

Artificial Intelligence Techniques in Crop Yield Estimation Based on Sentinel-2 Data: A Comprehensive Survey

Muhammet Fatih Aslan ^{1,*} , Kadir Sabanci ¹  and Busra Aslan ²

¹ Faculty of Engineering, Department of Electrical and Electronics Engineering, Karamanoglu Mehmetbey University, Karaman 70100, Türkiye; kadirSabanci@kmu.edu.tr

² Graduate School of Natural and Applied Sciences, Department of Mechatronics Engineering, Karamanoglu Mehmetbey University, Karaman 70100, Türkiye; busraaslan@kmu.edu.tr

* Correspondence: mfatihhaslan@kmu.edu.tr

Abstract: This review explores the integration of Artificial Intelligence (AI) with Sentinel-2 satellite data in the context of precision agriculture, specifically for crop yield estimation. The rapid advancements in remote sensing technology, particularly through Sentinel-2's high-resolution multispectral imagery, have transformed agricultural monitoring by providing critical data on plant health, soil moisture, and growth patterns. By leveraging Vegetation Indices (VIs) derived from these images, AI algorithms, including Machine Learning (ML) and Deep Learning (DL) models, can now predict crop yields with high accuracy. This paper reviews studies from the past five years that utilize Sentinel-2 and AI techniques to estimate yields for crops like wheat, maize, rice, and others. Various AI approaches are discussed, including Random Forests, Support Vector Machines (SVM), Convolutional Neural Networks (CNNs), and ensemble methods, all contributing to refined yield forecasts. The review identifies a notable gap in the standardization of methodologies, with researchers using different VIs and AI techniques for similar crops, leading to varied results. As such, this study emphasizes the need for comprehensive comparisons and more consistent methodologies in future research. The work underscores the significant role of Sentinel-2 and AI in advancing precision agriculture, offering valuable insights for future studies that aim to enhance sustainability and efficiency in crop management through advanced predictive models.



Citation: Aslan, M.F.; Sabanci, K.; Aslan, B. Artificial Intelligence Techniques in Crop Yield Estimation Based on Sentinel-2 Data: A Comprehensive Survey. *Sustainability* **2024**, *16*, 8277. <https://doi.org/10.3390/su16188277>

Academic Editors: Minxuan Lan, Song Shu and Bo Yang

Received: 21 August 2024

Revised: 17 September 2024

Accepted: 19 September 2024

Published: 23 September 2024



Copyright: © 2024 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/>).

Keywords: AI; crop yield estimation; precision agriculture; Sentinel-2; VI

1. Introduction

As the global human population continues to rise, the demand for food and agricultural products is also increasing at a comparable rate. However, this surge in demand exerts additional pressure on natural resources. To address this challenge, precision agriculture (PA) has emerged as a viable solution. The idea of precision agriculture has a relatively long history, starting in 1988. PA involves the optimization of agricultural inputs to enhance crop production while minimizing losses [1]. The growing need for increased crop yields has raised concerns about the production of safe and healthy food. On a global scale, the monitoring of food products by national governments and consumers is becoming increasingly difficult. Consequently, traceability and verification have recently gained significant importance in ensuring safe agricultural practices, with a strong demand for spatial information to achieve these goals [2]. Spatial information plays a crucial role in modern agriculture by enabling precise resource management, optimizing yields, and supporting sustainable practices. It provides detailed data on various factors, such as soil conditions, moisture levels, crop health, and terrain, allowing farmers to make informed decisions. By utilizing spatial information, farmers can implement site-specific management practices, reduce waste, and increase productivity. This data-driven approach contributes to enhanced agricultural productivity and sustainability by facilitating efficient

water use, proper application of fertilizers and pesticides, and timely interventions for pest and disease control [3].

2. Remote Sensing and Technologies in PA

Remote sensing (RS), a key technology in acquiring spatial information, has recently revolutionized agricultural monitoring and management. RS involves the detection and monitoring of an area's physical characteristics through reflection or imaging without direct contact. Measurements, estimations, and other activities conducted without direct physical interaction fall within the scope of RS [4]. RS has become increasingly significant and practical, particularly with advancements in satellite and aircraft technology, and provides accurate and timely data on crop conditions, growth patterns, and environmental changes [5]. It allows for continuous monitoring of extensive agricultural areas, offering insights into crop health, phenology, and stress factors. RS also facilitates the mapping of crop patterns, detection of anomalies, and assessment of the impacts of climate change on agriculture [6]. This technology supports PA by enabling farmers to optimize their practices, reduce costs, and minimize environmental impacts, thereby ensuring food security and promoting sustainable agricultural development.

In agriculture, RS technologies such as aerial vehicles (airplanes, helicopters, Unmanned Aerial Vehicles (UAVs)), sensors, and satellites play a crucial role in modern farming practices by providing detailed and timely information on crop health, soil conditions, and field variability [7]. Over the past 10 years, UAVs, commonly known as drones, have provided revolutionary benefits to agriculture [8]. Equipped with cameras and sensors that capture high-resolution images and data from above, they allow farmers to monitor crop growth, detect diseases, and optimize irrigation [9]. Ground-based sensors (ex. soil moisture sensors, soil temperature sensors, soil nutrient sensors, etc.) measure soil moisture, temperature, and nutrient levels, providing real-time data that supports precise irrigation and fertilization. Satellites, with their extensive coverage and advanced imaging capabilities, facilitate large-scale monitoring and analysis of agricultural lands, assisting in tracking crop development, assessing damage from pests or weather events, and predicting yields [10,11]. By leveraging data from these technologies, farmers can make informed decisions that enhance productivity, reduce resource use, and minimize environmental impact. The choice of technology for data collection depends on the specific application. For instance, satellite imagery can identify underperforming areas in a field, encouraging targeted interventions, while UAVs can conduct detailed inspections of specific crops. Ground sensors contribute to PA by providing accurate, site-specific information that optimizes input applications [12].

RS technologies provide a wide range of valuable data for agriculture, and the interpretation of these data is of significant importance. Traditional techniques used in data interpretation primarily rely on manual and semi-automated methods to analyze satellite and aerial imagery. These techniques include visual interpretation, where experts analyze images to identify crop types, assess their health, and predict yields based on color, texture, and patterns. Other common methods involve the use of Vegetation Indices (VI), such as the Normalized Difference Vegetation Index (NDVI), which measures plant health by comparing the reflectance of red and near-infrared light, as well as statistical analysis. Ground truthing, which involves collecting real-world data to validate remote sensing information, is a critical component of these traditional approaches. While effective, these methods can be time-consuming and labor-intensive and may be limited by the subjective nature of human interpretation and the complexity of data analysis [7]. However, Artificial Intelligence (AI) techniques, particularly Machine Learning (ML) and Deep Learning (DL), have recently had a significant impact on the evaluation of RS data. AI is a broad field of science that enables machines to perform tasks with human-like intelligence. ML, as a sub-branch of AI, includes methods that allow machines to learn from data. In this process, machines learn from examples and make predictions without writing specific rules. DL, on the other hand, is a sub-branch of ML and works especially with large datasets and multi-layered

neural networks. DL has the capacity to learn more complex data patterns using structures similar to how neurons in the human brain work. These three areas are interconnected; AI provides a general framework, while ML represents the learning process of this framework, and DL represents the deeper and more complex learning structure. DL, in particular, is capable of predicting complex relationships between environmental parameters through its advanced learning capabilities [13,14]. AI techniques enable the analysis of complex datasets from various sources, such as satellite imagery, drones, and sensors, facilitating precision crop monitoring, yield prediction, and disease detection. These methods can identify patterns and anomalies that traditional approaches may overlook, leading to more informed decision-making and optimized resource management. Additionally, AI models can continuously learn and improve from new data, enhancing their predictive capabilities and adaptability to changing agricultural conditions [15].

3. Yield Estimation and AI in PA

The success and impartiality of AI in interpreting agricultural data have recently made it a focal point for researchers, particularly in the area of crop yield estimation and forecasting. Yield estimation is a critical aspect of agriculture for farmers and agricultural planners alike. Accurate yield forecasts enable farmers to make informed decisions regarding resource allocation, crop management, and market strategies. By predicting potential crop output, farmers can optimize the use of fertilizers, water, and pesticides, which not only enhances productivity but also minimizes environmental impact. Additionally, yield prediction assists in financial planning, allowing farmers to estimate income and manage risks associated with crop failures or market fluctuations [16]. This foresight contributes to ensuring food security and stability within the agricultural sector. In this context, many countries around the world are making significant investments to obtain yield data [17]. The use of RS technologies to predict field and yield variability is becoming increasingly prevalent in PA due to their relatively low costs and non-invasive approaches [18]. AI applications can provide precise and real-time yield predictions by utilizing large datasets from various sources, such as satellite imagery, soil conditions, and crop health indicators. ML algorithms can analyze these datasets to identify patterns and correlations that traditional methods may overlook.

Recent studies in the literature have demonstrated that satellite data can be utilized for yield estimation at both the field and farm scales. Earth observation data obtained through satellites, with high spatial (~10 m) and temporal (every 5 days) resolutions, allow for effective monitoring of crop growth on a regional scale Franch, et al. [19]. In this context, Sentinel satellites (Sentinel-2A (2015), Sentinel-2B (2017)) have been frequently studied by researchers in yield prediction applications. These satellites were launched under the Copernicus Program to obtain high spatial resolution optical images. Due to their high-resolution multispectral imaging capabilities, they have recently become vital tools in predicting crop yield at the field level. The satellites' ability to capture data across 13 spectral bands, ranging from visible to shortwave infrared, allows for detailed analysis of vegetation health, soil properties, and water content [20]. These attributes are crucial for accurately monitoring crop growth stages, detecting stress factors, and forecasting yield. The high temporal resolution, with a revisit time of approximately five days, enables the prompt detection of changes in crop conditions, allowing for timely interventions. This frequent monitoring aids farmers in optimizing resource use, improving agricultural practices, and ultimately enhancing food security [21].

4. Sentinel-2 in Yield Estimation

Sentinel-2 stands out in crop yield estimation due to its superior combination of spatial, spectral, and temporal resolution compared to other satellite systems such as MODIS, Landsat, and EnMAP. While MODIS offers frequent revisit times, its coarse spatial resolution (250–500 m) limits its effectiveness for detailed field-level analysis. Landsat provides good spatial resolution (30 m) but has a longer 16-day revisit cycle, making it less effective for

timely monitoring. EnMAP, with its advanced hyperspectral imaging capabilities, offers rich spectral data but is constrained by lower spatial resolution and longer revisit times. In contrast, Sentinel-2 offers high spatial resolution (10 m), a comprehensive range of 13 spectral bands, and a frequent five-day revisit period, enabling precise, detailed, and timely crop monitoring. These features make Sentinel-2 particularly advantageous for PA, as it allows for accurate assessment of crop health and yield prediction [22,23].

In the literature, Sentinel-2 satellite data have been widely used in recent years to improve crop yield prediction models. Studies have demonstrated the effectiveness of Sentinel-2 data in mapping crop types, assessing biomass, and predicting yields for various crops, including wheat, corn, and rice. For instance, researchers have used Sentinel-2 data to develop ML models that incorporate spectral indices such as NDVI and EVI (Enhanced Vegetation Index) to predict yields with high accuracy [24] (For a more complete list, see Appendix A or <https://www.indexdatabase.de/> (accessed on 22 September 2024)). These vegetation indices (VIs) provide quantitative measures of crop health, vigor, and stress—key indicators that correlate directly with potential yield.

However, the methods for utilizing Sentinel-2 data can vary. Some studies rely solely on VIs such as NDVI and EVI, which are standard metrics for monitoring vegetation health. Others incorporate additional indices, such as the Normalized Difference Red Edge Index (NDRE) and the Canopy Chlorophyll Content Index (CCCI), which offer more nuanced insights into crop conditions, particularly in terms of chlorophyll content and photosynthetic activity. These indices, while related, measure different aspects of crop performance, which can lead to variations in yield predictions depending on the crops being analyzed and the agricultural practices involved. The high-resolution, multispectral capabilities of Sentinel-2 allow for the calculation of these indices, making it possible to track changes in crop conditions over time. This enables the detection of early indicators of crop performance, providing the basis for more accurate yield predictions [25,26]. In comparison to other methods that utilize different RS platforms or rely on ground-based observations alone, Sentinel-2-based models tend to offer higher spatial resolution and more detailed spectral data.

Another key difference lies in the timing and application of these methods. For instance, some studies utilize Sentinel-2 data only during specific growth stages, conducting seasonal analyses, while others adopt a continuous monitoring approach that covers the entire growing cycle, leading to more dynamic models. Seasonal analyses often yield quicker results and are favored when time and resources are limited. In contrast, continuous monitoring provides a more comprehensive dataset by capturing long-term changes in crop conditions. Additionally, the ML models applied also vary between studies. Some employ regression-based models, which are more interpretable and explainable, while others use DL algorithms that can process larger datasets, potentially yielding higher accuracy. These differences in approach are significant as they reflect the varying research goals and available resources, directly influencing the choice of methodology.

Additionally, researchers often integrate Sentinel-2 data with other RS platforms or ground-based observations to create comprehensive agricultural monitoring systems. This integration approach varies between studies, with some focusing on multisource data fusion for enhanced accuracy, while others prioritize ease of implementation and cost-efficiency [27]. These differences highlight the need for a tailored approach to agricultural monitoring based on the specific crops, environments, and resources available. The continued integration of Sentinel-2 data into Precision Agriculture (PA) has played a pivotal role in promoting sustainable and efficient farming practices, leading to improvements in crop management and yield predictions.

5. Previous Review Studies and Contribution of the Study

Given the significance of yield prediction in agriculture and the advantages of AI methods in this area, a substantial body of research has emerged focused on AI-based yield prediction. Numerous review studies have examined these efforts. For instance, Oikono-

midis, et al. [28] provided a comprehensive analysis of the application of DL techniques in crop yield prediction, emphasizing the importance of various data sources such as RS imagery, weather data, and soil information. The authors particularly highlighted the significant advantages of DL models, including Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, due to their ability to capture complex, non-linear relationships within large datasets, thereby improving the accuracy of yield predictions. Similarly, Van Klompenburg, et al. [29] investigated ML algorithms used for crop yield prediction, identifying Random Forests (RF) and neural networks as the most commonly employed methods across 50 analyzed studies. Desloires, Ienco and Botrel [21] also identified RFs, Artificial Neural Networks (ANNs), and ensemble learning approaches as frequently applied ML algorithms in crop yield prediction. Another review by Luo, et al. [30] examined the applications of ML techniques such as Support Vector Machines (SVM), Decision Trees (DT), and neural networks in improving crop yield and nitrogen status prediction, discussing the potential of ML to optimize nitrogen management practices through accurate nitrogen level predictions. Khairunniza-Bejo, et al. [31] explored the use of ANNs in crop yield prediction, highlighting their strengths in processing non-linear and multidimensional data to analyze complex patterns and relationships related to crop growth, environmental conditions, and management practices. Dharani, et al. [32], focused on the application of DL techniques for crop yield prediction, emphasizing the advancements and effectiveness of these methods in agricultural forecasting. They summarized past studies on various DL models, including CNNs and Recurrent Neural Networks (RNNs), used to analyze and predict crop yields based on complex datasets such as satellite imagery and climate data.

The number of studies mentioned in the previous paragraph can be extended. These summarized review studies have addressed AI-supported yield prediction research. However, despite the evident importance and effectiveness of Sentinel-2 in agricultural yield prediction, there has yet to be a review study specifically focused on Sentinel-2. Particularly in the past five years, research involving Sentinel-2-based yield prediction has gained significant momentum. For various crop yield predictions, different vegetation indices (VIs) and ML techniques have been explored by researchers. The following section will extensively cover yield studies conducted over the past five years that utilized Sentinel-2 data with AI models. These studies demonstrate the importance of Sentinel-2 in yield prediction and its compatibility with AI techniques. Given the novelty of these studies, there is a growing need for research that can guide future efforts by addressing questions such as which VIs are used, which AI models are most effective, and for which crops they have been applied. This need arises because, despite similar methodologies across most previous studies, the specific VIs, ML techniques, and application steps often vary, even for the same crops like rice and wheat. This review will broadly examine such past studies, considering the methodological differences and variations in the VIs used. In this sense, this study aims to serve as a guide for future AI-supported yield prediction research utilizing Sentinel-2 data. In summary, the fundamental features of this study are as follows:

- No previous review articles have examined Sentinel-2-based yield prediction studies. Therefore, this study is significant.
- The reviewed studies are recent (2019–2024).
- The crops used in previous yield studies and the types of VIs calculated are presented in a simple overview with a table.

6. Previous Sentinel-2 and Crop Yield Estimation Works

Previous crop yield estimation studies developed using Sentinel-2 data and AI techniques have increased significantly in the last 5 years (2019–2024). These studies calculate various VIs (such as NDVI, EVI, NDRE) that monitor plant health and development processes using Sentinel-2's high-resolution and multispectral imaging capabilities. AI algorithms analyze these data, learn complex and non-linear relationships affecting crop yields, and create prediction models. Studies usually make more precise and localized

predictions by monitoring within-field variability and changes over time, which provides significant advantages in agricultural management and decision-making processes. Such case studies can be explained as follows. (Since there are many abbreviations, especially those originating from VIs, abbreviation explanations are shared in Appendices A and B Section).

6.1. Studies Using Only ML Techniques

ML techniques are commonly used methods in agricultural yield prediction. These techniques provide high accuracy in predicting agricultural yields by learning complex relationships within large datasets. Algorithms frequently used in agriculture, such as RF and SVM, have shown successful results, particularly when combined with satellite imagery and vegetation indices (e.g., NDVI, EVI). By processing agricultural data, ML helps in understanding the relationship between crop growth processes and environmental factors, leading to better decision-making.

Hunt, et al. [33] utilized Sentinel-2 data to map the yield variability within a wheat field at a 10-m resolution over the course of a year. They presented a model calibrated with combined harvester yield monitor data. Wheat yield data were collected during the 2016 harvest season, from 6 August to 9 September, using combines equipped with a Global Positioning System (GPS) and an optical yield monitor. Over 8000 points from 39 wheat fields were used for training and validation. The yield data underwent several preprocessing steps, including resampling to 10 m and 20 m resolutions using an Inverse Distance Weighting function. This resampling allowed the yield to be aligned with the Sentinel-2 data used in the study, facilitating the evaluation of the optimal resolution for yield prediction. The study utilized only the 10 m or 20 m resolution bands of Sentinel-2 (i.e., B2, B3, B4, B5, B6, B7, B8, B8a, B11, and B12). Five different Vis, GCVI, GNDVI, NDVI, SRI, and WDRVI, were calculated using these bands. Yield prediction was performed with various combinations of data, integrating Sentinel-2 data from different periods of the growing season with environmental data (including meteorological, topographic, and soil moisture data). The highest performance, using RF and normalized two-band ratios, occurred during the maturity stage. The results demonstrated that it is possible to produce accurate intra-field yield variability maps at a 10 m resolution using Sentinel-2 data (RMSE = 0.66 t/ha). Accuracy increased when environmental data were incorporated (RMSE = 0.61 t/ha).

Kayad, et al. [34] conducted a study to predict the spatial variability of maize yield. Yield data were monitored using a grain yield monitor mounted on a harvester across a 22-hectare study area in Italy between 2016 and 2018. Sentinel-2-derived VIs and ML techniques (RF, SVM, and MLR) were employed for yield prediction. After all Sentinel-2 bands were resampled to 10 m, nine different VIs (NDVI, NDRE1, NDRE2, GNDVI, GARVI, EVI, WDRVI, mWDRVI, and GCVI) were extracted from the Sentinel-2 imagery. Among these, GNDVI was identified as the best parameter for monitoring the intra-field variability of maize grain yield. Additionally, RF was found to be the most accurate ML technique for maize yield monitoring, with R^2 values reaching up to 0.6.

Gómez, et al. [35] conducted a potato yield prediction study using images from Sentinel-2 satellites over three growing seasons, applying nine different ML models. They focused on the learning models that showed the best performance for the months of July, August, and September, continuing their experiments with these top-performing models. The models were evaluated based on test data from 33 different sites in Spain collected over three years. The minimum yield sample across all three years was 17.547 tons/ha, the maximum was 85.678 tons/ha, and the average potato yield was 57.950 tons/ha with a standard deviation of 16.747. Potatoes exhibiting green areas, measuring less than 28 mm in diameter, showing deformities, or having physical damage were excluded from the harvest and not considered in the crop yield calculations. Field assessments indicated that around 3–5% of the overall production had some form of defect, resulting in those potatoes being left unharvested. A total of 44 Sentinel-2 band images were resampled to 10-m

resolution for the years 2016, 2017, and 2018, with cloud cover removed through masking. Seven different VIs were calculated from these images: ARI2, CRI2, IRECI2, LCC, NDVI, PSRI, and WdVI. The authors emphasized the critical importance of feature selection in the application. Successful predictions were achieved using models such as Regression Quantile Lasso, RF, SVM Radial, and Leap Backwards. After removing data with a high correlation (0.5 or above), the Leap Backwards model achieved RMSE = 10.94%, $R^2 = 0.89$, and MAE = 8.95%. The SVM method, however, provided better predictions without feature selection, with prediction values of RMSE = 11.7%, $R^2 = 0.93$, and MAE = 8.64%.

Zhao, et al. [36] utilized observations from 103 study sites during the 2016 and 2017 harvest seasons in northeastern Australia. They incorporated indices derived from Sentinel-2 (S2) data and a modeled crop water stress index (SI) to predict drought-prone wheat yields at the field scale. To determine wheat yield variability among fields, they employed a total of 14 spectral VIs, including NDVI, OSAVI, EVI, CCCI, SR, and DVI. Predictive models were developed using data from 89 fields, while data from 14 fields were used for performance testing. A linear regression (LR) model was applied to the yield values. The study compared the predicted yields based on vegetation indices with the actual values. The VIs derived from S2 demonstrated moderately high accuracy in yield prediction, explaining more than 70% of yield variability. Specifically, CIred (RMSE = 0.88 t/ha) and OSAVI (RMSE = 0.91 t/ha) provided the best correlation with field yields. Additionally, combining the SI derived from the crop model with both structural and chlorophyll indices significantly improved predictability.

Franch, Bautista, Fita, Rubio, Tarrazó-Serrano, Sánchez, Skakun, Vermote, Becker-Reshef and Uris [19] used Sentinel-2 spectral reflectance bands and VI data to monitor within-field yield variability in a field in Valencia during the 2020 season. Yield measurements were conducted using harvesters equipped with sensors, covering a total of 52 fields and 66 hectares. The authors utilized Sentinel-2 spectral bands with spatial resolutions of 10 m and 20 m (B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12). They combined these spectral bands to achieve the best correlation with yield values. The derived VIs included WDRVI, SAVI, RVI, LSWI, NDBI, TVI, IPVI, RGVI, CIred, NDRE1, and NDRE2. Equation combinations were developed separately for the 10 m and 20 m spatial resolution bands based on LR. The study produced numerous results for different equations at both 10 m and 20 m resolutions. The findings demonstrated a strong relationship between rice yield and spectral bands using various combinations.

Nazir, et al. [37] aimed to measure rice yield at various phenological stages using VIs calculated from 2016 Sentinel-2 time series images. The study focused on 137 plots in the Sheikhpura region of Pakistan, where the average annual rice production is approximately 2.5 million tons. These plots were closely monitored until the harvest period, and rice yield values were measured with precision. The researchers used NDVI, EVI, SAVI, and REP vegetation indices to estimate rice yield. To analyze the relationship between these VIs and yield, they applied the Partial Least Squares Regression (PLSR) technique. The authors reported that the integration of PLSR and VIs successfully predicted rice yield with an RMSE of 0.12 t/ha. Additionally, they identified the late vegetative and flowering stages as the most suitable times for rice yield prediction.

Son, et al. [38] utilized Sentinel-2 images to predict rice yield across 671,772 hectares in Taiwan. They analyzed these images using ML techniques. Crop yield data were obtained from the Taiwan Agricultural Research Institute. Data from the 2019 to 2020 crop seasons were processed using ML algorithms (RF, SVM, ANN). Monthly EVI data were used for rice yield prediction, and the results were validated against the 2019–2020 season data from planting to maturity. The SVM model, which used EVI data, outperformed the other models (RF and ANN) in field-level rice yield prediction, with MAPE values of 4.5% and 3.5% for the two respective seasons. The authors concluded that rice crop yields could be predicted one month before harvest using ML models.

Marshall, et al. [39] compared the performance of PRISMA and Sentinel-2 spectral bands in predicting biomass and yield for maize, rice, soybeans, and wheat. Experimental

analyses were conducted on fields totaling 3800 hectares. The comparison considered key crop development stages, including vegetative, reproductive, and maturity phases. Three data-driven methods were selected: TBVIs, PLSR, and RF. PRISMA and Sentinel-2 data were preprocessed (e.g., outlier removal, gap-filling with a smoothing-spline filter). Sentinel-2 bands 1, 9, and 10, which are necessary for coastal/atmospheric studies, were excluded from the analysis. Band 2 was also excluded due to significant atmospheric scattering. The PRISMA and Sentinel-2 models explained approximately 20% more spectral variability in biomass and yield when using RF compared to TBVI and PLSR.

Crusiol, et al. [40] collected within-field soybean yield data using a yield monitor-equipped harvester across 15 different fields, each over 500 hectares, in Brazil. To monitor these yield data using RS, they utilized Sentinel-2 spectral bands and eight different VIs: BNDVI, GNDVI, NDVI, NDRE, NDII, NDII 2, EVI1, and EVI2. They applied PLSR and SVR algorithms to correlate the actual measurements with the Sentinel-2 observations. A 10-fold cross-validation method was used to evaluate the performance of the PLSR and SVR models. The spectral bands and VIs were fed into these regression models individually and in combination. SVR outperformed PLSR, showing a stronger correlation between observed and predicted yields. While the performance of individual VIs was inferior to that of the Sentinel-2 bands, combining all VI images reduced the prediction error. Moreover, the highest performance was achieved when all VI images were used in conjunction with VIS/NIR/SWIR spectral bands.

Ashourloo, et al. [41] conducted measurements for wheat yield in 2020, sampling across 6285.8 hectares in the Hamedan province of Iran. To monitor wheat yield prediction, they used Sentinel-2 bands and VIs. They employed the B3 (Green), B4 (Red), and B8 (VIR) bands along with various VIs derived from these bands (NDVI, SR, GCVI, GNDVI, WDRVI, DVI, EVI, SAVI, GRR1, NGBDI). Subsequently, they applied feature selection based on correlation coefficients between the bands and VIs, observing that the VIs provided higher correlations than the spectral bands. The most correlated features were also used to remove outliers. In the final step, they used multiple regression models, including KNN, NN, DT, SVR, GPR, RF, LR, and SR, to predict wheat yield. The regression results indicated that the GPR model outperformed the others, with an RMSE of 228.56 kg/ha.

Bebie, et al. [42] conducted an experimental study from 2017 to 2020 across 66 fields in Greece to analyze durum wheat yield. The actual yield data were collected using a yield mapping system on a combine harvester. To relate the actual yield values to RS data, they applied two different models. First, they employed an MLR model based on VIs, specifically EVI and NMDI, derived from Sentinel-2 bands. Secondly, they used Sentinel-2 spectral bands as inputs to ML algorithms, including RF, k-Nearest Neighbors KNN, and BR. The study emphasized the superior predictive accuracy of RF and KNN, with an RMSE of less than 360 kg/ha.

Segarra, et al. [43] collected wheat grain data over three seasons (2017–2019) from eight different fields in the Burgos province of Spain using a combine harvester. To develop a model for predicting these actual yield values, they calculated the LAI derived from radiative transfer models, along with seven different VIs (GNDVI, NDVI, RVI, EVI, TGI, NGRDI, and CVI) derived from Sentinel-2 bands. The ML algorithms used for modeling were RF, SVM, and BR. The results indicated that LAI was more successful in yield prediction than the indices derived from Sentinel-2 bands. Among the algorithms, RF outperformed the others with an R^2 value of 0.89.

Abebe, et al. [44] adopted an approach that combined Sentinel-2 and Landsat 8 data to predict sugarcane (*Saccharum officinarum* L.) yield across a 12,000-hectare area in Ethiopia. They collected actual yield data in 't/ha' units from the Wonji Sugarcane Research and Development Center for the 2016/2017 to 2018/2019 seasons. The authors investigated the performance of integrating Landsat 8 and Sentinel-2 data compared to using Sentinel-2 data alone. To achieve this, they registered the Landsat 8 images to the corresponding Sentinel-2 images using image registration techniques, utilizing similar bands from both sensors (Sentinel-2: B2, B3, B4, B8, B11, B12; Landsat 8: B2, B3, B4, B5, B6, B7). They used

these bands to calculate VIs such as NDVI, EVI, SAVI, MSAVI, SR, GNDVI, and SIRI from both Landsat 8 and Sentinel-2. These data were then input into SVR, MLPNN, and MLR methods to predict sugarcane yield. The performance of these ML methods was compared using ten-fold cross-validation. The authors concluded that the Sentinel-2 and Landsat 8 predictions (RMSE = 12.95 t/ha) outperformed predictions using only Sentinel-2 data (RMSE = 14.71 t/ha).

Bhumiphan, Nontapon, Kaewplang, Srihanu, Koedsin and Huete [17] aimed to predict rubber yield using Sentinel-2 satellite imagery for 213 fields between December 2020 and November 2021. They used VIs such as GSAVI, MSRI, NBR, NDVI, NR, and RVI to estimate monthly rubber yield. These index values were input into LR and MLR models to predict yield. The actual yield data from agricultural areas were obtained from farmers' sales invoices, which provided weight data (in kilograms). The prediction results indicated that the red edge spectral band (band 5) outperformed all other spectral bands and VIs ($R^2 = 0.79$, RMSE = 29.63 kg/ha). The MSRI index was the highest-performing VI with an R^2 of 0.62 and RMSE of 39.25 kg/ha. Combining band 5 and MSRI data further improved the results slightly ($R^2 = 0.8$, RMSE = 29.42 kg/ha).

Faqe Ibrahim, et al. [45] developed a wheat yield prediction model for a region in northeastern Iraq. They divided the total area into 11 plots and collected yield data through field visits. To predict the collected yield values, they created a LR model using VIs derived from Landsat 8 and Sentinel-2 (EVI, NDVI, NDWI, SAVI, SRI, RVI, GRVI, NDRE, CMFI, chlorophyll, LAI). The LR model results highlighted the superiority of Sentinel-2 over Landsat 8. Among the Sentinel-2 indices, LAI exhibited the highest correlation with the actual yield values (RMSE = 0.57).

Amankulova, et al. [46] conducted a study based on Sentinel-2 data to predict sunflower crop yield at the pixel or field level. Yield data were obtained from 20 different sunflower fields in Hungary in 2020. The authors employed the RF algorithm to correlate yield values obtained from a yield monitoring system, which is equipped with a combine harvester, with NDVI index values derived from Sentinel-2 spectral bands. The data from 10 fields were used for training, while the remaining 10 fields were reserved for testing. The authors concluded that the RMSE values ranging between 121.9 and 284.5 kg/ha were promising, indicating that sunflower seed yield could be predicted 3–4 months before harvest. The study identified the optimal period for yield prediction to be between 85 and 105 days into the growing season, which corresponded to the flowering stage of sunflowers.

Nuraeni and Manessa [47] examined the effectiveness of spatial ML methods to monitor tea leaf conditions and estimate crop yield at the Gunung Mas plantation in Bogor, Indonesia. The research utilized Sentinel-2 satellite data, which provided high-resolution multispectral imagery for the study area. ML algorithms (SVM, RF) were applied to process the satellite images and extract VIs, such as NDVI (Normalized Difference Vegetation Index), which are indicators of plant health. These indices were used to predict the tea yield by correlating them with ground truth data collected from the plantation. The study found that using spatial ML with Sentinel-2 imagery could accurately predict tea yields, offering a non-invasive and efficient method for agricultural monitoring. The results demonstrated that integrating RS data with ML techniques holds significant potential for optimizing resource use and improving crop management practices in PA.

Madugundu, et al. [48] investigated the best times for monitoring carrot crops and assessing yield using Sentinel-2 satellite data. The research employed various ML models to analyze the spectral and temporal data provided by Sentinel-2 imagery. By identifying key growth stages of the carrot crop, the study aimed to determine the optimal periods for collecting data that would most accurately predict yield outcomes. Sentinel-2's high-resolution images were processed to calculate VIs such as the NDVI, GNDVI, RDVI, GLI, and SIPI which were correlated with field-based yield data. The RF model was trained and validated using this combined satellite and ground data. Five scenarios adopting the RF algorithm were conducted. Among the five scenarios examined, the algorithm with the

highest predictive accuracy was achieved through the combination of individual S2 bands, VIs, and SPAD values (RMSE = 7.8 t/ha).

Kamenova, et al. [49] focused on mapping crop types and predicting winter wheat yield in the Upper Thracian Lowland, Bulgaria, using Sentinel-2 satellite imagery. The primary crops in this region include winter wheat, rapeseed, sunflower, and maize. Researchers employed ML techniques, specifically SVM and RF, to classify crop types accurately. They created temporal image composites to identify the best times to distinguish between different crops. Ground truth data from the Integrated Administration and Control System (IACS) were used to train the classifiers and validate the accuracy of the crop maps. The study found that both SVM and RF classifiers performed well, with SVM showing a slight edge in accuracy. The researchers masked winter wheat fields using the most accurate classification algorithm and then predicted yields using regression models calibrated with in situ data. The GNDVI from the April composite image emerged as the best predictor of winter wheat yield. This approach demonstrated the potential of combining Sentinel-2 data with ML techniques to enhance agricultural monitoring and yield prediction.

de Freitas, et al. [50] thoroughly investigated the potential of using texture measures derived from Sentinel-2 imagery to predict soybean yield variability. The researchers employed an RF model, leveraging unique Grey Level Co-occurrence Matrix (GLCM) texture measures as an alternative to traditional empirical models. They meticulously evaluated eleven different GLCM texture models, each based on eight texture measures of a single spectral layer, to represent soybean field yield variation across two distinct sites and seasons. The findings revealed that several models achieved high accuracy, with R^2 values ranging from 0.90 to 0.95 and RMSE values between 0.06 and 0.26 tons per hectare. The study highlighted that models with window sizes larger than 15 pixels were particularly effective for predicting soybean yield, as the window size significantly influenced the performance of the GLCM texture measures. Furthermore, the research indicated that models derived from individual spectral bands (EVI, GNDVI, GRNDVI, NDMI, NDRE, NDVI, SFDVI), such as red, red-edge, near-infrared, and short wavelength infrared, were more sensitive to changes in window size compared to those derived from vegetation indices. Ultimately, the study concluded that aggregating data using texture measures enhanced the predictive power of individual spectral responses, offering a viable and improved method for predicting within-field soybean yield variation using RF models. This approach provided a promising alternative to traditional methods, potentially leading to more accurate and efficient agricultural yield predictions.

6.2. Studies Using DL Techniques

DL is an artificial intelligence technique particularly successful with complex and multidimensional datasets. In agricultural yield prediction, DL models like CNN and RNN provide high accuracy, especially by processing satellite imagery and time series data. DL models have the capacity to learn complex and non-linear relationships in large datasets, making them a powerful tool for better understanding agricultural processes and predicting future crop yields.

Fernandez-Beltran, et al. [51] utilized Sentinel-2 data to predict rice crop yield. They developed a large-scale rice crop database for Nepal (RicePAL), comprising Sentinel-2 data and actual rice yield records from 2016 to 2018. For yield prediction, a 3D CNN was developed. The study analyzed the impact of using different temporal and climatic/soil data settings with the proposed approach. The Sentinel-2 data included four bands resampled to 20 m (B02–B04 and B08) and bands with a nominal spatial resolution of 20 m (B05–B07, B8A, B11, and B12). NDVI images derived from these bands were used. Given the significant influence of climatic and soil variables on rice crop production, these data were also considered. The proposed CNN model was compared with different ML methods (LR, RID, SVR, GPR) across four different scenarios. The results demonstrated that the proposed CNN consistently outperformed other methods across all tested scenarios. Furthermore, the

incorporation of auxiliary climatic and soil data were found to enhance the yield prediction process.

Narin, et al. [52] conducted a study to monitor sunflower yield values. In 2018, they collected yield data from 48 sunflower parcels in Tokat, Türkiye, based on information obtained from farmers. They correlated these data with NDVI and NDVI_{red} indices derived from Sentinel-2 bands. To predict actual yield values based on the VIs, they employed three different learning algorithms. For the first algorithm, they created LR functions. The second and third algorithms involved implementing ANN and 1D-CNN models. The results indicated that the 1D-CNN model, fed with NDVI data, outperformed the other algorithms, achieving an RMSE value of 20.874 kg/da.

Perich, et al. [53] aimed to model crop yield using AI-based techniques by utilizing five years of small grain crop yield data (2017–2021) collected from combine harvesters. These data, representing various crops over five years, was obtained from farms in Western Switzerland. High-resolution Sentinel-2 time series data were used for modeling crop yields at a spatial scale corresponding to the Sentinel-2-pixel level. Four different methods were applied and compared for modeling crop yield at the Sentinel-2 level, including 'Partial integral at peak' and 'Smoothed NDVI' based on spectral indices, 'Four S2 scenes' based on all spectral bands of Sentinel-2, and pixel-based crop yield modeling using an RNN. The RNN method produced superior results compared to the others, offering advantages in preprocessing and feature extraction while also performing well with cloudy data. When using data from all years, the models achieved higher accuracy. For winter wheat, the 'Four S2 scenes' method achieved an R^2 of 0.88 and an RMSE of 10.49%.

Xiao, et al. [54] conducted an AI-based yield prediction study using winter wheat yield data collected from Henan Province, China, between 2019 and 2020. The data consisted of 2885 field yield samples and 3500 processed Sentinel-2 image scenes. They developed an attention-based algorithm for yield prediction, which was fed with VIs derived from Sentinel-2, including LSWI, IRECI, GCVI, and NDVI. The proposed algorithm design combined a 1D-CNN model with a multi-channel attention mechanism. For comparison, 1D-CNN and RF models were also employed. The developed ACNN model outperformed the others, achieving an RMSE of 1460 kg/ha and an R^2 of 0.44.

Mancini, et al. [55] explored advanced methodologies for predicting durum wheat yield by leveraging Sentinel-2 satellite time series data. The researchers employed Functional Data Analysis (FDA) and DL models, focusing on developing a yield prediction system using both temporal and spectral domains of the data. By analyzing pixels from Sentinel-2 imagery and removing irrelevant data affected by clouds or shadows, they generated continuous spectral time series for each pixel. Functional Principal Component Analysis was applied to create NDVI and NDRE curves, which were then used to predict wheat yield through classical PLSR. Additionally, DL models such as CNNs, including VGG16, VGG19, and MobileNetv2, were tested. The results showed that image-based predictions, particularly using the VGG16 model, achieved the highest accuracy, with a RMSE of 0.047 kg/ha. This study demonstrated the potential of combining spectral indices with DL models for precise crop yield forecasting, thus providing valuable insights for organic durum wheat yield prediction and its applications in PA.

Amankulova, et al. [56] focused on developing an innovative approach to predict soybean yields by combining data from Sentinel-2 and PlanetScope satellites. Researchers collected normalized difference vegetation indices (NDVI, GNDVI, NDRE, EVI, SAVI) data from the S2 Level-2A and PS Level-3 surface reflectance during the soybean growing season. They applied the pyDMS algorithm, a decision tree-based technique, to enhance both low and high-resolution images with information from large-scale images. The study analyzed the robustness and flexibility of a multidimensional data fusion, DNN, and ML-based yield estimation model, considering the within-field variability in soybean yield. The fusion of data from both satellites significantly outperformed individual predictions, demonstrating higher accuracy and fewer errors. The study utilized various VIs and, during validation, compared crop forecasts using NDVI maps. The effectiveness of ANNs in predicting crop

yields was highlighted, showing superior performance across diverse datasets compared to other algorithms. The research underscored the potential of combining different satellite data sources to improve agricultural yield predictions, offering a more reliable and precise method for farmers and agricultural planners.

6.3. Studies Using Ensemble Methods

Ensemble methods aim to achieve stronger and more accurate predictions by using multiple models together. XGBoost, Stacked Models, and other ensemble approaches work by combining models composed of weak learners. Ensemble methods used in agricultural yield prediction reduce error margins and produce more precise results by bringing together the strengths of different ML algorithms. These methods have been particularly successful with large and complex datasets.

Pejak, et al. [57] utilized Sentinel-2 multispectral images and soil parameters to predict soybean yields over a total area of 411 hectares across 142 fields in Austria during the 2018–2020 growing seasons. They introduced the Polygon-Pixel Interpolation method, which optimizes yield values with satellite imagery. From the 12 bands of the Sentinel-2 multispectral image, they calculated 20 different VIs, including NDVI, EVI, ARVI, SAVI, NDVI_{Red}, VARI, NDWI, MNDWI, VDVI, NLI, MNLI, NMDI, GLI, ExG, CIVE, AWEI, GRVI, GARI, DVI, and LAI. These VIs were tested and compared using ML models such as MLR, SVM, XGBoost, SGD, and RF. The study concluded that the SGD algorithm outperformed the others, with an MAE of 0.436 t/ha.

Desloires, Ienco and Botrel [21] conducted a study to predict end-of-season maize yields at the field scale, using yield data collected from 1164 fields across two U.S. states between 2017 and 2021. They utilized satellite and environmental data, including spectral time series images derived from Sentinel-2 and temperature data obtained from ERA-5 climate reanalysis. Bands B1, B2, B9, and B10 were excluded from the analysis. All other raw band pixels were resampled to a 10-m resolution using nearest-neighbor interpolation. Cloud and shadow-classified pixels were disregarded using the SCL map, and images with more than 50% valid pixels were included in the study. From these images, three VIs (GNDVI, NDRE, NDWI) were calculated. Additionally, the LAI and LCC parameters were estimated using an ANN trained with Sentinel-2 data. To predict yield, they employed ML algorithms such as RR, RF, SVR, MLP, XGBoost, and STACK. Hyperparameters for each algorithm were determined using the Grid Search method. To observe yield predictions under different scenarios, they developed a resampling strategy based on calendar time and thermal time. They also conducted yield predictions under four scenarios: using bands, VIs, bands + VIs, and biophysical parameters. The study concluded that the STACK algorithm performed relatively better and that the use of thermal time was beneficial.

Zhang, Zhang, Liu, Lan, Gao and Li [16] collected winter wheat yield data using combined harvesters from 117 different fields in Henan Province, China, between 2019 and 2021. To automate pre-harvest yield predictions, they employed VI time series data from Landsat 8 and Sentinel-2. A comparison was made using common bands from both sensors, and the BRDF correction model was applied to the images to facilitate more accurate comparisons. The study found a strong correlation between the bands of Sentinel-2 and Landsat 8. The VIs derived from both sensors were integrated using bands B2, B3, B4, B8, B8A, B11, and B12 for Sentinel-2. All bands were resampled to a 10-m resolution, and five different VIs (NDVI, GNDVI, RVI, EVI2, and WDRVI) were calculated. The study identified WDRVI as the most suitable VI for integrating Landsat 8 and Sentinel-2 data. Subsequently, different ML methods, including Bayesian Optimization with CatBoost (BO-CatBoost), LASSO, SVM, and RF, were used to predict winter wheat yield. The study concluded that the BO-CatBoost model provided the best performance for yield prediction (RMSE = 0.62 t/ha). The authors suggested that the proposed method could predict winter wheat yield 40 days before harvest.

Darra, Espejo-Garcia, Kasimati, Kriezi, Psomiadis and Fountas [18] focused on predicting the yield of three different tomato varieties using vegetation indices derived from

Sentinel-2. The study was conducted over an area of 410.10 hectares in Greece, where yield data for 108 different fields with three tomato varieties were collected by farmers under the supervision of agricultural experts in 2021. To model the yield of these fields, five different types of vegetation indices (NDVI, WDI, PVI, RVI, SAVI) were calculated from Sentinel-2 spectral bands. The relationship between these VIs and yield values was modeled using an AutoML technique. This approach automates the selection of the most suitable ML methods from a wide range of options, forming ensembles to enhance performance by combining several ML methods. The study concluded that the AutoML technique successfully predicted yield values both individually and as ensembles. The combination of ARD regression and SVR was identified as the best-performing model ($R^2 = 0.67 \pm 0.02$). Additionally, the study observed that RVI and SAVI were the most effective VIs for yield prediction.

7. Interpretation of Previous Studies

In the previous section, summaries of studies that employed Sentinel-2 bands and VIs derived from these bands for yield prediction using various AI models were presented. Before discussing these studies, this section first provides a quantitative analysis of Sentinel-2's usage based on Web of Science (WOS) data. The bar chart shown in Figure 1 illustrates the number of studies, which are categorized into three distinct groups, over the years 2017 to 2024. These categories include studies using Sentinel-2, studies involving Sentinel-2 and yield prediction, and studies incorporating Sentinel-2, yield prediction, and AI. The studies discussed in Section 2 fall into the third category.

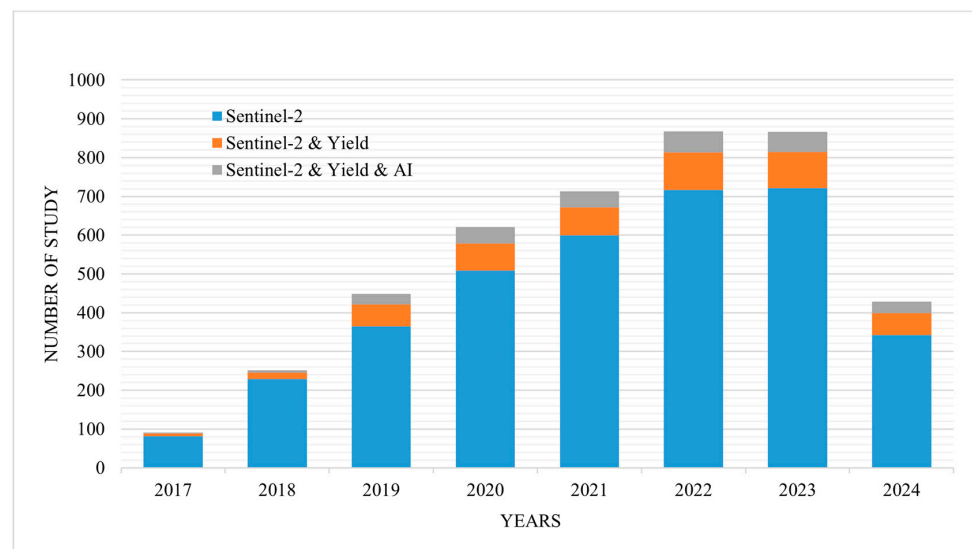


Figure 1. Number of Sentinel-2 related studies in WOS by year.

According to the graph in Figure 1, there is a noticeable and continuous increase in the number of studies utilizing Sentinel-2. While there were relatively few studies in 2017, this number peaked in 2022, surpassing 700. This trend can be attributed to the growing prevalence and importance of Sentinel-2 satellite imaging technology in scientific research. Secondly, there has also been a moderate increase in studies involving both Sentinel-2 and yield prediction. Such studies began to emerge in 2017 and have gradually increased each year, indicating a growing utilization of Sentinel-2 satellite data for yield prediction in agriculture. Thirdly, although the number of studies incorporating Sentinel-2, yield prediction, and AI has increased over the years, this growth has been more limited and less pronounced. This suggests that this area remains relatively new and is still an emerging field of research. The limited widespread integration of AI and yield prediction with Sentinel-2 indicates that this field is still in the early stages of development. In summary,

the graph demonstrates that Sentinel-2 has become a critical tool in scientific research, particularly in the context of yield prediction and AI integration in agriculture.

Table 1 presents the studies summarized in Section 2, categorized by year, crop type, the VIs used, and the learning methods employed. This table facilitates easier observation and comparison of the studies. The selection of these 30 studies was made with attention to the quality of the journals in their respective fields. The studies encompass various crop types and employ diverse VIs and learning methods.

Table 1. AI-based yield estimation studies using Sentinel-2 data conducted in the last 5 years.

No	Study	Publication Year	Crop Type	VI	Learning Method
1	Hunt, Blackburn, Carrasco, Redhead and Rowland [33]	2019	Wheat	GCVI, GNDVI, NDVI, SRI	RF
2	Kayad, Sozzi, Gatto, Marinello and Pirotti [34]	2019	Corn	NDVI, NDRE1, NDRE2, GNDVI, GARVI, EVI, WDRVI, mWDRVI, GCVI	RF, SVM, MLR
3	Gómez, Salvador, Sanz and Casanova [35]	2019	Potato	ARI2, CRI2, IRECI2, LCC, NDVI, PSRI, WdVI	RF, SVM, LR
4	Zhao, Potgieter, Zhang, Wu and Hammer [36]	2020	Wheat	NDVI, OSAVI, SR, DVI, EVI, EVI2, Clred, TCARI, TO, GCVI, GDVI, NDRE1, NDRE2, CCCI	LR
5	Fernandez-Beltran, Baidar, Kang and Pla [51]	2021	Rice	NDVI	CNN, LR, RID, SVR, GPR
6	Narin, Sekertekin, Saygin, Balik Sanli and Gullu [52],	2021	Sunflower	NDVI, NDVIred	LR, ANN, CNN
7	Franch, Bautista, Fita, Rubio, Tarrazó-Serrano, Sánchez, Skakun, Vermote, Becker-Reshef and Uris [19]	2021	Rice	WDRVI, SAVI, RVI, LSWI, NDBI, TVI, IPVI, RGVI, Clred, NDRE1, NDRE2	LR
8	Nazir, Ullah, Saqib, Abbas, Ali, Iqbal, Hussain, Shakir, Shah and Butt [37]	2021	Rice	NDVI, EVI, SAVI, REP	PLSR
9	Son, Chen, Cheng, Toscano, Chen, Chen, Tseng, Syu, Guo and Zhang [38],	2022	Rice	EVI	RF, SVM, ANN
10	Marshall, Belgiu, Boschetti, Pepe, Stein and Nelson [39]	2022	Corn, rice, soybean, wheat	NDVI	PLSR, RF
11	Crusiol, Sun, Sibalidelli, Junior, Furlaneti, Chen, Sun, Wuyun, Chen, Nanni, Furlanetto, Cezar, Nepomuceno and Farias [40]	2022	Soybean	BNDVI, GNDVI, NDVI, NDRE, NDII, NDII 2, EVI1, EVI2	PLSR, SVR
12	Ashourloo, Manafifard, Behifar and Kohandel [41]	2022	Wheat	NDVI, SR, GCVI, GNDVI, WDRVI, DVI, EVI, SAVI, GRRI, NGBDI	KNN, NN, DT, SVR, GPR, RF, LR, SR
13	Bebie, Cavalaris and Kyparissis [42]	2022	Wheat	EVI, NMDI	RF, KNN, BR
14	Segarra, Araus and Kefauver [43]	2022	Wheat	GNDVI, NDVI, RVI, EVI, TGI, NGRDI, CVI	RF, SVM, BR
15	Abebe, Tadesse and Gessesse [44]	2022	Sugarcane	NDVI, EVI, SAVI, MSAVI, SR, GNDVI, SIRI	SVR, MLPNN, MLR
16	Pejak, Lugonja, Antić, Panić, Pandžić, Alexakis, Mavrepis, Zhou, Marko and Crnojević [57]	2022	Soya	NDVI, EVI, ARVI, SAVI, NDVIred, VARI, NDWI, MNDWI, VDVI, NLI, MNLI, NMDI, GLI, ExG, CIVE, AWEI, GRVI, GARI, DVI, LAI	MLR, SVM, XGBoost, SGD

Table 1. Cont.

No	Study	Publication Year	Crop Type	VI	Learning Method
17	Perich, Turkoglu, Graf, Wegner, Aasen, Walter and Liebisch [53]	2023	Winter Wheat	NDVI, GCVI	Four S2 scenes, RNN
18	Bhumiphan, Nontapon, Kaewplang, Srihanu, Koedsin and Huete [17]	2023	Rubber	GSAVI, MSRI, NBR, NDVI, NR, and RVI	LR, MLR
19	Faqe Ibrahim, Rasul and Abdullah [45]	2023	Wheat	EVI, NDVI, NDWI, SAVI, SRI, RVI, GRVI, NDRE, CMFI, chlorophyll, LAI	LR
20	Desloires, Ienco and Botrel [21]	2023	Corn	GNDVI, NDRE, NDWI, LAI, LCC	RR, RF, SVR, MLP, XGBoost, STACK
21	Zhang, Zhang, Liu, Lan, Gao and Li [16]	2023	Wheat	NDVI, GNDVI, RVI, EVI2, WDRVI	BO-CatBoost, LASSO, SVM, RF
22	Darra, Espejo-Garcia, Kasimati, Kriezi, Psomiadis and Fountas [18]	2023	Tomato	NDVI, WDVI, PVI, RVI, SAVI	ARD&SVR (Ensemble)
23	Amankulova, Farmonov and Mucsi [46]	2023	Sunflower	NDVI	RF
24	Nuraeni and Manessa [46]	2023	Tea Leaves	NDVI	RF, SVM
25	Xiao, Zhang, Niu, Li, Li, Zhong and Huang [54]	2024	Wheat	LSWI, IRECI, GCVI, NDVI	ACNN, RF
26	Mancini, Solfanelli, Coviello, Martini, Mandolesi and Zanolli [55]	2024	Wheat	NDVI, NDRE	PLSR, VGG16, VGG19, MobileNetv2
27	Madugundu, Al-Gaadi, Tola, Edriss, Edrees and Alameen [48]	2024	Carrot	NDVI, RDVI, GNDVI, SIPI, GLI	RF
28	Kamenova, Chanev, Dimitrov, Filchev, Bonchev, Zhu and Dong [49]	2024	Winter Wheat	GNDVI	RF, SVM
29	Amankulova, Farmonov, Abdelsamei, Szatmári, Khan, Zhran, Rustamov, Akhmedov, Sarimsakov and Mucsi [56]	2024	Soybean	NDVI, GNDVI, NDRE, EVI, SAVI	ANN, DNN, KNN, RF, SVR, XGBoost
30	de Freitas, Oldoni, Joaquim, Pozzuto and do Amaral [50]	2024	Soybean	EVI, GNDVI, GRNDVI, NDMI, NDRE, NDVI, SFDVI	RF

According to Table 1, when considering the types of crops used in these studies, it is evident that the majority of research has focused on widely cultivated crops such as wheat, maize, and rice. However, other crops like soybean, sunflower, sugarcane, tomato, and potato have also been included. The studies exhibit a broad range of VIs, with NDVI being the most commonly used. It has served as a primary data source in many studies, either as a primary or secondary index. In addition, other indices such as EVI, GNDVI, GCVI, and SAVI have also been widely utilized. It is noteworthy that some studies, despite using the same crop (e.g., wheat), have employed entirely different indices for yield prediction. Most of these indices are used to provide information about vegetation density and health, which is critical for yield prediction. The widespread use of these indices underscores their acceptance as practical tools for assessing plant health and yield potential.

In terms of learning methods, the RF algorithm emerges as one of the most commonly employed techniques. Additionally, other ML methods such as SVM, LR, and PLSR have also been frequently used. In recent years, there has been an increase in the use of more complex DL approaches (e.g., CNN, RNN, Attention-based CNN (ACNN)) and modern methods like XGBoost and VGG19. This trend indicates a new phase in the development of AI-based prediction models. The variations in learning methods highlighted in the table

are particularly striking. For instance, some studies aim to optimize model performance by combining multiple ML techniques, while others focus on a specific algorithm. Additionally, certain studies have concentrated on developing more complex prediction models using advanced DL models (e.g., VGG16, VGG19). This demonstrates that researchers are actively working to identify the most suitable combinations of algorithms and indices for specific crop types and conditions.

Although the studies in Table 1 focus on agricultural yield prediction using Sentinel-2 satellite data and AI models, there are several important limitations and weaknesses in these studies. First, methodological differences between studies using different VIs and AI techniques for the same crop type make it difficult to compare results. For instance, while Hunt, Blackburn, Carrasco, Redhead and Rowland [33] used indices such as NDVI, GNDVI, and RF for wheat yield prediction, Zhao, Potgieter, Zhang, Wu and Hammer [36], in their study on wheat, preferred indices like EVI and OSAVI along with an LR model. These different approaches for the same crop reveal the lack of a standard methodology and limit the comparability of the results. Furthermore, while most studies rely solely on satellite data for yield prediction, important environmental factors such as soil moisture, nutrients, and irrigation, which affect these predictions, are often overlooked. For example, in the potato yield prediction study by Gómez, Salvador, Sanz and Casanova [35], it was emphasized that integrating such additional data could improve prediction accuracy. This suggests that existing models need to be approached more holistically without being limited to only VIs. Lastly, while significant progress has been made in using AI and Sentinel-2 data for agricultural yield prediction, most studies have focused on ML methods (e.g., RF, SVM). The potential of DL techniques has not yet been fully explored. The success of the CNN model in the rice yield prediction study by Fernandez-Beltran, Baidar, Kang and Pla [51] compared to other methods indicates that DL methods could be used more in the future. However, the high computational cost and data requirements of these models limit their widespread use. In light of these critiques, future research should develop more standardized methodologies and integrate more diverse datasets to enhance the generalizability of these models.

8. Discussion

The integration of advanced ML algorithms with RS data has enabled researchers to capture the complex patterns and relationships between VIs and crop yields. By incorporating diverse datasets, including weather conditions and soil properties, these models have enhanced their accuracy and reliability. Such approaches have significantly improved the precision of yield predictions, contributing to more informed decision-making in agricultural management, resource allocation, and food security planning.

AI-based crop yield prediction using VIs derived from Sentinel-2 imagery has shown remarkable potential in recent years. These studies have effectively leveraged Sentinel-2's high spatial, temporal, and spectral resolution to provide accurate and timely yield forecasts. Meghraoui, et al. [58] highlight the efficacy of DL models in enhancing prediction accuracy and address various challenges such as model generalizability and data quality. The integration of advanced ML algorithms with RS data has enabled researchers to capture the complex patterns and relationships between VIs and crop yields. However, the study also emphasizes that many models are limited by their applicability to diverse agricultural contexts and may not fully account for all relevant factors influencing crop yield. To complement these efforts, effective nutrient management strategies are essential. According to Pandey [59], optimizing nutrient delivery and water use in soilless cultures can significantly impact crop productivity. Incorporating such nutrient management strategies into AI-based models could enhance their predictive accuracy by integrating additional agronomic data, thus addressing some of the limitations related to soil properties and nutrient availability.

Although it has been proven that AI methods are highly effective in agricultural yield prediction, several important limitations must be acknowledged. One of the main concerns

is the generalizability of AI models. Many studies are often limited to specific geographical regions or conducted under controlled conditions, which restricts the applicability of these models in diverse agricultural environments. This issue worsens when models are trained on historical yield data that do not accurately reflect current farming practices, climate changes, or the evolution of environmental conditions. The heterogeneity between different crop types, growth stages, and environmental factors creates additional challenges in terms of the robustness of these models and leads to difficulties in accounting for the complex and dynamic nature of agricultural systems.

Additionally, many AI models in this field are typically based on VIs derived from RS data, such as Sentinel-2 imagery. While VIs are effective proxy indicators for plant health and productivity, they only capture a subset of the factors affecting crop yield. Critical variables, such as soil fertility, moisture levels, pest infestations, disease outbreaks, and changes in agricultural practices, are often excluded from models, leading to biased or incomplete yield predictions. Moreover, external factors like atmospheric conditions, sensor errors, or changes in lighting can compromise the accuracy of VIs, introducing data noise that could significantly impact the performance of AI models. The sensitivity of AI systems to such data issues highlights the importance of rigorous preprocessing, data cleaning, and validation for reliable predictions.

In response to these limitations, integrating various data sources, such as ground-based observations, meteorological data, soil moisture sensors, and multispectral or hyperspectral remote sensing from other satellites, could provide a more holistic view of the agricultural environment. This multisource data fusion would likely enhance the robustness and generalizability of AI models by capturing a broader range of factors that influence crop yield. Moreover, improving model explainability remains a critical step toward the broader adoption of AI in agriculture. Especially with DL approaches functioning as 'black boxes', it becomes difficult for agricultural engineers and farmers to understand the rationale behind these models' predictions. The development of transparent and explainable models will build trust and allow experts to make informed decisions based on model outputs.

Finally, ensuring the validation of these AI models under various geographical and climatic conditions will help fine-tune the models for global applicability. Collaborations between data scientists, agricultural engineers, and farmers will be crucial not only for developing accurate AI-driven solutions but also for ensuring their practicality in real-world agricultural settings. The continuous refinement of these techniques, improvement in data quality, and integration of the latest sensor technologies will significantly contribute to sustainable and efficient farming practices in the future.

9. Conclusions

In the context of sustainability, the integration of Sentinel-2 data with AI models holds immense potential for fostering more environmentally responsible agricultural practices. By enabling more accurate and timely yield predictions, these models can assist farmers in optimizing resource use, reducing waste, and minimizing environmental impact. The ability to predict yields with higher precision can lead to better water management, reduced use of fertilizers and pesticides, and more efficient land use, all of which are critical components of sustainable agriculture. Moreover, the growing use of advanced AI techniques such as DL and ensemble methods paves the way for models that not only improve yield forecasts but also account for climatic variations and other factors essential for long-term agricultural resilience. As the agricultural sector faces increasing pressure to produce more food with fewer resources, these advancements will play a vital role in ensuring food security while adhering to sustainability principles. This makes AI-driven yield prediction a key contributor to both the economic and environmental pillars of sustainable agriculture.

This review study has evaluated Sentinel-2-based yield prediction studies, which are still new but effective in the field. It comprehensively analyzes 30 studies conducted over the past five years, which employed AI for yield forecasting. Notably, our study features a table summarizing these studies and a graph illustrating the utilization of Sentinel-2 data

by researchers over the years, which are key highlights of our work. The graph in Figure 1 demonstrates the increasing significance of Sentinel-2 in academia. The studies presented in Table 1 reveal that AI-based yield prediction using Sentinel-2 data is a highly diverse and evolving area of research. These studies have developed various approaches using different crop types, VIs, and AI techniques. Particularly noteworthy are the variations in VIs and the extensive use of different types of vegetation indices in these studies. There is significant diversity in the learning methods employed. While traditional methods like RF remain popular, other techniques such as SVM and regression methods are also frequently used. Additionally, the use of DL and ensemble techniques is seen as heralding a new era in agricultural yield prediction. Earlier studies (2019–2020) focused on simpler ML methods such as RF, LR, and SVM. More recent studies (2022–2024) indicate a shift toward more advanced ML techniques, including CNNs, RNNs, attention networks, and complex ensemble methods. This diversity allows researchers to select the most suitable model based on different data types and problem definitions.

Overall, the studies reviewed demonstrate significant diversity and advancement in AI-based yield prediction. In most studies, Sentinel-2 data and VIs were used as primary inputs for yield prediction, and their integration with AI models has created substantial potential in the agricultural field. Based on this information, it is evident that there is a trend toward achieving more advanced and accurate predictions in agricultural yield forecasting, a trend that is likely to continue in the coming years, with Sentinel-2 and AI applications playing a crucial role. Progress in this field will contribute significantly to agricultural production and management strategies. The fact that it is still a nascent and developing area suggests that many new researchers will emerge in this field. In this context, our study, as the first review in this domain, is intended to serve as an important guide for researchers.

Author Contributions: Conceptualization, M.F.A.; methodology, M.F.A.; investigation, M.F.A., K.S. and B.A.; resources, M.F.A. and B.A.; writing—original draft preparation, M.F.A. and B.A.; writing—review and editing, M.F.A., K.S. and B.A.; visualization, B.A.; supervision, M.F.A. and K.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data sharing is not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

The definitions and formulas of the VI abbreviations used in the studies shown in Table 1 are presented in Table A1. The references for each formula are the studies provided in Table 1 and <https://www.indexdatabase.de/> (accessed on 22 September 2024).

Table A1. Explanations of VI abbreviations.

Abbreviation	Definition	Formula
ARI2	Anthocyanin Reflectance Index	$(1/550nm) - (1/700nm)$
ARVI	Atmospherically Resistant Vegetation Index	$\frac{(NIR - (RED - 1.7 \times (BLUE - RED)))}{(NIR + (RED - 1.7 \times (BLUE - RED)))}$
AWEI	Automated Water Extraction Index	$4 \times (Green - SWIR2) - (0.25 \times NIR + 2.75 \times SWIR3)$
BNDVI	Blue Normalized Difference Vegetation Index	$(NIR - Blue) / (NIR + Blue)$
CCCI	Canopy Chlorophyll Content Index	$NDRE1/OSAVI$
CIRed	Chlorophyll Index red edge	$(NIR/RedEdge) - 1$
CIVE	Color Index of Vegetation Extraction	$0.441 \times Red - 0.881 \times Green + 0.385 \times Blue + 18.78745$

Table A1. Cont.

Abbreviation	Definition	Formula
CMFI	Cropping Management Factor Index	$\frac{Red}{NIR+Red}$
CRI2	Carotenoid Reflectance Index	$(1/510nm) - (1/700nm)$
CVI	Chlorophyll vegetation index	$NIR \times (Red / (Green^2))$
DVI	Difference Vegetation Index	$NIR - Red$
EVI	Enhanced Vegetation Index	$2.5 \times \frac{(NIR-Red)}{(NIR+6 \times Red-7.5 \times Blue+1)}$
EVI2	Enhanced Vegetation Index 2	$2.5 \times \frac{(NIR-Red)}{(NIR+2.4 \times Red+1)}$
ExG	Excess Green	$2 \times Green - Red - Blue$
GARI	Green Atmospherically Resistant Index	$\frac{NIR-(Green-1.7 \times (Blue-Red))}{NIR+(Green-1.7 \times (Blue-Red))}$
GCVI	Green chlorophyll vegetation index	$NIR/Green - 1$
GDVI	Green Difference Vegetation Index	$NIR - Green$
GLI	Green Leaf Index	$\frac{(Green-Red)(Green-Blue)}{(2 \times Green)+Red+Blue}$
GNDVI	Green normalized difference vegetation index	$\frac{NIR-Green}{NIR+Green}$
GRR	Green-red ratio Index	$Green / Red$
GRVI	Green-Red Vegetation Index	$\frac{Green-Red}{Green+Red}$
GSAVI	Green Soil Adjusted Vegetation Index	$\frac{NIR-Green}{(NIR+Green+0.5)} \times (1 + 0.5)$
IPVI	Infrared percentage vegetation index	$(NIR / (NIR + Red) / 2) \times (NDVI + 1)$
IRECI	Inverted Red-Edge Chlorophyll Index	$\frac{(RedEdge3-Red)}{RedEdge1 / RedEdge2}$
LAI	Leaf Area index	$3.618 \times EVI - 0.118$
LSWI	Land surface water index	$\frac{NIR-SWIR}{NIR+SWIR}$
MNLI	Modified Non-linear Index	$\frac{(NIR^2-Red)(1+0.5)}{NIR^2+Red+0.5}$
MNDWI	Modified Normalized Difference Vegetation Index	$\frac{Green-SWIR2}{Green+SWIR2}$
MSAVI	Modified soil adjusted vegetation index	$\frac{2 \times NIR + 1 - \sqrt{((2 \times NIR + 1)^2 - 8 \times (NIR - Red))}}{2}$
MSRI	Modified Simple Ratio Index	$\frac{800nm-445nm}{680nm-445nm}$
NBR	Normalized Burn Ratio	$\frac{NIR-SWIR2}{NIR+SWIR2}$
NDBI	Normalized difference built-up index	$\frac{SWIR2-NIR}{SWIR2+NIR}$
NDII	Normalized Difference Infrared Index 1	$\frac{NIR-SWIR2}{NIR+SWIR2}$
NDII2	Normalized Difference Infrared Index 2	$\frac{NIR-SWIR3}{NIR+SWIR3}$
NDRE1	Normalized difference red edge1	$\frac{NIR-RedEdge}{NIR+RedEdge}$
NDRE2	Normalized difference red edge2	$\frac{NIR-RedEdge2}{NIR+RedEdge2}$
NDVI	Normalized Difference Vegetation Index	$\frac{(NIR-Red)}{(NIR+Red)}$
NDVIRE	Normalized Difference Vegetation Index Red-edge	$\frac{(NIR-RedEdge)}{(NIR+RedEdge)}$
NDWI	Normalized Difference Water Index	$\frac{(NIR-SWIR2)}{(NIR+SWIR2)}$
NGBDI	Normalized green-blue difference Index	$\frac{Green-Blue}{Green+Blue}$
NGRDI	Normalized Green-Red Difference Index	$\frac{Green-Red}{Green+Red}$

Table A1. Cont.

Abbreviation	Definition	Formula
NLI	Non-linear Index	$\frac{(NIR^2 - Red)}{(NIR^2 + Red)}$
NMDI	Normalized Multiband Drought Index	$\frac{NIR2 - (SWIR2 - SWIR3)}{NIR2 + (SWIR2 - SWIR3)}$
OSAVI	Optimized Soil Adjusted Vegetation Index	$1.16 \times \frac{NIR - Red}{NIR + Red + 0.16}$
PSRI	Plant Senescence Reflectance Index	$\frac{Red - Blue}{RedEdge2}$
PVI	Perpendicular Vegetation Index	$\frac{NIR - a \times Red - b}{\sqrt{(a^2 + 1)}}$
RDVI	Renormalized Difference Vegetation Index	$\frac{NIR - Red}{\sqrt{NIR + Red}}$
REP	Red Edge Position	$704 + 35 \times \left(\frac{RedEdge - RedEdge1}{RedEdge2 - RedEdge3} \right)$
RGVI	Rice growth vegetation index	$1 - \frac{Blue - Red}{NIR + SWIR2 + SWIR3}$
RVI	Ratio vegetation index	Red / NIR
SAVI	Soil-adjusted vegetation index	$\frac{NIR - Red}{(NIR + Red + 0.5)} \times (1 + 0.5)$
SIPI	Structure Intensive Pigment Vegetation Index	$\frac{NIR - Blue}{NIR - Red}$
SRI	Simple ratio index	NIR / Red
TCARI	Transformed Chlorophyll Absorption in Reflectance Index	$3 \times [(RedEdge1 - Red) - 0.2 \times (RedEdge1 - Green) \times RedEdge1 / Red]$
TGI	Triangular Greenness Index	$-0.5 \times (190 \times (Red - Green) - 120 \times (Red - Blue))$
TO	TCARI/OSAVI	$TCARI / OSAVI$
TVI	Triangular vegetation index	$\sqrt{\frac{NIR - Red}{NIR + Red} + 0.5}$
VARI	Visible Atmospherically Resistant Index	$\frac{(Green - Red)}{(Green + Red - Blue)}$
VDVI	Visible-Band Difference Vegetation Index	$\frac{2 \times Green - Red - Blue}{2 \times Green + Red + Blue}$
WDRVI	Wide Dynamic Range Vegetation Index	$\frac{0.1 \times NIR - Red}{0.1 \times NIR + Red}$
WDVI	Weighted Difference Vegetation Index	$NIR - a \times Red$

Appendix B

Definitions of other abbreviations, except VI, are shown in Table A2.

Table A2. Other abbreviations.

Abbreviation	Definition
ANN	Artificial Neural Network
ARD	Automatic relevance determination
ACNN	Attention-Based One-Dimensional Convolutional Neural Network
BO-CatBoost	Bayesian optimized CatBoost
BR	Boosting regression
BRDF	Bidirectional Reflectance Distribution Function
CNN	Convolutional Neural Network
DT	Decision Tree
DNN	Deep Neural Network
GPR	Gaussian process regression
KNN	K-Nearest Neighbor

Table A2. Cont.

Abbreviation	Definition
LASSO	Least Absolute Shrinkage and Selection Operator
LR	Linear Regression
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MLPNN	Multilayer perceptron neural network
MLR	Multiple linear regression
MSE	Mean Squared Error
PLSR	Partial Least Squares Regression
RF	Random Forest
RID	Ridge Regression
RMSE	Root Mean Squared Error
RR	Ridge Regression
SCL	Scene Classification
SGD	Stochastic Gradient Descent
SPAD	Soil Plant Analysis Development
SR	Stepwise Regression
STACK	Stacked Averaging Ensemble
SVM	Support Vector Model
SVR	Support Vector Regression
VI	Vegetation indices
XGBoost	Extreme Gradient Boosting

References

- Aslan, M.F.; Durdu, A.; Sabanci, K.; Ropelewska, E.; Gültekin, S.S. A Comprehensive Survey of the Recent Studies with UAV for Precision Agriculture in Open Fields and Greenhouses. *Appl. Sci.* **2022**, *12*, 1047. [[CrossRef](#)]
- Bégué, A.; Arvor, D.; Bellon, B.; Betbeder, J.; De Abelleira, D.; P. D. Ferraz, R.; Lebourgeois, V.; Lelong, C.; Simões, M.; R. Verón, S. Remote Sensing and Cropping Practices: A Review. *Remote Sens.* **2018**, *10*, 99. [[CrossRef](#)]
- Denton, O.; Aduramigba-Modupe, V.; Ojo, A.; Adeoyolanu, O.; Are, K.; Adelana, A.; Oyedele, A.; Adetayo, A.; Oke, A. Assessment of spatial variability and mapping of soil properties for sustainable agricultural production using geographic information system techniques (GIS). *Cogent Food Agric.* **2017**, *3*, 1279366. [[CrossRef](#)]
- Navalgund, R.R.; Jayaraman, V.; Roy, P. Remote sensing applications: An overview. *Curr. Sci.* **2007**, *93*, 1747–1766.
- Matese, A.; Toscano, P.; Di Gennaro, S.F.; Genesisio, L.; Vaccari, F.P.; Primicerio, J.; Belli, C.; Zaldei, A.; Bianconi, R.; Gioli, B. Intercomparison of UAV, aircraft and satellite remote sensing platforms for precision viticulture. *Remote Sens.* **2015**, *7*, 2971–2990. [[CrossRef](#)]
- Winkler, K.; Gessner, U.; Hochschild, V. Identifying Droughts Affecting Agriculture in Africa Based on Remote Sensing Time Series between 2000–2016: Rainfall Anomalies and Vegetation Condition in the Context of ENSO. *Remote Sens.* **2017**, *9*, 831. [[CrossRef](#)]
- Sishodia, R.P.; Ray, R.L.; Singh, S.K. Applications of remote sensing in precision agriculture: A review. *Remote Sens.* **2020**, *12*, 3136. [[CrossRef](#)]
- Aslan, M.F.; Durdu, A.; Sabanci, K. Goal distance-based UAV path planning approach, path optimization and learning-based path estimation: GDRRT*, PSO-GDRRT* and BiLSTM-PSO-GDRRT*. *Appl. Soft Comput.* **2023**, *137*, 110156. [[CrossRef](#)]
- Furlanetto, J.; Dal Ferro, N.; Longo, M.; Sartori, L.; Polese, R.; Caceffo, D.; Nicoli, L.; Morari, F. LAI estimation through remotely sensed NDVI following hail defoliation in maize (*Zea mays* L.) using Sentinel-2 and UAV imagery. *Precis. Agric.* **2023**, *24*, 1355–1379. [[CrossRef](#)]
- Binte Mostafiz, R.; Noguchi, R.; Ahamed, T. Agricultural land suitability assessment using satellite remote sensing-derived soil-vegetation indices. *Land* **2021**, *10*, 223. [[CrossRef](#)]
- Ali, U.; Esau, T.J.; Farooque, A.A.; Zaman, Q.U.; Abbas, F.; Bilodeau, M.F. Limiting the Collection of Ground Truth Data for Land Use and Land Cover Maps with Machine Learning Algorithms. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 333. [[CrossRef](#)]

12. Khanal, S.; Kc, K.; Fulton, J.P.; Shearer, S.; Ozkan, E. Remote sensing in agriculture—Accomplishments, limitations, and opportunities. *Remote Sens.* **2020**, *12*, 3783. [[CrossRef](#)]
13. Yuan, Q.; Shen, H.; Li, T.; Li, Z.; Li, S.; Jiang, Y.; Xu, H.; Tan, W.; Yang, Q.; Wang, J.; et al. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* **2020**, *241*, 111716. [[CrossRef](#)]
14. Lin, W.; Zhang, D.; Liu, F.; Guo, Y.; Chen, S.; Wu, T.; Hou, Q. A Lightweight Multi-Label Classification Method for Urban Green Space in High-Resolution Remote Sensing Imagery. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 252. [[CrossRef](#)]
15. Aslan, M.F. A hybrid end-to-end learning approach for breast cancer diagnosis: Convolutional recurrent network. *Comput. Electr. Eng.* **2023**, *105*, 108562. [[CrossRef](#)]
16. Zhang, H.; Zhang, Y.; Liu, K.; Lan, S.; Gao, T.; Li, M. Winter wheat yield prediction using integrated Landsat 8 and Sentinel-2 vegetation index time-series data and machine learning algorithms. *Comput. Electron. Agric.* **2023**, *213*, 108250. [[CrossRef](#)]
17. Bhumiphan, N.; Nontapon, J.; Kaewplang, S.; Srihanu, N.; Koedsin, W.; Huete, A. Estimation of Rubber Yield Using Sentinel-2 Satellite Data. *Sustainability* **2023**, *15*, 7223. [[CrossRef](#)]
18. Darra, N.; Espejo-Garcia, B.; Kasimati, A.; Kriezi, O.; Psomiadis, E.; Fountas, S. Can Satellites Predict Yield? Ensemble Machine Learning and Statistical Analysis of Sentinel-2 Imagery for Processing Tomato Yield Prediction. *Sensors* **2023**, *23*, 2586. [[CrossRef](#)]
19. Franch, B.; Bautista, A.S.; Fita, D.; Rubio, C.; Tarrazó-Serrano, D.; Sánchez, A.; Skakun, S.; Vermote, E.; Becker-Reshef, I.; Uris, A. Within-Field Rice Yield Estimation Based on Sentinel-2 Satellite Data. *Remote Sens.* **2021**, *13*, 4095. [[CrossRef](#)]
20. Revel, C.; Lonjou, V.; Marcq, S.; Desjardins, C.; Fougny, B.; Coppolani-Delle Luche, C.; Guillemot, N.; Lacamp, A.-S.; Lourme, E.; Miquel, C.; et al. Sentinel-2A and 2B absolute calibration monitoring. *Eur. J. Remote Sens.* **2019**, *52*, 122–137. [[CrossRef](#)]
21. Desloires, J.; Ienco, D.; Botrel, A. Out-of-year corn yield prediction at field-scale using Sentinel-2 satellite imagery and machine learning methods. *Comput. Electron. Agric.* **2023**, *209*, 107807. [[CrossRef](#)]
22. Wang, J.; Wang, P.; Tian, H.; Tansey, K.; Liu, J.; Quan, W. A deep learning framework combining CNN and GRU for improving wheat yield estimates using time series remotely sensed multi-variables. *Comput. Electron. Agric.* **2023**, *206*, 107705. [[CrossRef](#)]
23. Clark, M.L. Comparison of multi-seasonal Landsat 8, Sentinel-2 and hyperspectral images for mapping forest alliances in Northern California. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 26–40. [[CrossRef](#)]
24. Liang, J.; Ren, C.; Li, Y.; Yue, W.; Wei, Z.; Song, X.; Zhang, X.; Yin, A.; Lin, X. Using Enhanced Gap-Filling and Whittaker Smoothing to Reconstruct High Spatiotemporal Resolution NDVI Time Series Based on Landsat 8, Sentinel-2, and MODIS Imagery. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 214. [[CrossRef](#)]
25. Xue, J.; Su, B. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *J. Sens.* **2017**, *2017*, 1353691. [[CrossRef](#)]
26. Roznik, M.; Boyd, M.; Porth, L. Improving crop yield estimation by applying higher resolution satellite NDVI imagery and high-resolution cropland masks. *Remote Sens. Appl. Soc. Environ.* **2022**, *25*, 100693. [[CrossRef](#)]
27. Wang, Q.; Blackburn, G.A.; Onojeghuo, A.O.; Dash, J.; Zhou, L.; Zhang, Y.; Atkinson, P.M. Fusion of Landsat 8 OLI and Sentinel-2 MSI Data. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3885–3899. [[CrossRef](#)]
28. Oikonomidis, A.; Catal, C.; Kassahun, A. Deep learning for crop yield prediction: A systematic literature review. *N. Z. J. Crop Hortic. Sci.* **2023**, *51*, 1–26. [[CrossRef](#)]
29. Van Klompenburg, T.; Kassahun, A.; Catal, C. Crop yield prediction using machine learning: A systematic literature review. *Comput. Electron. Agric.* **2020**, *177*, 105709. [[CrossRef](#)]
30. Luo, L.; Sun, S.; Xue, J.; Gao, Z.; Zhao, J.; Yin, Y.; Gao, F.; Luan, X. Crop yield estimation based on assimilation of crop models and remote sensing data: A systematic evaluation. *Agric. Syst.* **2023**, *210*, 103711. [[CrossRef](#)]
31. Khairunniza-Bejo, S.; Mustaffha, S.; Ismail, W.I.W. Application of artificial neural network in predicting crop yield: A review. *J. Food Sci. Eng.* **2014**, *4*, 1.
32. Dharani, M.K.; Thamilselvan, R.; Natesan, P.; Kalaivaani, P.C.D.; Santhoshkumar, S. Review on Crop Prediction Using Deep Learning Techniques. *J. Phys. Conf. Ser.* **2021**, *1767*, 012026. [[CrossRef](#)]
33. Hunt, M.L.; Blackburn, G.A.; Carrasco, L.; Redhead, J.W.; Rowland, C.S. High resolution wheat yield mapping using Sentinel-2. *Remote Sens. Environ.* **2019**, *233*, 111410. [[CrossRef](#)]
34. 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](#)]
35. Gómez, D.; Salvador, P.; Sanz, J.; Casanova, J.L. Potato Yield Prediction Using Machine Learning Techniques and Sentinel 2 Data. *Remote Sens.* **2019**, *11*, 1745. [[CrossRef](#)]
36. Zhao, Y.; Potgieter, A.B.; Zhang, M.; Wu, B.; Hammer, G.L. Predicting Wheat Yield at the Field Scale by Combining High-Resolution Sentinel-2 Satellite Imagery and Crop Modelling. *Remote Sens.* **2020**, *12*, 1024. [[CrossRef](#)]
37. Nazir, A.; Ullah, S.; Saqib, Z.A.; Abbas, A.; Ali, A.; Iqbal, M.S.; Hussain, K.; Shakir, M.; Shah, M.; Butt, M.U. Estimation and Forecasting of Rice Yield Using Phenology-Based Algorithm and Linear Regression Model on Sentinel-II Satellite Data. *Agriculture* **2021**, *11*, 1026. [[CrossRef](#)]
38. Son, N.-T.; Chen, C.-F.; Cheng, Y.-S.; Toscano, P.; Chen, C.-R.; Chen, S.-L.; Tseng, K.-H.; Syu, C.-H.; Guo, H.-Y.; Zhang, Y.-T. Field-scale rice yield prediction from Sentinel-2 monthly image composites using machine learning algorithms. *Ecol. Inform.* **2022**, *69*, 101618. [[CrossRef](#)]
39. Marshall, M.; 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](#)]

40. Crusiolo, L.G.T.; Sun, L.; Sibaldelli, R.N.R.; Junior, V.F.; Furlaneti, W.X.; Chen, R.; Sun, Z.; Wuyun, D.; Chen, Z.; Nanni, M.R.; et al. Strategies for monitoring within-field soybean yield using Sentinel-2 Vis-NIR-SWIR spectral bands and machine learning regression methods. *Precis. Agric.* **2022**, *23*, 1093–1123. [[CrossRef](#)]
41. Ashourloo, D.; Manafifard, M.; Behifar, M.; Kohandel, M. Wheat yield prediction based on Sentinel-2, regression, and machine learning models in Hamedan, Iran. *Sci. Iran.* **2022**, *29*, 3230–3243. [[CrossRef](#)]
42. Bebie, M.; Cavalaris, C.; Kyparissis, A. Assessing Durum Wheat Yield through Sentinel-2 Imagery: A Machine Learning Approach. *Remote Sens.* **2022**, *14*, 3880. [[CrossRef](#)]
43. Segarra, J.; Araus, J.L.; Kefauver, S.C. Farming and Earth Observation: Sentinel-2 data to estimate within-field wheat grain yield. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *107*, 102697. [[CrossRef](#)]
44. Abebe, G.; Tadesse, T.; Gessesse, B. Combined Use of Landsat 8 and Sentinel 2A Imagery for Improved Sugarcane Yield Estimation in Wonji-Shoa, Ethiopia. *J. Indian Soc. Remote Sens.* **2022**, *50*, 143–157. [[CrossRef](#)]
45. Faqe Ibrahim, G.R.; Rasul, A.; Abdullah, H. Sentinel-2 accurately estimated wheat yield in a semi-arid region compared with Landsat 8. *Int. J. Remote Sens.* **2023**, *44*, 4115–4136. [[CrossRef](#)]
46. Amankulova, K.; Farmonov, N.; Mucsi, L. Time-series analysis of Sentinel-2 satellite images for sunflower yield estimation. *Smart Agric. Technol.* **2023**, *3*, 100098. [[CrossRef](#)]
47. Nuraeni, D.; Manessa, M.D.M. Spatial machine learning for monitoring tea leaves and crop yield estimation using sentinel-2 imagery, (A Case of Gunung Mas Plantation, Bogor). *Int. J. Remote Sens. Earth Sci. (IJReSES)* **2023**, *19*, 133–142. [[CrossRef](#)]
48. Madugundu, R.; Al-Gaadi, K.A.; Tola, E.; Edriss, M.K.; Edrees, H.F.; Alameen, A.A. Optimal Timing of Carrot Crop Monitoring and Yield Assessment Using Sentinel-2 Images: A Machine-Learning Approach. *Appl. Sci.* **2024**, *14*, 3636. [[CrossRef](#)]
49. Kamenova, I.; Chaney, M.; Dimitrov, P.; Filchev, L.; Bonchev, B.; Zhu, L.; Dong, Q. Crop Type Mapping and Winter Wheat Yield Prediction Utilizing Sentinel-2: A Case Study from Upper Thracian Lowland, Bulgaria. *Remote Sens.* **2024**, *16*, 1144. [[CrossRef](#)]
50. de Freitas, R.G.; Oldoni, H.; Joaquim, L.F.; Pozzuto, J.V.F.; do Amaral, L.R. Predicting on-farm soybean yield variability using texture measures on Sentinel-2 image. *Precis. Agric.* **2024**. [[CrossRef](#)]
51. Fernandez-Beltran, R.; Baidar, T.; Kang, J.; Pla, F. Rice-Yield Prediction with Multi-Temporal Sentinel-2 Data and 3D CNN: A Case Study in Nepal. *Remote Sens.* **2021**, *13*, 1391. [[CrossRef](#)]
52. Narin, O.G.; Sekertekin, A.; Saygin, A.; Balik Sanli, F.; Gullu, M. Yield Estimation of Sunflower Plant with CNN and ANN Using Sentinel-2. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *46*, 385–389. [[CrossRef](#)]
53. Perich, G.; Turkoglu, M.O.; Graf, L.V.; Wegner, J.D.; Aasen, H.; Walter, A.; Liebisch, F. Pixel-based yield mapping and prediction from Sentinel-2 using spectral indices and neural networks. *Field Crops Res.* **2023**, *292*, 108824. [[CrossRef](#)]
54. Xiao, G.; Zhang, X.; Niu, Q.; Li, X.; Li, X.; Zhong, L.; Huang, J. Winter wheat yield estimation at the field scale using sentinel-2 data and deep learning. *Comput. Electron. Agric.* **2024**, *216*, 108555. [[CrossRef](#)]
55. Mancini, A.; Solfanelli, F.; Coviello, L.; Martini, F.M.; Mandolesi, S.; Zanolli, R. Time Series from Sentinel-2 for Organic Durum Wheat Yield Prediction Using Functional Data Analysis and Deep Learning. *Agronomy* **2024**, *14*, 109. [[CrossRef](#)]
56. Amankulova, K.; Farmonov, N.; Abdelsamei, E.; Szatmári, J.; Khan, W.; Zhran, M.; Rustamov, J.; Akhmedov, S.; Sarimsakov, M.; Mucsi, L. A Novel Fusion Method for Soybean Yield Prediction Using Sentinel-2 and PlanetScope Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 13694–13707. [[CrossRef](#)]
57. Pejak, B.; Lugonja, P.; Antić, A.; Panić, M.; Pandžić, M.; Alexakis, E.; Mavrepis, P.; Zhou, N.; Marko, O.; Crnojević, V. Soya Yield Prediction on a Within-Field Scale Using Machine Learning Models Trained on Sentinel-2 and Soil Data. *Remote Sens.* **2022**, *14*, 2256. [[CrossRef](#)]
58. Meghraoui, K.; Sebari, I.; Pilz, J.; Ait El Kadi, K.; Bensiali, S. Applied Deep Learning-Based Crop Yield Prediction: A Systematic Analysis of Current Developments and Potential Challenges. *Technologies* **2024**, *12*, 43. [[CrossRef](#)]
59. Pandey, K. Nutrient Management Strategies for Water and Nutrient Saving in Substrate Soilless Culture under Protected Cultivation. In *Artificial Intelligence and Smart Agriculture: Technology and Applications*; Pandey, K., Kushwaha, N.L., Pande, C.B., Singh, K.G., Eds.; Springer Nature: Singapore, 2024; pp. 369–386.

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.