amazon, Google Earth Engine, and Microsoft AI for Earth has significantly enhanced crop monitoring capabilities by overcoming previous constraints related to data processing and information extraction [6].

Pre-harvest yield estimation insights provide farmers and experts involved in farming with essential information to guide critical decisions, such as determining the ideal quantity of fertilizers to apply [60]. The predictive efficacy of various vegetation indices (VIs) in relation to wheat yield exhibits considerable variability across research investigations, with each index presenting distinct benefits contingent upon the specific context and methodological approach utilized. The Normalized Difference Vegetation Index (NDVI) is frequently used due to its simplicity and suitability for predicting winter wheat yield, particularly when combined with meteorological data [61]. Other indices such as the Modified Soil-Adjusted Vegetation Index (MSAVI) and the Green Normalized Difference Vegetation Index (GNDVI) have illustrated improved correlation coefficients with real yield metrics, especially during critical growth periods including Flowering [62]. The Enhanced Vegetation Index (EVI) has been competently used together with machine learning methodologies, including Random Forest (RF), to estimate winter wheat yields, demonstrating a strong relation with yield estimation (an R^2 of 0.81 and an RMSE of 1250 kg/ha) and providing substantial insights for winter wheat yield forecasting, especially when data from various growth intervals are factored in [63].

For durum wheat, GNDVI outperformed all other tested indices, consistent with findings from previous studies [64]. Furthermore, GNDVI also demonstrated superior predictive capabilities over NDVI, which ranked as the second most effective predictor [65]. For soft wheat, LAI, derived from EVI, demonstrates superior performance over all other examined indices. However, its predictive capability shows only a marginal difference when compared to NDVI, as corroborated by findings in previous studies [66].

4.2. Web Application Development

The developed web application represents a significant advancement in wheat growth monitoring by utilizing Sentinel-2 satellite data to deliver actionable agricultural insights. This advancement holds significant implications for smallholder agricultural producers, agronomists, and professionals in the agricultural sector who encounter both financial limitations and technical challenges. By integrating open-source frameworks and leveraging freely available Sentinel-2 data, the application offers an accessible and cost-effective solution that addresses the economic barriers often associated with traditional crop monitoring systems. This affordability is complemented by the automation of data processing workflows, which provide real-time insights into wheat growth across phenological stages. These capabilities empower users to make informed decisions regarding resource allocation, such as optimizing fertilizer application and irrigation schedules, thereby enhancing productivity and promoting sustainable agricultural practices.

The application's scalability and flexibility further extend its utility across diverse agricultural contexts. Users can customize their analyses by selecting vegetation indices, adjusting parameters, and focusing on specific growth stages, enabling the tool to accommodate both small-scale and large-scale farming operations. The intuitive graphical user interface facilitates accessibility for individuals possessing diverse degrees of technical

proficiency, thereby effectively connecting sophisticated geospatial technologies with their pragmatic applications in the field. Through these features, the web application fosters a more inclusive approach to precision agriculture, providing a robust framework for systematic crop monitoring.

4.3. Future Work

Future research should include repeated analyses in subsequent cultivation periods to further validate the findings of this study, as well as an expansion of the study to incorporate a larger number of fields for the cross-validation of results. Additionally, evaluating the model's applicability across different geographical regions with varying soil properties and climatic conditions is essential to assess its sensitivity to these factors. Another important direction for future research is extending the analysis to include other crops, such as maize and sunflowers, to evaluate the model's adaptability and predictive capabilities.

Since the application is designed to assist small-scale farmers and agronomists with limited knowledge of geospatial technology, future work could focus on further automation. This includes estimating phenological stages based on sowing dates provided by farmers combined with historical climate data—such as regional climatic zones—to determine the necessary imagery for each stage. These advancements will enhance the application's ability to deliver more precise and actionable insights to farmers. Additionally, integrating Natural Language Processing (NLP) [67] into the web application will further enhance usability. By leveraging NLP, the system will minimize the time and effort required for farmers to interact with the application, allowing for intuitive communication with minimal input—such as sowing dates—while improving accessibility and ease of use. Moreover, the application possesses significant potential for future advancements beyond its existing functionalities, especially in the realm of incorporating artificial intelligence (AI) to augment the analysis of geospatial data. The use of innovative machine intelligence frameworks, featuring both machine learning and deep learning approaches, offers a remarkable opportunity to enrich the functionality of systems in exploring massive datasets, uncovering essential correlations and crafting accurate prediction models. For example, incorporating AI-driven methods for disease detection, pest management, and multi-crop monitoring would further enhance its utility and relevance. Such technological advancements would facilitate the application's capacity to tackle a wider array of challenges inherent in precision agriculture, thereby augmenting global initiatives directed towards the enhancement of food security and the promotion of environmental sustainability.

4.4. Limitations

A significant limitation of this study is the availability of data for every phenological stage. Adverse meteorological conditions in a given region may hinder the acquisition of satellite imagery during one or more phenological stages. This issue can be partially mitigated by utilizing alternative satellite imagery sources (e.g., Landsat), which offer different revisit times. However, the limitation persists, as unfavorable weather conditions may still prevent the capture of high-quality data over the study area.

Another limitation concerns the spatial resolution of freely available satellite data. In the case of very small fields, the presence of mixed pixels—where a single pixel encompasses multiple land cover types—may reduce model accuracy.

Lastly, a critical limitation is the presence of weeds, as they can significantly alter vegetation index values, leading to results that do not accurately reflect true crop growth. Consequently, this may result in false predictions, as the AUC values will be artificially inflated.

5. Conclusions

In this study, the monitoring of wheat crops, including both durum and soft wheat, was conducted throughout the growing season to investigate the relationship between yield and the independent variable "Area Under the Curve" (AUC) of vegetation indices. This research demonstrates the effectiveness of specific vegetation indices (VIs) in estimating wheat yield, with GNDVI emerging as the most reliable predictor for durum wheat and LAI for soft wheat. Phenological stages are widely recognized for their pivotal role in determining crop productivity, and the integration of AUC offers a robust method for correlating yield with dynamic changes observed across these stages. The findings of this study highlight the significance of specific phenological stages for yield prediction. The results indicate that the time period spanning from Tillering to Stem Elongation and extending through Grain Filling plays a pivotal role in wheat crop yield monitoring. This observation aligns with the existing literature reinforcing the critical importance of these growth stages in determining crop productivity. The results of this investigation emphasize the resilience of the suggested methodology; however, additional inquiry is essential to enhance its relevance. The expansion of the dataset to incorporate a larger array of wheat cultivation areas, additional growth periods, and diverse geographic regions will substantially enhance the external validity and rigor of the results. Such efforts will provide deeper insights into the spatial and temporal dynamics of wheat crop yield prediction, paving the way for more comprehensive monitoring and management strategies in varied agricultural contexts. While the proposed framework shows strong predictive capabilities, certain limitations exist, including the impact of adverse weather conditions on data availability, spatial resolution constraints, and potential interference from weeds. Additionally, the web application exemplifies a robust and user-centered methodology for agricultural monitoring, providing real-time actionable intelligence while maintaining cost-efficiency and accessibility. Its capacity to facilitate data-informed decision making accentuates its significance in endorsing sustainable practices and enhancing agricultural productivity. The integration of artificial intelligence for geospatial data analysis signifies a promising pathway for augmenting the application's functionalities, thereby laying the groundwork for more comprehensive and intelligent solutions within the realm of precision agriculture.

Author Contributions: Conceptualization, K.P. and K.N.; methodology, K.N.; software, G.G.; validation, K.P. and K.N.; formal analysis, K.N.; investigation, V.D.-P.; resources, V.D.-P.; data curation, V.D.-P. and G.G.; writing—original draft preparation, K.P.; writing—review and editing, K.P. and K.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data are not available for publication as they remain the private property of the farmer.

Acknowledgments: The authors express their gratitude to Pavlos Alexandridis, a dedicated farmer, for generously providing the invaluable data that made this study possible.

Conflicts of Interest: Author Georgios Gkologkinas was employed by NubiGroup Geoservices & Research Private Company. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A

```
import x, y, z  . . .  # Import libraries (e.g., streamlit, ee, folium, plotly, etc.)
st.set_page_config(. . .)  # Configure page (title, layout, sidebar size, etc.)
st.markdown(custom_css)  # Add CSS styling for appearance and customization of buttons/fields
-------------------------------------------------------------------------------------
# Functions
-------------------------------------------------------------------------------------
def custom_notification(message, color):
        # Display a custom "alert" box in Streamlit
        # e.g., red for error, green for success, etc.
def authenticate_gee():
        # Check if there's an active Google Earth Engine (GEE) session
        # If not, run ee.Authenticate() and ee.Initialize()
def kml_to_full_geojson(kml_path):
        # Load and convert a KML file into a GeoJSON
        # Retain all available fields (name, description, extended data, etc.)
def search_images(start_date, end_date, cloud_limit):
        # Search Sentinel-2 images for the ROI (Region of Interest)
        # based on time range [start_date, end_date] and cloud coverage (cloud_limit)
        # Returns a DataFrame with image IDs, cloud coverage, and dates
def calculate_index(image, index):
        # Compute a vegetation index, e.g., NDVI, EVI, LAI, etc.
        # using the appropriate spectral bands (B2, B4, B8, etc.)
def calculate_min_max(image, index):
        # Calculate min and max values of the index for better visualization on the map
        # Uses reduceRegion with histogram and applies cutoffs (e.g., 2–98%)
def add_legend(map_obj, title, min_val, max_val, colors):
        # Create and position a dynamic legend (e.g., a linear color scale) on the map
def log_download(filename):
        # Keep track of the name of the file the user downloads
def plot_histogram_on_streamlit(bin_edges, bin_counts, index_name, image_label):
        # Use the Plotly library (e.g., go.Figure, go.Bar) to build a histogram
        # and display it with st.plotly_chart
# (NEW) Cloud and Shadow Masking function
def maskImageWithS2Cloudless(sr_image, cloud_prob_threshold, shadow_threshold = 0.1):
        """
        Removes cloudy pixels based on S2 Cloud Probability
        and dark (shadow) pixels based on a brightness threshold.
        Ensures the final result is an ee.Image (not a ComputedObject).
        """
        # 1) Retrieve the matching COPERNICUS/S2_CLOUD_PROBABILITY image by system:index
        # 2) Create 'cloud_mask' = (probability < cloud_prob_threshold)
```

```
        # 3) Create 'shadow_mask' = (min(B2,B11,B12) > shadow_threshold)
        # 4) Combine with logical AND
        # 5) Wrap ee.Algorithms.If(...) with ee.Image(...) to avoid 'ComputedObject' errors

-----------------------------------------------------------------------------------
# Session Initialization/Variables
-----------------------------------------------------------------------------------
ee.Initialize()                          # Initial connection to Earth Engine (if not already done)
Map = geemap_folium.Map()                # Create a base map for display in Streamlit
st.session_state.roi = None              # ROI is not set initially
st.session_state.parcels = None          # Parcels have not been uploaded yet
st.session_state.selected_images = []    # No selected images for now
...    # (Define additional session_state variables as needed)
-----------------------------------------------------------------------------------
# Sidebar: Parameter inputs, file uploads, buttons
-----------------------------------------------------------------------------------
start_date = st.sidebar.date_input(...)      # Input start date
end_date = st.sidebar.date_input(...)        # Input end date
cloud_slider = st.sidebar.slider(...)        # Slider for cloud coverage (e.g., 0–100%)
# (NEW) Checkbox and sliders for cloud and shadow masking
apply_cloud_mask = st.sidebar.checkbox("Apply Cloud/Shadow Mask", value = False)
cloud_prob_threshold = st.sidebar.slider("Cloud Probability Threshold", 0, 100, 40)
shadow_threshold = st.sidebar.slider("Shadow Threshold", 0.0, 0.5, 0.1, step = 0.01)
# Uploader for ROI (GeoJSON/KML)
uploaded_roi = st.sidebar.file_uploader(...)
if uploaded_roi is not None:
        # Read the file
        # Convert to GeoJSON if KML
        # Convert to ee.Geometry/ee.FeatureCollection for use in GEE
        st.session_state.roi = geojson_to_ee(roi_geojson)

# Uploader for Parcels (GeoJSON/KML)
uploaded_parcels = st.sidebar.file_uploader(...)
if uploaded_parcels is not None:
        # Read the file
        # Convert to full GeoJSON
        # Save in st.session_state.parcels as ee.FeatureCollection

# Button to search images
if st.sidebar.button('Search Images'):
        # Call search_images(...)
        # Display the results (date, ID, cloud cover) in a DataFrame

# Multiselect to pick specific images
# e.g., st.sidebar.multiselect('Select Images', list_of_image_ids)
# Select Vegetation Index (e.g., NDVI, EVI, etc.)
selected_index = st.sidebar.selectbox(...)
# "Display Images" button
if st.sidebar.button('Display Images'):
        # For each selected image ID:
        #    1) Load ee.Image
```

```
        #    (NEW) 2) If apply_cloud_mask is True, call maskImageWithS2Cloudless(...)
        #   e.g. image = maskImageWithS2Cloudless(image, cloud_prob_threshold, shadow_
threshold)
        #    3) Compute the index (calculate_index)
        #    4) Optionally remove/clear old layers
        #    5) Add layer to the map (Map.addLayer)
        #    6) Create legend (add_legend)

# "Generate Histograms" button
if st.sidebar.button('Generate Histograms'):

        # For each selected image:
        #    1) reduceRegion with histogram
        #    2) Retrieve bin_edges, bin_counts
        #    3) plot_histogram_on_streamlit(...)

# "Export Images" button
if st.sidebar.button('Export Images'):

        # Export the computed images (e.g., NDVI) to GeoTIFF with geemap.ee_export_image
        # Temporarily save, read into bytes
        # Create download_button for each image

-------------------------------------------------------------------------------------
# Compute Index per Parcel
-------------------------------------------------------------------------------------
if st.sidebar.button('Calculate Mean Index over Parcels'):

        # For each feature/parcel in st.session_state.parcels:
        #    1) Create geometry
        #    2) Use index_image.reduceRegion(reducer = ee.Reducer.mean(), geometry = ...,
scal e= 10)
        #    3) Store the result in a dictionary (st.session_state.ndvi_results)
    if st.sidebar.button('Download Mean Index over Parcels to CSV'):
        # Convert st.session_state.ndvi_results to a DataFrame
        # Use st.download_button to download CSV

-------------------------------------------------------------------------------------
# Graphical Visualizations
-------------------------------------------------------------------------------------
if st.sidebar.button('Plot Results'):

        # Use the Plotly library
        # For each feature_id: draw a line of Mean Index vs. Date
        # Possibly compute the integral (e.g., area under the curve) if needed
        # Display results in Streamlit (st.plotly_chart)
    if st.sidebar.button('Plot Subset Results'):
        # Select a subset of dates (e.g., 03/01–05/31)
        # Plot only the index values within that period
        # Display the chart

-------------------------------------------------------------------------------------
# Yield Estimation
-------------------------------------------------------------------------------------

    if st.sidebar.button('Yield Prediction'):
        # Use precomputed areas under the curve (AUC) of the index
```

```
        # or other estimation functions
        # Multiply by coefficients (a-values) depending on the index
        # Create a DataFrame with {Feature ID, Yield, MAPE, etc.}
        # Provide st.download_button for CSV results

-------------------------------------------------------------------------------
# Update Parcels
-------------------------------------------------------------------------------

    if st.sidebar.button('Update parcels with info'):
        # For each parcel:
         #   1) From the dictionary st.session_state.ndvi_results, add properties (Mean-
NDVI_date)
        #   2) From the computed yield, add properties ("Total Yield", "Subset Yield")
        #   3) Create a new GeoJSON
        # Style (folium GeoJson style_function & highlight_function)
        # Export as GeoJSON/KML
        # Add a layer to the map

-------------------------------------------------------------------------------
# Map and Metadata Integration
-------------------------------------------------------------------------------
Map.to_streamlit(height = 600) # Embed the folium/Geemap map in Streamlit
# "Download Metadata" button
if st.sidebar.button('Download Metadata'):

        # Create a text file with session information:
        #   -When the Metadata was created
        #   -Which ROI/Parcels files were uploaded
        #   -Which images were selected (dates, IDs)
        #   -Which downloads took place
        #   -Possible yield estimations
        #   -(NEW) Optionally record if apply_cloud_mask was True and the thresholds
used
        # Use st.download_button to output "metadata.txt"
```

References

1. Aslan, M.F.; Sabancı, K.; Aslan, B. Artificial Intelligence Techniques in Crop Yield Estimation Based on Sentinel-2 Data: A Comprehensive Survey. *Sustainability* **2024**, *16*, 8277. [CrossRef]
2. Teste, F.; Gangloff, H.; Chen, M.; Ciais, P.; Makowski, D. Leveraging satellite data with machine and deep learning techniques for corn yield and price forecasting. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–16. [CrossRef]
3. Marshall, M.S.; Belgiu, M.; Boschetti, M.; Pepe, M.; Stein, A.; Nelson, A. Field-level crop yield estimation with PRISMA and Sentinel-2. *ISPRS J. Photogramm. Remote Sens.* **2022**, *187*, 191–210. [CrossRef]
4. Chen, D.; Hu, H.; Liao, C.; Ye, J.; Bao, W.; Mo, J.; Wu, Y.; Dong, T.; Fan, H.; Pei, J. Crop NDVI time series construction by fusing Sentinel-1, Sentinel-2, and environmental data with an ensemble-based framework. *Comput. Electron. Agric.* **2023**, *215*, 108388. [CrossRef]
5. Farmonov, N.; Amankulova, K.; Szatmári, J.; Urinov, J.; Narmanov, Z.; Nosirov, J.; Mucsi, L. Combining PlanetScope and Sentinel-2 images with environmental data for improved wheat yield estimation. *Int. J. Digit. Earth* **2023**, *16*, 847–867. [CrossRef]
6. Wu, B.; Zhang, M.; Zeng, H.; Tian, F.; Potgieter, A.; Qin, X.; Yan, N.; Chang, S.; Zhao, Y.; Dong, Q.; et al. Challenges and opportunities in remote sensing-based crop monitoring: A review. *Natl. Sci. Rev.* **2022**, *10*, nwac290. [CrossRef]
7. Skakun, S.; Kalecinski, N.I.; Brown, M.G.L.; Johnson, D.M.; Vermote, E.; Roger, J.-C.; Franch, B. Assessing within-Field Corn and Soybean Yield Variability from WorldView-3, Planet, Sentinel-2, and Landsat 8 Satellite Imagery. *Remote Sens.* **2021**, *13*, 872. [CrossRef]

8. Dhillon, M.S.; Kübert-Flock, C.; Dahms, T.; Rummler, T.; Arnault, J.; Steffan-Dewenter, I.; Ullmann, T. Evaluation of MODIS, Landsat 8 and Sentinel-2 Data for Accurate Crop Yield Predictions: A Case Study Using STARFM NDVI in Bavaria, Germany. *Remote Sens.* **2023**, *15*, 1830. [CrossRef]
9. Li, F.; Miao, Y.; Chen, X.; Sun, Z.; Stueve, K.M.; Yuan, F. In-Season Prediction of Corn Grain Yield through PlanetScope and Sentinel-2 Images. *Agronomy* **2022**, *12*, 3176. [CrossRef]
10. Amankulova, K.; Farmonov, N.; Omonov, K.; Abdurakhimova, M.; Mucsi, L. Integrating the Sentinel-1, Sentinel-2 and Topographic data into soybean yield modelling using Machine Learning. *Adv. Space Res.* **2024**, *73*, 4052–4066. [CrossRef]
11. Ghazaryan, G.; Skakun, S.; König, S.; Rezaei, E.E.; Siebert, S.; Dubovyk, O. Crop Yield Estimation Using Multi-Source Satellite Image Series and Deep Learning. In Proceedings of the International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020. [CrossRef]
12. Torino, M.S.; Ortiz, B.V.; Fulton, J.P.; Balkcom, K.S.; Wood, C.W. Evaluation of vegetation indices for early assessment of corn status and yield potential in the southeastern United States. *Agron. J.* **2014**, *106*, 1389–1401. [CrossRef]
13. Bognár, P.; Kern, A.; Pásztor, S.; Steinbach, P.; Lichtenberger, J. Testing the robust yield estimation method for winter wheat, corn, rapeseed, and sunflower with different vegetation indices and meteorological data. *Remote Sens.* **2022**, *14*, 2860. [CrossRef]
14. Marino, S.; Alvino, A. Vegetation indices data clustering for dynamic monitoring and classification of wheat yield crop traits. *Remote Sens.* **2021**, *13*, 541. [CrossRef]
15. Mukiibi, A.; Machakaire, A.T.B.; Franke, A.C.; Steyn, J.M. A Systematic Review of Vegetation Indices for Potato Growth Monitoring and Tuber Yield Prediction from Remote Sensing. *Potato Res.* **2024**, *67*, 1–40. [CrossRef]
16. İrik, H.A.; Ropelewska, E.; Çetin, N. Using spectral vegetation indices and machine learning models for predicting the yield of sugar beet (*Beta vulgaris* L.) under different irrigation treatments. *Comput. Electron. Agric.* **2024**, *221*, 109019. [CrossRef]
17. Ma, Y.; Ma, L.; Zhang, Q.; Huang, C.; Yi, X.; Chen, X.; Hou, T.; Lv, X.; Zhang, Z. Cotton yield estimation based on vegetation indices and texture features derived from RGB image. *Front. Plant Sci.* **2022**, *13*, 925986. [CrossRef]
18. Ruwanpathirana, P.P.; Sakai, K.; Jayasinghe, G.Y.; Nakandakari, T.; Yuge, K.; Wijekoon, W.M.C.J.; Priyankara, A.C.P.; Samaraweera, M.D.S.; Madushanka, P.L.A. Evaluation of Sugarcane Crop Growth Monitoring Using Vegetation Indices Derived from RGB-Based UAV Images and Machine Learning Models. *Agronomy* **2024**, *14*, 2059. [CrossRef]
19. Fu, Z.; Jiang, J.; Gao, Y.; Krienke, B.; Wang, M.; Zhong, K.; Cao, Q.; Tian, Y.; Zhu, Y.; Cao, W.; et al. Wheat growth monitoring and yield estimation based on multi-rotor unmanned aerial vehicle. *Remote Sens.* **2020**, *12*, 508. [CrossRef]
20. Kayad, A.; Sozzi, M.; Gatto, S.; Marinello, F.; Pirotti, F. Monitoring within-field variability of corn yield using Sentinel-2 and machine learning techniques. *Remote Sens.* **2019**, *11*, 2873. [CrossRef]
21. Vidican, R.; Mălinaș, A.; Ranta, O.; Moldovan, C.; Marian, O.; Ghețe, A.; Ghișe, C.R.; Popovici, F.; Cătunescu, G.M. Using remote sensing vegetation indices for the discrimination and monitoring of agricultural crops: A critical review. *Agronomy* **2023**, *13*, 3040. [CrossRef]
22. Van Diepen, C.A.; Wolf, J.V.; Van Keulen, H.; Rappoldt, C. WOFOST: A simulation model of crop production. *Soil Use Manag.* **1989**, *5*, 16–24. [CrossRef]
23. Ren, S.; Chen, H.; Hou, J.; Zhao, P.; Dong, Q.; Feng, H. Based on historical weather data to predict summer field-scale maize yield: Assimilation of remote sensing data to WOFOST model by ensemble Kalman filter algorithm. *Comput. Electron. Agric.* **2024**, *219*, 108822. [CrossRef]
24. Liu, L.; Wallach, D.; Li, J.; Liu, B.; Zhang, L.; Tang, L.; Zhang, Y.; Qiu, X.; Cao, W.; Zhu, Y. Uncertainty in wheat phenology simulation induced by cultivar parameterization under climate warming. *Eur. J. Agron.* **2018**, *94*, 46–53. [CrossRef]
25. Bergjord, A.K.; Bonesmo, H.; Skjelvåg, A.O. Modelling the course of frost tolerance in winter wheat: I. Model development. *Eur. J. Agron.* **2008**, *28*, 321–330. [CrossRef]
26. Chen, J.; Zhang, P.; Liu, J.; Deng, J.; Su, W.; Wang, P.; Li, Y. Study on the impact of low-temperature stress on winter wheat based on multi-model coupling. *Food Energy Secur.* **2024**, *13*, e543. [CrossRef]
27. Saltelli, A.; Tarantola, S.; Chan, K.S. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics* **1999**, *41*, 39–56. [CrossRef]
28. Li, X.; Tan, J.; Li, H.; Wang, L.; Niu, G.; Wang, X. Sensitivity Analysis of the WOFOST Crop Model Parameters Using the EFAST Method and Verification of Its Adaptability in the Yellow River Irrigation Area, Northwest China. *Agronomy* **2023**, *13*, 2294. [CrossRef]
29. Ren, Y.; Li, Q.; Du, X.; Zhang, Y.; Wang, H.; Shi, G.; Wei, M. Analysis of corn yield prediction potential at various growth phases using a process-based model and deep learning. *Plants* **2023**, *12*, 446. [CrossRef]
30. Zhuo, W.; Huang, H.; Gao, X.; Li, X.; Huang, J. An improved approach of winter wheat yield estimation by jointly assimilating remotely sensed leaf area index and soil moisture into the Wofost model. *Remote Sens.* **2023**, *15*, 1825. [CrossRef]
31. Wu, Y.; Xu, W.; Huang, H.; Huang, J. Bayesian posterior-based winter wheat yield estimation at the field scale through assimilation of Sentinel-2 data into WOFOST model. *Remote Sens.* **2022**, *14*, 3727. [CrossRef]

32. Xu, L.; Liu, H.; Jiang, L.; Zhang, F.; Li, X.; Feng, X.; Huang, J.; Bai, T. WOFOST-N: An improved WOFOST model with nitrogen module for simulation of Korla Fragrant pear tree growth and nitrogen dynamics. *Comput. Electron. Agric.* **2024**, *220*, 108860. [CrossRef]
33. Gavasso-Rita, Y.L.; Papalexiou, S.M.; Li, Y.; Elshorbagy, A.; Li, Z.; Schuster-Wallace, C. Crop models and their use in assessing crop production and food security: A review. *Food Energy Secur.* **2024**, *13*, e503. [CrossRef]
34. Ji, F.; Meng, J.; Cheng, Z.; Fang, H.; Wang, Y. Crop yield estimation at field scales by assimilating time series of Sentinel-2 data into a modified CASA-WOFOST coupled model. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–14. [CrossRef]
35. Wang, Y.; Zhang, Q.; Yu, F.; Zhang, N.; Zhang, X.; Li, Y.; Wang, M.; Zhang, J. Progress in Research on Deep Learning-Based Crop Yield Prediction. *Agronomy* **2024**, *14*, 2264. [CrossRef]
36. Lokesh, S.; Madhavan, A.; Ramanathan, R.P.; Anand, K. Intelligent Systems for Data Driven Agriculture: Enhancing Farmer Productivity Through Automation and Artificial Intelligence. In Proceedings of the 2024 International Conference on Smart Systems for Electrical, Electronics, Communication and Computer Engineering (ICSSEECC), Coimbatore, India, 28–29 June 2024; IEEE: Piscataway, NJ, USA; pp. 433–438. [CrossRef]
37. Qi, H.; Qian, X.; Shang, S.; Wan, H. Multi-year mapping of cropping systems in regions with smallholder farms from Sentinel-2 images in Google Earth engine. *GISci. Remote Sens.* **2024**, *61*, 2309843. [CrossRef]
38. Hnatushenko, V.V.; Sierikova, K.Y.; Sierikov, I.Y. Development of a cloud-based web geospatial information system for agricultural monitoring using Sentinel-2 data. In Proceedings of the 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), Lviv, Ukraine, 11–14 September 2018; IEEE: Piscataway, NJ, USA; Volume 1, pp. 270–273. [CrossRef]
39. Pandit, A.; Sawant, S.A.; Agrawal, R.; Mohite, J.D.; Pappula, S. Development of Automated Satellite Data Downloading and Processing Pipeline on Aws Cloud for Near-Real-Time Agriculture Applications. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *4*, 189–196. [CrossRef]
40. Kpienbaareh, D.; Kansanga, M.; Luginaah, I. Examining the potential of open source remote sensing for building effective decision support systems for precision agriculture in resource-poor settings. *GeoJournal* **2019**, *84*, 1481–1497. [CrossRef]
41. Stratoulias, D.; Tolpekin, V.; De By, R.A.; Zurita-Milla, R.; Retsios, V.; Bijker, W.; Hasan, M.A.; Vermote, E. A workflow for automated satellite image processing: From raw VHSR data to object-based spectral information for smallholder agriculture. *Remote Sens.* **2017**, *9*, 1048. [CrossRef]
42. Bontemps, S.; Arias, M.; Cara, C.; Dedieu, G.; Guzzonato, E.; Hagolle, O.; Inglada, J.; Morin, D.; Rabaute, T.; Savinaud, M.; et al. "Sentinel-2 for agriculture": Supporting global agriculture monitoring. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; IEEE: Piscataway, NJ, USA; pp. 4185–4188. [CrossRef]
43. Brook, A.; Micco, V.D.; Battipaglia, G.; Erbaggio, A.; Ludeno, G.; Catapano, I.; Bonfante, A. A Smart Multi-scale and Multi-temporal System to Support Precision and Sustainable Agriculture from Satellite Images. *Proceedings* **2019**, *30*, 17. [CrossRef]
44. Valero, S.; Morin, D.; Inglada, J.; Sepulcre, G.; Arias, M.; Hagolle, O.; Dedieu, G.; Bontemps, S.; Defourny, P. Processing Sentinel-2 image time series for developing a real-time cropland mask. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; IEEE: Piscataway, NJ, USA; pp. 2731–2734. [CrossRef]
45. CopernicusEU. Copernicus Data Space Ecosystem. 2024. Available online: https://dataspace.copernicus.eu/ (accessed on 23 December 2024).
46. Lancashire, P.D.; Bleiholder, H.; Boom, T.V.D.; Langelüddeke, P.; Stauss, R.; Weber, E.; Witzenberger, A. A uniform decimal code for growth stages of crops and weeds. *Ann. Appl. Biol.* **1991**, *119*, 561–601. [CrossRef]
47. Djamai, N.; Fernandes, R. Comparison of SNAP-derived Sentinel-2A L2A product to ESA product over Europe. *Remote Sens.* **2018**, *10*, 926. [CrossRef]
48. QGIS Development Team. QGIS Geographic Information System. QGIS Association. 2024. Available online: http://www.qgis.org (accessed on 23 December 2024).
49. Sanner, M.F. Python: A programming language for software integration and development. *J. Mol. Graph. Model.* **1999**, *17*, 57–61.
50. Jupyter Team. Jupyter Notebook Documentation. 2024. Available online: https://jupyter-notebook.readthedocs.io/en/v7.0.6/index.html (accessed on 23 December 2024).
51. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]
52. Khorasani, M.; Abdou, M.; Fernández, J.H. *Web Application Development with Streamlit*; Software Development; Apress: Berkeley, CA, USA, 2022; pp. 105–128. [CrossRef]
53. Rouse, J.W., Jr.; Haas, R.H.; Schell, J.A.; Deering, D.W. *Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation (No. NASA-CR-132982)*; NASA: Washington, DC, USA, 1973.
54. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [CrossRef]

55. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [CrossRef]
56. Fang, H.; Baret, F.; Plummer, S.; Schaepman-Strub, G. An overview of global leaf area index (LAI): Methods, products, validation, and applications. *Rev. Geophys.* **2019**, *57*, 739–799. [CrossRef]
57. Boegh, E.; Soegaard, H.; Broge, N.; Hasager, C.B.; Jensen, N.O.; Schelde, K.; Thomsen, A. Airborne multispectral data for quantifying leaf area index, nitrogen concentration, and photosynthetic efficiency in agriculture. *Remote Sens. Environ.* **2002**, *81*, 179–193. [CrossRef]
58. Gitelson, A.A.; Zygielbaum, A.I.; Arkebauer, T.J.; Walter-Shea, E.A.; Solovchenko, A. Stress detection in vegetation based on remotely sensed light absorption coefficient. *Int. J. Remote Sens.* **2024**, *45*, 259–277. [CrossRef]
59. Croft, H.; Chen, J.M.; Zhang, Y. The applicability of empirical vegetation indices for determining leaf chlorophyll content over different leaf and canopy structures. *Ecol. Complex.* **2014**, *17*, 119–130. [CrossRef]
60. Karmakar, P.; Teng, S.W.; Murshed, M.; Pang, S.; Li, Y.; Lin, H. Crop monitoring by multimodal remote sensing: A review. *Remote Sens. Appl. Soc. Environ.* **2024**, *33*, 101093. [CrossRef]
61. Zsebő, S.; Bede, L.; Kukorelli, G.; Kulmány, I.M.; Milics, G.; Stencinger, D.; Teschner, G.; Varga, Z.; Vona, V.; Kovács, A.J. Yield Prediction Using NDVI Values from GreenSeeker and MicaSense Cameras at Different Stages of Winter Wheat Phenology. *Drones* **2024**, *8*, 88. [CrossRef]
62. Kaya, Y.; Polat, N. A linear approach for wheat yield prediction by using different spectral vegetation indices. *Int. J. Eng. Geosci.* **2023**, *8*, 52–62. [CrossRef]
63. Wang, Z.; Zhang, C.; Gao, L.; Fan, C.; Xu, X.; Zhang, F.; Zhou, Y.; Niu, F.; Li, Z. Time Phase Selection and Accuracy Analysis for Predicting Winter Wheat Yield Based on Time Series Vegetation Index. *Remote Sens.* **2024**, *16*, 1995. [CrossRef]
64. Badagliacca, G.; Messina, G.; Praticò, S.; Lo Presti, E.; Preiti, G.; Monti, M.; Modica, G. Multispectral vegetation indices and machine learning approaches for durum wheat (*Triticum durum* Desf.) Yield Prediction across Different Varieties. *AgriEngineering* **2023**, *5*, 2032–2048. [CrossRef]
65. Cavalaris, C.; Megoudi, S.; Maxouri, M.; Anatolitis, K.; Sifakis, M.; Levizou, E.; Kyparissis, A. Modeling of durum wheat yield based on sentinel-2 imagery. *Agronomy* **2021**, *11*, 1486. [CrossRef]
66. Devkota, K.P.; Bouasria, A.; Devkota, M.; Nangia, V. Predicting wheat yield gap and its determinants combining remote sensing, machine learning, and survey approaches in rainfed Mediterranean regions of Morocco. *Eur. J. Agron.* **2024**, *158*, 127195. [CrossRef]
67. Nadkarni, P.M.; Ohno-Machado, L.; Chapman, W.W. Natural language processing: An introduction. *J. Am. Med. Inform. Assoc.* **2011**, *18*, 544–551. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.