

BIG DATA IN AGRONOMY: TECHNOLOGIES, APPLICATIONS, BENEFITS AND CHALLENGES

Kodavali Aswini¹, *Raju Marimuthu¹, Guru Gurusamy¹, Kannan Pandian², Anandham Rangasamy³

¹Department of Agronomy, Tamil Nadu Agricultural University, Coimbatore 641003, India

²Soil Science & Agricultural Chemistry, Agricultural College & Research Institute, Chettinad-630102, India

³Department of Agricultural Microbiology, Tamil Nadu Agricultural University, Coimbatore- 641003, India

*Correspondence email: raju.m@tnau.ac.in, E-mail ID and ORCID

Sl. No.	Full name	Abbreviated name	E-mail ID	ORCID**	Author Status (student, faculty, scientist, etc.)
1.	Kodavali Aswini	KA	aswinikodavali4@gmail.com	0009-0002-6739-8300	PG Scholar
2.	Raju Marimuthu	RM	raju.m@tnau.ac.in	0000-0001-9026-5106	Professor
3.	Guru Gurusamy	GG	guru.g@tnau.ac.in	0000-0002-3202-378X	Associate Professor
4.	Kannan Pandian	KP	Kannan.p@tnau.ac.in	0000-0001-7003-3535	Associate Professor
5.	Anandham Rangasamy	AR	anandhamranga@gmail.com	0000-0002-3496-564X	Associate Professor

ABSTRACT

Big Data is becoming an important component of modern agronomy by enabling the integration, processing and analysis of large, diverse and continuously generated datasets for improved agricultural decision-making. Agricultural Big Data is generated from multiple sources, including remote sensing satellites, Unmanned Aerial Vehicles (UAVs), Internet of Things (IoT) sensors, weather stations, farm machinery, field observations and historical databases. The integration of these data sources with Geographic Information Systems (GIS), cloud computing, Artificial Intelligence (AI) and Machine Learning (ML) enables the identification of spatial and temporal patterns that are difficult to detect using conventional approaches. Big Data applications in agronomy include climate and weather prediction, pest and disease management, soil health and nutrient management, irrigation scheduling, crop yield prediction, precision agriculture, crop insurance and weed management. Examples from India and other regions demonstrate the application of satellite imagery, UAV-based sensing, IoT systems, crop-growth models and machine-learning algorithms for field-scale and regional agricultural decision-making. Big Data analytics can improve resource-use efficiency, support predictive and real-time monitoring, strengthen agricultural risk assessment and facilitate site-specific management. However, effective implementation is constrained by challenges such as high infrastructure and operational costs, data quality and interoperability issues, data privacy and ownership concerns, limited rural digital infrastructure, inadequate technical skills and increasing requirements for data storage and computing. Therefore, the effective integration of Big Data with agronomic knowledge and decision-support systems can contribute to more precise, efficient and sustainable agricultural production while addressing emerging challenges related to climate variability, resource scarcity and food security.

KEYWORDS: Big Data, Precision Agriculture, Internet of Things, Remote Sensing, Artificial Intelligence, Machine Learning, Decision Support Systems

1. INTRODUCTION

Agriculture is undergoing a major transformation with the rapid advancement of digital technologies and data-driven systems. In recent years, the concept of Big Data has emerged as a powerful tool for improving agricultural productivity, sustainability, and resource management. Big Data refers to extremely large and complex datasets that are generated at high speed from multiple sources and cannot be effectively processed using traditional data management systems (Bhat & Huang, 2021). In agriculture, these datasets are produced through various digital technologies such as remote sensing satellites, unmanned aerial vehicles (UAVs), Internet of Things (IoT) sensors, weather monitoring systems, farm machinery, and mobile-based advisory platforms. The integration of these

technologies generates vast volumes of information related to crop growth, soil conditions, weather patterns, pest incidence, and farm management practices. Traditionally, agricultural decision-making has largely relied on farmers' experience, field observations, and historical data. While these approaches have supported crop production for many decades, they often fail to address modern agricultural challenges such as climate variability, spatial variability within fields, resource scarcity, and increasing food demand.

Agricultural systems are highly complex and dynamic, influenced by multiple interacting factors including soil properties, climatic conditions, crop genetics, and management practices. Conventional data analysis methods are often insufficient to analyze such complex interactions. Big Data analytics provides new opportunities to analyze large-scale agricultural datasets and extract meaningful insights that can support efficient farm management and policy decision-making (Kamilaris et al., 2017). The development of precision agriculture has further accelerated the adoption of Big Data technologies in modern farming systems. Precision agriculture aims to optimize crop production by managing spatial and temporal variability within agricultural fields. By collecting and analyzing detailed field-level data, farmers can apply inputs such as water, fertilizers, and pesticides more precisely according to crop requirements. Technologies such as global positioning systems (GPS), geographic information systems (GIS), remote sensing, drones, and soil sensors enable the collection of high-resolution spatial data that can be used for site-specific crop management. Big Data platforms allow these diverse datasets to be integrated and analyzed to provide accurate recommendations for irrigation scheduling, nutrient management, pest control, and yield prediction (Liakos et al., 2018). Another important driver of Big Data adoption in agriculture is the rapid growth of Internet of Things (IoT) technologies. IoT-based devices such as soil moisture sensors, temperature sensors, and weather stations continuously collect real-time environmental data from agricultural fields. These sensors transmit data to cloud-based platforms where advanced analytical algorithms process the information and generate decision-support recommendations. For example, soil moisture sensors can help determine the optimal irrigation time, while weather-based predictive models can forecast pest and disease outbreaks. Such real-time monitoring systems enable farmers to respond quickly to changing field conditions and improve resource use efficiency (Bhat & Huang, 2021). In addition to descriptive analytics, the application of artificial intelligence (AI) and machine learning (ML) has substantially advanced the analytical capability of Big Data systems in agriculture. Techniques such as artificial neural networks, support vector machines, random forests, and deep learning architectures-including convolutional neural networks for image-based tasks-are now widely applied for crop yield prediction, pest and disease detection, and crop suitability assessment (Liakos et al., 2018; Zhang et al., 2020). These AI-driven models can process heterogeneous data streams, from satellite imagery to sensor time-series, to uncover nonlinear relationships between environmental variables and crop performance that traditional statistical approaches often fail to capture.

Big Data analytics also plays an important role in addressing major global agricultural challenges such as climate change, water scarcity, and food security. Climate change has increased the frequency of extreme weather events such as droughts, floods, and heat waves, which significantly affect crop production. By analyzing long-term climate data and satellite observations, Big Data models can predict weather patterns and help farmers plan appropriate cropping strategies. Similarly, advanced data analytics can improve irrigation management by estimating crop water requirements based on weather conditions and crop growth stages. These applications contribute to improved water-use efficiency and sustainable agricultural production systems (Kamilaris et al., 2017).

Despite its numerous benefits, the implementation of Big Data in agriculture also faces several challenges. These include limited digital infrastructure in rural areas, lack of data standardization, data privacy concerns, and limited digital literacy among farmers. The effective use of Big Data requires reliable internet connectivity, advanced computing infrastructure, and skilled personnel capable of handling large datasets and complex analytical tools. In many developing countries, including India, these limitations can hinder the widespread adoption of digital agriculture technologies. Therefore, strengthening digital infrastructure, improving farmer awareness, and developing user-friendly decision support systems are essential for maximizing the benefits of Big Data in agriculture (Liakos et al., 2018).

Beyond farm-level decision-making, Big Data technologies are increasingly being integrated into agricultural supply chains to improve transparency, traceability, and stakeholder trust. Blockchain-based systems, when combined with Big Data platforms, enable secure and tamper-proof recording of information across the food supply chain, from farm-level production data through post-harvest handling and distribution (Thinakaran et al., 2023). Such integration proved particularly valuable during supply chain disruptions such as the COVID-19 pandemic, where transparent and traceable data flows helped sustain the resilience of agricultural markets (Khan et al., 2022).

Overall, Big Data has the potential to revolutionize modern agriculture by enabling data-driven decision-making, improving crop productivity, and promoting sustainable resource management. The integration of advanced technologies such as artificial intelligence, machine learning, IoT, and remote sensing is transforming agriculture into a more efficient and resilient system. As agricultural data continues to grow in volume and complexity, Big Data analytics will play an increasingly important role in addressing global food security challenges and supporting the transition toward smart and sustainable farming systems.

2. Concept of Big data

Big Data refers to extremely large, complex and rapidly generated datasets that exceed the capacity of conventional data management systems to efficiently store, process and analyze them. The concept is associated with the increasing generation of digital information from multiple sources and the development of technologies capable of storing, integrating and extracting useful information from such datasets. In agriculture, the emergence of Big Data has been closely associated with the development of digital technologies such as sensors, remote sensing, Internet of Things (IoT), cloud computing, unmanned aerial vehicles (UAVs), GPS-enabled machinery and artificial intelligence. These technologies enable the generation and processing of large quantities of agricultural information for improved decision-making (Wolfert et al., 2017).

Agricultural Big Data is generated from a wide range of sources and includes different forms of information. Data may originate from databases, sensors, cameras, GPS systems, remote sensing platforms, weather stations, mobile devices and farm machinery. These datasets can be stored and processed using platforms such as cloud computing systems, Hadoop-based systems and geospatial data-processing platforms. The diversity of data sources and formats is an important feature of agricultural Big Data because agricultural systems generate information at different spatial and temporal scales (Krisnawijaya et al., 2022).

In agricultural production, Big Data is particularly important because crop performance is influenced by the interaction of several factors, including soil properties, weather conditions, crop characteristics, water availability, nutrient supply, pest and disease incidence and management practices. These factors also exhibit considerable spatial and temporal variability. Modern digital technologies allow agricultural systems to be monitored continuously, generating large quantities of information that can be analyzed to identify patterns and relationships among these variables. Such analysis can contribute to a better understanding of agricultural ecosystems and support improved productivity and resource management (Kamilaris et al., 2017).

An important feature of agricultural Big Data is therefore the integration of heterogeneous datasets. For example, soil sensors can provide information on soil moisture and nutrient status, weather stations can provide meteorological observations, UAVs can provide high-resolution crop images, satellites can provide large-scale spatial observations, and farm machinery can generate information on field operations and input application. When these datasets are integrated, they provide a more comprehensive representation of field conditions than any single data source can provide. The integration of these diverse data streams is an important component of modern precision and digital agriculture (Benos et al., 2021; Krisnawijaya et al., 2022).

The nature of agricultural Big Data has also expanded beyond conventional numerical observations. Modern agricultural datasets may contain structured, semi-structured and unstructured information, including numerical sensor readings, geospatial datasets, satellite images, UAV photographs, videos, text, market information and farmer-generated data. The combination of these different data types increases the analytical potential of Big Data but also creates additional requirements for data integration, processing and interpretation (Maffezzoli et al., 2022). The rapid development of IoT technologies has further increased the generation of agricultural Big Data. IoT systems connect sensors, devices and communication networks to enable the continuous collection and transmission of information from agricultural fields. Soil moisture, temperature, humidity, weather and crop-related measurements can be collected at frequent intervals and transmitted to cloud-based platforms for storage and analysis. The integration of IoT with remote sensing, cloud computing and Big Data analytics has therefore become an important component of digital agriculture (Ayaz et al., 2019; Elijah et al., 2018). More recent reviews also identify IoT, wireless sensor networks, remote sensing, cloud computing and Big Data analytics as interconnected components of digital agricultural systems.

Table 1. Data Sources and Analytical Techniques Used in Agricultural Applications (Ghosh et al., 2018)

Data Source	Analysis techniques	Application
Multispectral, RGB, thermal sensors; vegetation indices	Decision Trees	Fertilizer efficiency monitoring
Plant images with weed–soil background	Support Vector Machine (SVM)	Weed detection
Plant images under varying outdoor conditions	k-Nearest Neighbour (k-NN)	Plant detection
Vegetation indices (NDVI, EVI, GNDVI, MCARI)	SVM	Crop and land-use classification
Crop images (healthy vs diseased)	Convolutional Neural Networks (CNN)	Crop disease detection
Time-series disease images and indices	SVM	Early disease detection
C-band radar satellite data	Image processing, Deep learning	Crop type detection

Big Data is not valuable merely because of the large quantity of information generated. Its importance lies in the ability to transform raw observations into useful information and subsequently into actionable agricultural decisions. Big Data analytics involves processing and analyzing large datasets to identify patterns, relationships, trends and predictive information. In agriculture, these analytical approaches can be used for crop yield prediction,

crop monitoring, soil assessment, irrigation management, pest and disease prediction, resource optimization and risk assessment (Kamilaris et al., 2017).

The development of artificial intelligence (AI) and machine learning (ML) has further increased the potential of agricultural Big Data. Machine learning algorithms can analyze large and heterogeneous datasets and identify complex relationships among agricultural variables. Techniques such as random forests, support vector machines, artificial neural networks, decision trees and deep learning have been applied to crop yield prediction, disease detection, soil classification, crop monitoring and resource management. Recent systematic evidence shows that the combination of Big Data and machine learning is increasingly being used across crop, soil and water management to support site-specific agricultural decisions (Ahanou et al., 2026).

Big Data has consequently become an important foundation of Agriculture 4.0 and Smart Farming. Agriculture 4.0 involves the integration of digital technologies such as IoT, Big Data analytics, artificial intelligence, cloud computing, robotics and advanced sensing systems into agricultural production. The objective is to move from conventional and reactive management towards more connected, automated and data-driven production systems (Maffezzoli et al., 2022).

From an agronomic perspective, the concept of Big Data can therefore be viewed as a continuous pathway from data generation to agricultural decision-making. Sensors, satellites, UAVs, weather stations and machinery generate large quantities of information; communication networks and cloud platforms facilitate data transfer and storage; Big Data analytics and AI/ML techniques extract meaningful patterns; and decision-support systems convert these results into recommendations for crop, soil, water and input management. Recent research describes this transition as a movement towards integrated, data-driven precision agriculture in which heterogeneous data are converted into actionable information for site-specific management and resource optimization (Ahanou et al., 2026).

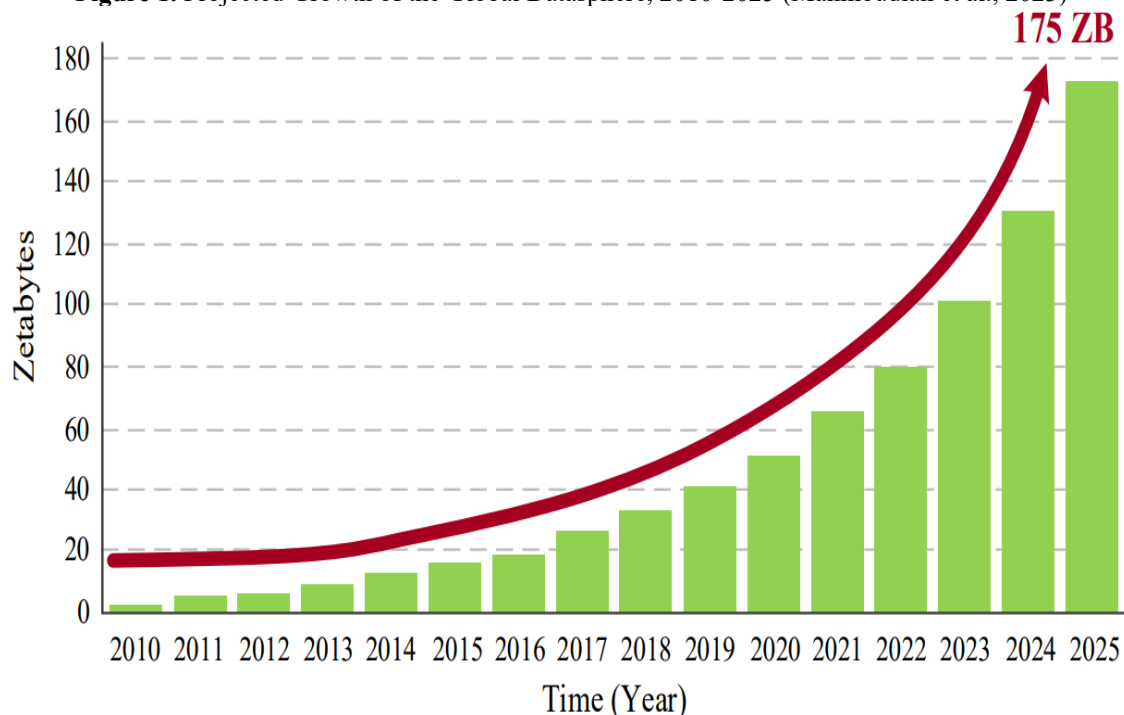
Thus, Big Data in agronomy can be defined as the systematic generation, collection, integration, storage, processing and analysis of large, diverse and continuously generated agricultural datasets to obtain actionable information for improving crop production, resource-use efficiency, risk management and sustainability. It represents not a single technology but an integrated ecosystem connecting data-generation technologies, computational infrastructure, analytical methods and agricultural decision-making.

3. Global Scenario of Big Data Growth

3.1. Growth of Global Data (2010–2025)

The volume of data generated worldwide has increased dramatically over the past decade due to rapid advancements in digital technologies, internet connectivity, cloud computing, AI, and IoT devices. The global digital ecosystem continuously produces enormous datasets from sources such as smartphones, sensors, satellites, social media, online transactions, scientific research, and industrial operations. These massive datasets form the foundation for big data analytics across multiple sectors, including agriculture, healthcare, finance, and transportation. The graph illustrates the global growth of data from 2010 to 2025, measured in zettabytes (ZB). One zettabyte equals one trillion gigabytes, indicating the extremely large scale of data generation worldwide.

Figure 1. Projected Growth of the Global Datasphere, 2010-2025 (Mahmoudian et al., 2023)



Global data growth can be broadly categorized into three major phases

Early Growth Phase (2010-2015)

During this period, global data volume increased from approximately 2 ZB in 2010 to about 16-18 ZB by 2015. Although the growth was relatively moderate compared to later years, this phase established the technological foundation for large-scale data generation and digital transformation.

Rapid Expansion Phase (2016-2020)

These technologies generated massive datasets used in sectors such as environmental monitoring, disaster management, and precision agriculture. As a result, global data volume increased sharply and reached approximately 50 ZB by the year 2020.

Exponential Growth Phase (2021-2025)

According to global estimates, the total volume of digital data is projected to reach approximately 175 ZB by 2025, demonstrating the rapid expansion of the global data ecosystem. This enormous growth of data highlights the importance of advanced data management systems, big data analytics platforms, and high-performance computing infrastructure to effectively store, process, and analyze such massive datasets.

3.2. Distribution of Data Centres in the Asia-Pacific Region

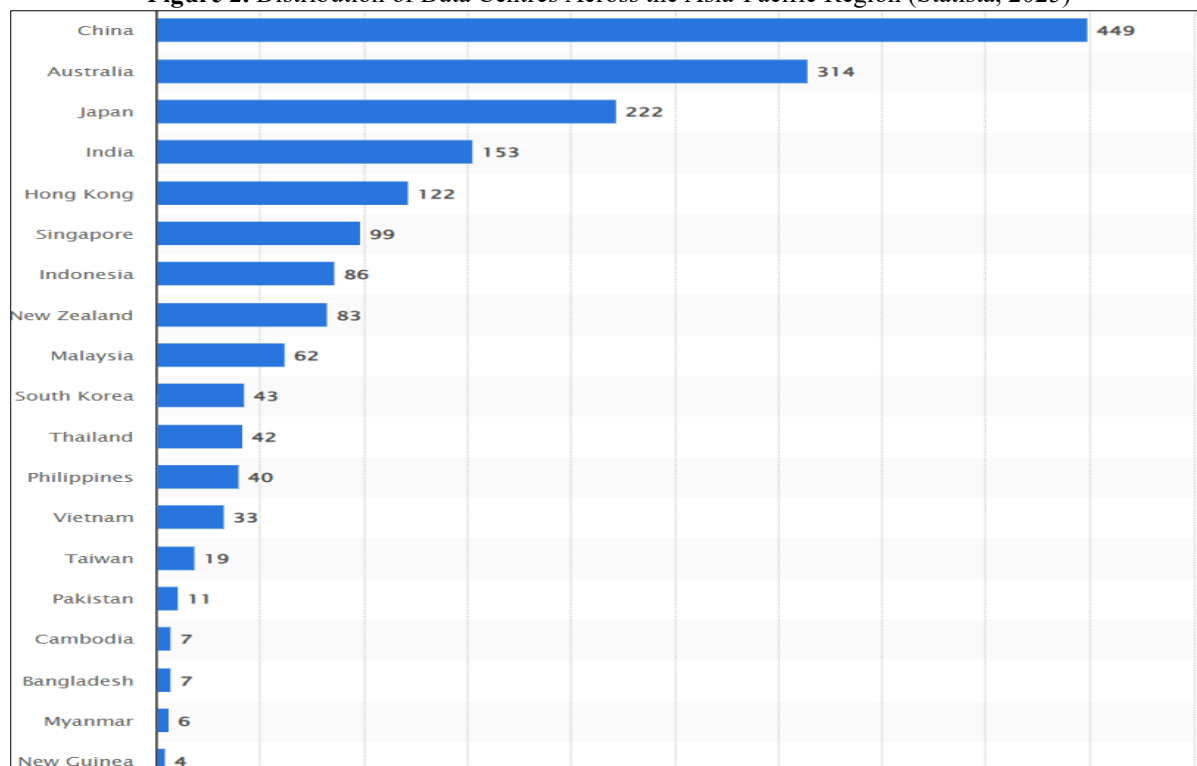
The rapid increase in global data generation requires powerful digital infrastructure capable of storing and processing enormous datasets. Data centres play a critical role in this process. A data centre is a specialized facility that houses computer servers, storage systems, networking equipment, and cooling infrastructure designed to manage large-scale data operations. Data centres serve as the backbone of big data technologies, enabling cloud computing services, artificial intelligence applications, and real-time data analytics. They provide high-capacity storage and computing power required to process massive datasets generated from different sectors, including agriculture.

The chart illustrates the distribution of data centres across countries in the Asia-Pacific region. The Asia-Pacific region has become one of the fastest-growing hubs for digital infrastructure due to increasing internet penetration, rapid economic growth, and expansion of digital services.

Among the countries shown in the chart, China has the largest number of data centres, with approximately 449 facilities. China's dominance is largely due to its massive digital economy, extensive internet user base, and strong investment in cloud computing and artificial intelligence technologies.

Following China, Australia has about 314 data centres, reflecting its strong digital infrastructure and high demand for cloud-based services. Japan ranks third with approximately 222 data centres, supported by its advanced technology sector and digital innovation. India has around 153 data centres, and the number is rapidly increasing due to the growth of digital services, e-commerce platforms, mobile internet usage, and government digital initiatives such as Digital India. India is emerging as a major data hub in the Asia-Pacific region. Other countries such as Hong Kong, Singapore, Indonesia, New Zealand, Malaysia, and South Korea also have significant numbers of data centres. Singapore, in particular, serves as a major regional data hub due to its strong connectivity, stable regulatory environment, and advanced digital infrastructure.

Figure 2. Distribution of Data Centres Across the Asia-Pacific Region (Statista, 2025)



4. Characteristics of Big Data

Big Data is commonly described using a set of core characteristics known as the “5Vs of Big Data.” These characteristics help explain the unique nature of large-scale datasets and the challenges involved in managing and analyzing them. The five major characteristics are Volume, Variety, Velocity, Veracity, and Value (Keskar et al., 2020).

1. Volume

Volume refers to the massive amount of data generated from numerous sources. With the rapid expansion of digital technologies, enormous quantities of data are continuously produced from devices such as smartphones, sensors, satellites, drones, social media platforms, and internet-based applications.

2. Variety

Variety refers to the different types and formats of data that are generated from diverse sources. Unlike traditional data systems that mainly deal with structured data stored in tables and databases, Big Data includes multiple forms of data such as:

Structured data (databases, spreadsheets)
Semi-structured data (XML files, JSON files, log files)
Unstructured data (images, videos, satellite imagery, sensor outputs, social media content)

3. Velocity

Velocity refers to the speed at which data is generated, transmitted, and processed. In modern digital systems, data is produced continuously and often needs to be analyzed immediately to support real-time decision-making.

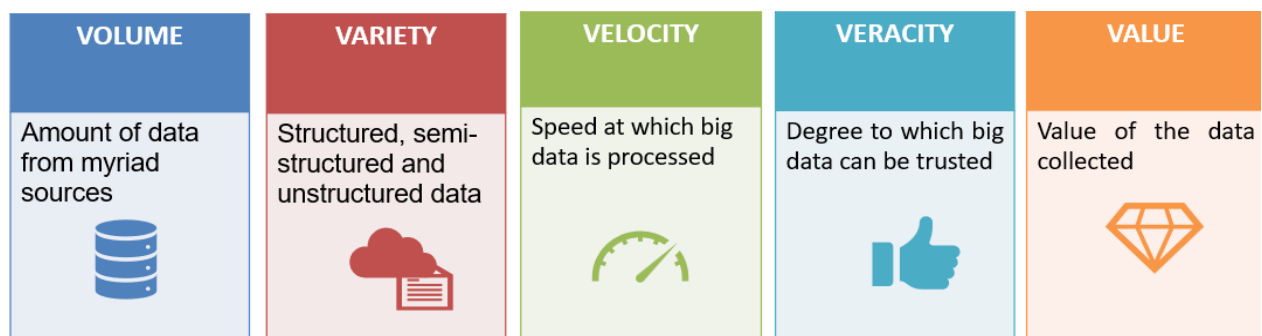
4. Veracity

Veracity refers to the reliability, accuracy, and trustworthiness of data. Because Big Data is collected from multiple sources, the quality of the data can sometimes vary due to measurement errors, missing values, inconsistencies, or noise.

5. Value

Value represents the useful insights and benefits that can be extracted from data. While Big Data involves massive amounts of information, the true importance lies in converting raw data into meaningful knowledge that can support decision-making and problem-solving.

Figure 3. Five Characteristics (5Vs) of Big Data (Keskar et al., 2020)



5. Applications of Big Data in Agronomy

Big Data has emerged as an important component of modern agronomy by enabling the integration and analysis of information collected from diverse sources such as soil sensors, weather stations, satellites, UAVs, farm machinery, field observations and historical agricultural databases. The integration of these datasets with AI, ML, GIS, remote sensing and cloud computing enables the identification of spatial and temporal patterns that are difficult to detect using conventional approaches. Consequently, Big Data analytics is increasingly being applied to climate and weather prediction, pest and disease management, soil and nutrient management, irrigation scheduling, crop yield prediction, precision agriculture, crop insurance and agricultural decision support (Kamilaris et al., 2017; Wolfert et al., 2017).

5.1. Climate Change and Weather Prediction

Climate variability and increasing frequency of extreme weather events such as droughts, floods, heat waves and irregular rainfall create considerable uncertainty in agricultural production. Crop growth and yield are strongly influenced by temperature, rainfall, humidity, solar radiation and other weather variables. Big Data technologies provide opportunities to integrate long-term meteorological observations, satellite information, climate indices and crop data to identify weather patterns and develop predictive models.

Machine learning algorithms can process large historical datasets and establish nonlinear relationships between climatic variables and agricultural outcomes. Such models can be used for rainfall prediction, drought assessment, crop planning and climate-risk management. Weather-based information can further support decisions regarding sowing dates, irrigation scheduling, fertilizer application and crop protection.

An example from Tamil Nadu is the study on monthly rainfall hind-casting for Coimbatore, which used 42 years of monthly rainfall data from 1976-2017. K-means clustering and Principal Component Analysis (PCA) were used

to identify patterns and reduce dimensionality, while Southern Oscillation Index (SOI) and Indian Ocean Dipole (IOD) information was incorporated to account for large-scale climatic influences. Four machine-learning algorithms—Decision Tree Regression, Gradient Boosting, AdaBoost and Random Forest Regression—were evaluated. Random Forest produced the highest reported prediction accuracy of 89.2%, demonstrating the potential of data-driven models for rainfall prediction and climate-resilient agricultural planning (Geethalakshmi et al., 2022).

Similar data-driven approaches have been applied elsewhere in Tamil Nadu and South India to strengthen drought risk assessment and climate-resilient planning. Rahim and Chokkalingam (2025) developed a multi-dimensional drought evaluation framework for Tiruchirappalli District, Tamil Nadu, by integrating Google Earth Engine cloud computing with machine learning and geospatial analysis. Meteorological, hydrological, agricultural, and socioeconomic vulnerability indicators were combined into a composite Multi-Drought Severity Index spanning 2014–2025, and Random Forest-based land-use classification was further used to project future drought manifestation up to 2035, demonstrating how large-scale satellite, groundwater, and socioeconomic datasets can be integrated for regional climate-resilient planning. At a broader agro-climatic scale, Sharma et al. (2025) applied ensemble machine learning models, including Bagging and Random Forest classifiers, to Sentinel-2 multispectral satellite indices to detect drought conditions across three distinct Indian regions, including Thanjavur in Tamil Nadu, during the Rabi cropping season. Twelve vegetation, soil-moisture, and water-stress indices were evaluated, achieving classification accuracies above 83% individually and reaching 94% through seasonal majority voting, with SHAP-based feature attribution identifying the Normalized Multi-band Drought Index and crop growth stage as the most influential predictors. Collectively, these studies illustrate how the integration of satellite remote sensing, meteorological records, and machine learning algorithms enables scalable, near-real-time drought monitoring and early-warning systems, supporting more climate-resilient agricultural planning across diverse agro-climatic regions of India.

5.2. Pest and Disease Management

Pests and diseases are major constraints to agricultural productivity, and their occurrence is influenced by weather conditions, crop growth stage, host characteristics and environmental factors. Conventional pest monitoring often depends on periodic field observations, which may not provide sufficient information for early detection of outbreaks. Big Data analytics enables the integration of weather observations, sensor data, remote sensing imagery, historical pest records and field observations to identify conditions favourable for pest and disease development. Machine-learning and statistical forecasting models can be used to establish relationships between environmental variables and pest populations. Such predictive systems can provide early warnings and allow farmers to implement control measures at the appropriate time. This can reduce unnecessary pesticide applications and improve the efficiency of integrated pest management.

A relevant example is the weather-based prediction of rice yellow stem borer (*Scirpophaga incertulas* Walker) in the Cauvery Delta zone of Tamil Nadu. Yellow stem borer population data obtained from light-trap catches were analysed together with maximum temperature, minimum temperature, relative humidity, wind velocity, evaporation and rainfall. Relative humidity showed a positive association with pest population, whereas maximum temperature, minimum temperature and evaporation showed negative relationships. An ARIMAX model incorporating historical pest observations and weather variables was used for forecasting. The approach demonstrated the potential of integrating pest and weather datasets for early warning and timely pest-management decisions (Anandhi et al., 2025).

Beyond weather-based statistical forecasting, image-based deep learning has become a widely used Big Data approach for pest and disease diagnosis at the field level. Ramcharan et al. (2017) applied transfer learning with deep convolutional neural networks to field-collected cassava leaf images in Tanzania to detect three major diseases and two pest-damage types. The model achieved disease-specific accuracies of up to 98% and an overall accuracy of 93% on unseen field data, demonstrating that mobile-deployable, image-based deep learning models can offer a fast, affordable, and scalable alternative to manual field scouting, particularly relevant for smallholder farming systems in sub-Saharan Africa.

At a transboundary scale, the Food and Agriculture Organization's Desert Locust Information Service (DLIS) illustrates how Big Data integration supports pest early-warning systems over vast geographic areas. Since 1978, DLIS has continuously combined field survey data transmitted via handheld eLocust devices, satellite-derived rainfall and vegetation greenness data, and historical locust records within GIS-based forecasting platforms to predict the timing, scale, and location of locust breeding and migration up to six weeks in advance. More recently, machine-learning-based forecasting tools have been developed to complement these mechanistic models, using remote sensing and historical locust distribution data to improve prediction during recession periods when locust populations are low but breeding risk persists (FAO Desert Locust Information Service, 2024). Collectively, these examples show that Big Data-driven pest and disease management spans a spectrum of approaches, from statistical weather-based forecasting of specific pest species to image-based deep learning diagnostics and continental-scale early-warning platforms, all aimed at enabling timely, data-informed intervention.

5.3. Soil Health and Nutrient Management

Efficient soil and nutrient management are essential for maintaining crop productivity and long-term soil health. Excessive or imbalanced fertilizer application can increase production costs, reduce nutrient-use efficiency and contribute to environmental degradation. Big Data technologies enable the integration of soil-test results, sensor

measurements, remote sensing observations, crop information and historical yield data to develop site-specific nutrient-management recommendations.

IoT-based soil sensors can provide information on soil nutrient and moisture conditions, while machine-learning models can analyse relationships among soil properties, crop requirements and yield. These approaches support variable-rate fertilizer application and help farmers apply nutrients according to spatially variable soil fertility.

An IoT-based nutrient management system developed using NPK sensors, a microcontroller, cloud platform and mobile application demonstrated the potential for real-time soil nutrient monitoring. A Mamdani fuzzy-logic expert system was used to interpret N, P and K levels and generate fertilizer recommendations. Such systems can improve nutrient-use efficiency by converting sensor observations into practical recommendations for farmers (Madhumathi et al., 2022).

Another example is GeaGrow, a mobile-based decision-support system that uses soil and environmental datasets with Artificial Neural Networks (ANN) to predict soil nutrient status and provide crop-specific fertilizer recommendations. The system incorporates SCORPAN-based feature extraction and ANN modelling and provides recommendations for crops such as yam, maize, cassava and rice. The approach demonstrates how Big Data and machine learning can support site-specific nutrient management and reduce excessive fertilizer application (Folorunso et al., 2025).

Digital soil mapping using machine learning has also been applied to generate spatially explicit soil information at regional scale. Kumaraperumal et al. (2022) compared multiple machine-learning algorithms for qualitative and quantitative digital soil mapping across the eastern districts of Tamil Nadu, integrating soil survey data with terrain and environmental covariates to predict the spatial distribution of key soil attributes. Their findings demonstrated that machine-learning-based digital soil maps can capture within-region soil variability more effectively than conventional soil survey methods, providing a scalable basis for site-specific nutrient and land-management recommendations across large geographic areas where field-level soil testing is not feasible for every farm. Collectively, these examples illustrate a progression of Big Data applications in soil and nutrient management, ranging from real-time, sensor-driven fertilizer recommendations at the individual-field level, to mobile-based decision-support tools for smallholder farmers, to regional-scale digital soil mapping that supports land-management planning across entire districts

5.4. Water Management and Irrigation Optimization

Water scarcity, irregular rainfall and increasing crop water demand have increased the need for efficient irrigation management. Big Data analytics can combine weather observations, rainfall, soil moisture, crop growth information, evapotranspiration and irrigation records to estimate crop water requirements and improve irrigation scheduling.

Weather-based models can estimate evapotranspiration and irrigation requirements at different crop growth stages. When combined with soil sensors and remote sensing, these datasets can support precision irrigation by identifying spatial and temporal variations in crop water demand. Consequently, Big Data can contribute to improved water-use efficiency while maintaining crop productivity.

A study on cotton evapotranspiration, irrigation requirements and water productivity evaluated rainfall, irrigation, effective rainfall and crop evapotranspiration across three cropping seasons. Season 2 received the highest irrigation of 336.1 mm and had the highest adjusted crop evapotranspiration of 607.2 mm, while Season 3 recorded the highest seed cotton yield of 2237 kg ha⁻¹ and lint yield of 761.7 kg ha⁻¹. Variation in water-use efficiency among seasons demonstrated the importance of matching irrigation supply with actual crop water requirements (Madhavan et al., 2025).

Sensor-driven, machine-learning-based irrigation scheduling has also demonstrated measurable water savings in precision agriculture settings. Maniam et al. (2026) developed an IoT-enabled smart irrigation management system using subsurface time-domain reflectometry soil moisture probes at multiple depths, connected via a wireless sensor network to a cloud platform, combined with a recurrent neural network-long short-term memory (RNN-LSTM) model for real-time soil moisture prediction. Deployed over a six-month field trial, the system achieved a prediction accuracy within $\pm 5\%$ for over 95% of observations and reduced overall water usage by approximately 30%, illustrating how continuous sensor data streams combined with deep learning can directly translate into more efficient irrigation scheduling and reduced production costs.

At a larger platform scale, Kamienski et al. (2019) developed the SWAMP (Smart Water Management Platform), an IoT-based precision irrigation system integrating soil, weather, and crop data across multiple agricultural pilot sites in Brazil and Europe. The platform combined heterogeneous sensor networks with a scalable, replicable software architecture (built on FIWARE components) to support real-time, data-driven irrigation decisions across diverse cropping systems and geographic contexts. This example highlights how Big Data platforms can be designed not just for single-farm optimization, but for scalable deployment across varied agro-climatic regions, supporting broader adoption of precision irrigation technologies.

Beyond sensor networks and IoT platforms, remote sensing-derived evapotranspiration estimates offer a complementary Big Data pathway for irrigation scheduling, particularly where ground-based meteorological data are limited or unavailable. Fong et al. (2025) developed deep learning models - including long short-term memory (LSTM), gated recurrent unit (GRU), and hybrid CNN-LSTM and CNN-GRU architectures - to estimate reference evapotranspiration (ET₀) using MODIS-derived land surface temperature, vegetation indices, surface reflectance bands and downward shortwave radiation as inputs, at two stations in Peninsular Malaysia. The hybrid CNN-

GRU model achieved the highest predictive accuracy ($R^2 = 0.805$), outperforming individual LSTM and GRU models, and demonstrated that satellite-derived variables can substitute for extensive ground-based meteorological networks in estimating crop water demand. This example illustrates how Big Data derived purely from satellite platforms can extend reliable irrigation scheduling support to data-scarce agricultural regions that lack dense ground-sensor infrastructure.

Collectively, these examples show that Big Data-driven water management spans field-level empirical water-productivity assessment, sensor-and-deep-learning-based irrigation control, satellite-based evapotranspiration modelling in data-scarce regions, and multi-site platform-scale deployment, all converging on the shared goal of matching irrigation supply more precisely to actual crop water demand.

5.5. Crop Yield Prediction

Crop yield prediction is one of the most important applications of Big Data in agriculture because reliable pre-harvest yield estimates are useful for farm planning, food-supply management, market decisions and agricultural policy. Yield prediction models can integrate weather data, soil characteristics, crop growth parameters, remote sensing observations, historical yield records and management information.

Remote sensing provides repeated observations of crop growth over large areas, while machine-learning algorithms can identify relationships between vegetation characteristics and final yield. The integration of these datasets allows yield to be predicted at field, farm, district or regional scales.

A study on cotton yield prediction using drone-derived Leaf Area Index (LAI) and chlorophyll content demonstrated this approach. Multispectral UAV imagery was used to obtain red, near-infrared and red-edge information. Vegetation indices including WDRVI and MCARI were used to estimate LAI and chlorophyll content, respectively. Regression relationships were established between remotely sensed crop parameters and seed cotton yield, and GIS was used to generate pixel-level spatial yield prediction maps. LAI and chlorophyll content showed positive relationships with cotton yield, demonstrating the potential of UAV-based Big Data analytics for field-scale yield prediction (Shanmugapriya et al., 2022).

Several other studies illustrate this same Big Data-driven approach across different crops and spatial scales. In rice, UAV-based multispectral and digital imagery was used to derive vegetation indices at multiple crop growth stages, with the booting stage identified as optimal for yield estimation; NDVI calculated from near-infrared and red bands showed a strong relationship with grain yield ($R^2 = 0.75$), confirming that carefully timed UAV acquisition can substantially improve pre-harvest yield estimates (Zhou et al., 2017). Building on this field-scale approach, a national-scale study integrated multiple heterogeneous data sources - climate variables, soil properties, and satellite-derived vegetation indices - across China, comparing conventional machine learning with deep learning models for rice yield prediction. The results showed that combining diverse Big Data sources and applying deep learning architectures improved accuracy over single-source, single-model approaches, particularly by capturing nonlinear relationships between inputs and yield (Cao et al., 2021).

Similar advances have been demonstrated for maize. A Long Short-Term Memory (LSTM) deep learning model was developed to predict county-level maize yield across the U.S. Corn Belt using only publicly available weather and remote sensing data, generating monthly pre-harvest forecasts that closely matched official USDA estimates and were made available through a public forecasting tool, highlighting the potential of Big Data-driven deep learning to democratize access to reliable yield forecasts (Jiang et al., 2020). Addressing the challenge of large, spatially heterogeneous growing regions, another study developed a deep learning model for county-level maize yield prediction across China using multiple satellite-derived variables, incorporating an over-complete dictionary matrix to represent inter-province variability and thereby reducing estimation error compared with conventional regression-based methods (Li et al., 2022).

Comparable approaches have also been applied to wheat, where synthesized Landsat and MODIS time-series imagery spanning two decades was used to predict yield at both county and pixel levels. Vegetation indices such as NDVI, NIRv, and kernel NDVI (kNDVI) were compared as predictors, with regression and LSTM-based models showing that emerging vegetation indices offered modest but consistent improvements over traditional NDVI-based approaches (Zhang et al., 2024).

Together, these studies - spanning cotton, rice, maize, and wheat, and ranging from individual fields to entire countries - demonstrate that integrating remote sensing (UAV or satellite), vegetation indices, and machine learning or deep learning algorithms consistently improves the accuracy and scalability of pre-harvest yield prediction compared with traditional survey- or model-based approaches.

5.6. Precision Agriculture and Site-Specific Management

Agricultural fields are rarely uniform. Soil fertility, moisture, crop growth and yield potential can vary considerably within a single field. Uniform application of seed, fertilizer, irrigation water and pesticides may therefore result in inefficient resource use. Big Data supports precision agriculture by integrating geospatial information, sensor observations, remote sensing and machinery data to identify field variability.

Multi-sensor data fusion is particularly useful because information from different sensors can be combined to develop a more comprehensive representation of field conditions. The resulting information can be used for variable-rate seeding, fertilizer application, irrigation and other site-specific management practices.

A multi-sensor data-fusion approach for site-specific potato seeding compared site-specific seeding (SSS) with uniform-rate seeding (URS). In Site 2, yield increased from 55.22 t ha⁻¹ under URS to 56.66 t ha⁻¹ under SSS. Site 3 showed the highest increase in net profit, approximately 27.21%, despite having the lowest absolute yield.

This demonstrates that site-specific management can improve resource-use efficiency and economic returns, particularly in heterogeneous fields (Munnaf et al., 2021).

Building on this same multi-sensor data-fusion approach, a related study extended site-specific management to maize production, delineating management zone maps from fused proximal soil sensor and remote sensing data to guide site-specific seeding rates. As with the earlier potato study, zone-based seeding produced measurable improvements in yield uniformity and resource-use efficiency compared with uniform-rate seeding, reinforcing that management-zone-based variable-rate seeding is transferable across different crops when the same underlying multi-sensor fusion workflow is applied (Munnaf et al., 2022).

Site-specific management has also been extended beyond seeding to variable-rate crop protection. A recent study on wheat production integrated UAV-based multispectral imagery to generate weed distribution maps, which were used to simulate variable-rate application (VRT) of herbicides and fertilizers, and compared this against conventional uniform-rate application using a full life-cycle assessment. The VRT approach reduced nitrogen use by 22.2%, phosphorus by 38.9%, potassium by 77%, and herbicide use by 82.4%, while lowering global warming potential by roughly 20% and freshwater ecotoxicity by 65%, without compromising crop performance. This example illustrates that Big Data-enabled site-specific management can generate substantial environmental and input-efficiency gains, not just yield or profit gains, reinforcing that precision agriculture's benefits extend across economic, agronomic and environmental dimensions (Jalilian et al., 2025).

Collectively, these examples show that Big Data-enabled precision agriculture - from multi-sensor management-zone delineation to UAV-guided variable-rate application - allows growers to move beyond uniform field management toward input strategies tailored to actual within-field variability, improving yield, profitability and environmental performance simultaneously.

5.7. Crop Insurance and Agricultural Risk Assessment

Crop insurance requires reliable information on cultivated area, crop condition, yield and crop losses. Traditional assessment methods may depend heavily on field surveys, which can be time-consuming and difficult to implement over large geographical areas. Big Data technologies, particularly satellite remote sensing, GIS, weather datasets and crop-growth models, provide an alternative approach for large-scale crop monitoring and loss assessment. Time-series satellite observations can detect changes in crop condition and identify areas affected by prevented sowing, crop failure, drought or other disturbances. When satellite observations are integrated with crop-growth models and weather information, spatial yield estimates can also be generated to support insurance assessment.

Pazhanivelan et al. (2019) integrated Sentinel-1A Synthetic Aperture Radar (SAR) data with the MAPscape-Rice system and the ORYZA2000 crop-growth model for rice mapping and yield estimation in the Cauvery Delta region of Tamil Nadu. Rice-area classification achieved 90.3-94.2% accuracy based on field validation. Estimated rice yields ranged from 2022 to 5494 kg ha⁻¹. The approach also identified prevented or failed sowing and total crop failure, demonstrating the usefulness of satellite Big Data for crop insurance assessment.

In a similar vein, Mandapati et al. (2023) integrated Sentinel-2 optical satellite data with the DSSAT CERES-Rice crop-growth model to estimate village-level rice yields in Karimnagar district, Telangana, India, explicitly for use in insurance premium pricing and loss assessment. Rice-growing area was mapped using supervised classification with an overall accuracy of 93.04% ($\kappa = 0.87$), closely matching government statistics (1.43 vs. 1.47 lakh ha). Leaf area index (LAI), measured in the field and simulated by DSSAT, correlated strongly with Sentinel-2-derived NDVI ($R^2 = 0.82$), and this relationship was used to generate spatial LAI and yield maps. Model-simulated yields agreed well with observed field yields ($R^2 = 0.80$, RMSE = 658 kg ha⁻¹, D-index = 0.88), and the district-average simulated yield (5350 kg ha⁻¹) closely matched official statistics (5014 kg ha⁻¹), reinforcing the value of assimilating remote sensing data with crop growth models for insurance-relevant yield prediction.

Beyond crop-specific applications, Vrieling et al. (2016) demonstrated the broader utility of satellite vegetation indices for index-based agricultural risk insurance in pastoral systems across Kenya and southern Ethiopia. Under an operational scheme, payouts are triggered when a “forage scarcity index” derived from 250 m, 10-day MODIS NDVI time series (2001-2014) falls below a drought threshold. By incorporating phenological modelling to define location-specific growing-season windows, the study showed that payout determination could be advanced by 2-3 months without compromising the index's reliability, enabling earlier, loss-mitigating interventions. Collectively, these three studies illustrate how satellite Big Data - whether SAR-based crop mapping, optical NDVI-LAI-crop model integration, or long-term vegetation index time series - can support faster, more objective, and more scalable assessment of cultivated area, crop condition, yield and losses than conventional field-survey-based approaches, forming a robust technological basis for modern crop and agricultural risk insurance.

5.8. Weed Detection and Management

Weeds compete with crops for light, nutrients and water, and are among the leading causes of yield loss worldwide. Conventional weed management relies on uniform, calendar-based herbicide spraying across entire fields, which is inefficient, costly, and environmentally damaging, since it applies chemicals regardless of whether weeds are actually present at a given location. Big Data technologies - particularly UAV-based multispectral and RGB imagery combined with deep learning and computer vision - now allow weeds to be detected, mapped and distinguished from crop plants at high spatial resolution, enabling targeted, site-specific herbicide application (precision spraying) rather than blanket treatment. This reduces herbicide volumes, cuts input costs, limits the

development of herbicide-resistant weed biotypes, and lowers the environmental and health burden of chemical use.

Kebede et al. (2025) developed a UAV- and deep-learning-based weed detection system for teff (*Eragrostis tef*), a staple cereal crop in Ethiopia where weeds are difficult to distinguish from the crop due to similar leaf morphology. Using a dataset of 1,308 high-resolution drone images collected across different growth stages and weather conditions at the University of Gondar Agricultural Research Farm, the authors evaluated several deep learning architectures - including MobileNetV2, InceptionResNetV2, DenseNet201, VGG16, ResNet50, Faster R-CNN and YOLOv8 - with and without fine-tuning, and incorporated shape-based features such as aspect ratio and solidity to improve discrimination between teff and weeds. The fine-tuned MobileNetV2 model achieved the highest classification accuracy, at 96.40%, demonstrating that lightweight, UAV-deployable deep learning models can reliably support selective, precision herbicide application even for crop–weed pairs that are visually very similar.

Seiche, Wittstruck and Jarmer (2024) took a complementary approach, focusing on making UAV-based weed detection more accessible by comparing a self-built, low-cost multispectral camera system against a high-end commercial sensor (MicaSense Altum) for weed mapping in maize fields. Using a U-Net semantic segmentation model on imagery from both sensors, the weed class was detected with an F1-score of 82% on the high-end system and 76% on the low-cost system, with precision reaching 90% on the low-cost setup despite some misclassification of small or overlapping weed patches. The study showed that affordable, self-assembled multispectral sensors can approach the performance of expensive commercial systems for weed classification, lowering the cost barrier for smallholder and resource-limited farmers to adopt precision weed-management technology.

Together, these studies illustrate two complementary directions in Big Data-enabled weed management: improving detection accuracy through advanced deep learning architectures on standard high-resolution imagery (Kebede et al., 2025), and improving accessibility and affordability of the sensing hardware itself without a major loss in detection performance (Seiche et al., 2024). Both point toward UAV-based, AI-driven weed mapping as a scalable alternative to manual scouting and uniform herbicide application, supporting more sustainable and cost-effective crop protection.

6. Advantages of Big Data Analytics

In agriculture and other sectors, Big Data analytics provides significant benefits by improving productivity, optimizing resource use, reducing risks, and enabling real-time monitoring of systems (Smaya, 2022). These benefits are outlined below.

Figure 4. Key Benefits of Big Data Analytics (Smaya, 2022)

Improved Efficiency and Productivity



By integrating data from multiple sources - satellite imagery, sensors, weather stations and field records - Big Data analytics allows farmers and agencies to identify inefficiencies and act on them precisely rather than uniformly. For instance, satellite-based crop and area mapping, as demonstrated by Pazhanivelan et al. (2019) and Mandapati et al. (2023), removes the need for exhaustive manual field surveys while still delivering accurate, field-level information, substantially reducing the time and labour required for crop assessment.

Enables Predictive Analysis and Forecasting

The integration of time-series satellite data with crop-growth models (e.g., ORYZA2000, DSSAT) allows analysts to forecast yields, anticipate crop failure, and predict pest or drought stress before they fully manifest. This predictive capacity, illustrated in the yield-forecasting work of Mandapati et al. (2023) and the early-warning forage-scarcity index of Vrieling et al. (2016), shifts agricultural risk management from reactive to anticipatory, giving stakeholders lead time to intervene.

Enhances Risk Assessment

Big Data analytics strengthens the accuracy and objectivity of risk assessment by replacing subjective or delayed field reporting with continuous, data-driven evidence. In crop insurance, this is evident in the use of satellite-derived indices to detect prevented sowing, crop failure and drought impact (Pazhanivelan et al., 2019), and in NDVI-based index insurance schemes that quantify drought risk directly from vegetation condition (Vrieling et al., 2016), enabling more transparent and defensible loss and premium calculations.

Support real time monitoring

High-revisit-frequency satellite constellations such as Sentinel-1 and Sentinel-2, together with UAV-based sensing platforms, enable frequent monitoring of crop condition, weed infestation, and field-level anomalies throughout the growing season. This capability supports timely agricultural decision-making and precision management. For example, UAV-based deep learning approaches have been used for weed detection in teff, where a fine-tuned MobileNetV2 model achieved 96.40% accuracy (Kebede et al., 2025), while multispectral UAV imagery combined with U-Net deep learning demonstrated effective weed detection for precision weed management (Seiche et al., 2024)

Optimises Resource Utilization

Rather than applying inputs such as water, fertilizer, or herbicide uniformly across a field, Big Data-driven precision agriculture enables site-specific application based on observed spatial and temporal variability. In weed management, UAV-based multispectral sensing can generate detailed weed maps that support targeted herbicide application. Seiche et al. (2024) demonstrated the potential of both high-end and low-cost multispectral UAV sensors for accurate weed detection, supporting site-specific weed control and more efficient herbicide use. Such targeted application can reduce unnecessary chemical inputs, thereby lower environmental impacts and improving resource-use efficiency.

Strengthens Planning and Strategic Decision-Making

At the policy and institutional level, aggregated Big Data outputs, such as spatial yield maps, crop-loss assessments, and drought indicators, support evidence-based decisions related to crop insurance, agricultural planning, and resource allocation. Mandapati et al. (2023) demonstrated that integrating Sentinel-2 remote sensing data with crop simulation models enabled spatial rice yield estimation and improved yield prediction for insurance companies and policy- and decision-makers. At the national level, India's Yield Estimation System based on Technology (YES-TECH) under the Pradhan Mantri Fasal Bima Yojana (PMFBY) incorporates technology-based yield estimates into yield-loss and insurance-claim assessment, with nationwide rollout initiated from Kharif 2023.

7. Disadvantages and Challenges of Big Data

While Big Data analytics offers substantial benefits for productivity, forecasting, and risk management, its implementation is constrained by a range of technical challenges spanning the full data lifecycle, from analysis to storage and security (Mahmoudian et al., 2023).

Figure 5. Major Challenges of Big Data Analytics (Mahmoudian et al., 2023)



Analysis

Extracting meaningful insight from Big Data requires advanced analytical methods, particularly machine learning and artificial intelligence, capable of identifying patterns across large, heterogeneous datasets. Processing datasets of this scale is computationally intensive, often demanding significant processing power and specialized infrastructure. Since the reliability of downstream predictions and decisions depends directly on the accuracy of this analysis, ensuring robust and well-validated analytical methods is essential.

Redundancy Reduction

Big datasets frequently contain duplicate, overlapping, or otherwise unnecessary information, particularly when data are aggregated from multiple sensors, platforms, or sources. Reducing such redundancy improves data-management efficiency by decreasing unnecessary storage and processing requirements. Data cleaning and preprocessing techniques can be applied to identify and remove duplicate or low-value information before subsequent analysis. In agricultural Big Data systems, this is particularly important when information from remote sensing, IoT sensors, UAVs, weather stations, and farm machinery is integrated into a common analytical framework (Mahmoudian et al., 2023).

Data Compression

The sheer volume of Big Data places substantial demands on storage capacity and network bandwidth. Data compression techniques reduce the size of data files while retaining the information required for analysis, thereby improving storage efficiency and reducing data-transmission requirements. This is particularly important in agricultural IoT systems, where large volumes of sensor observations must be transmitted through resource-constrained communication networks. Podnar et al. (2024) demonstrated the usefulness of compression algorithms for reducing data traffic and improving resource efficiency in agricultural IoT applications.

Data Lifecycle Management

Big Data must be actively managed across its entire lifecycle—from creation and storage through processing and eventual archiving or disposal. Proper lifecycle management ensures that data remain of high quality, accessible, and secure throughout their use. It also helps organizations maintain the long-term usability of agricultural datasets rather than allowing information to become disorganized, outdated, or difficult to manage (Mahmoudian et al., 2023).

Scalability

As data volumes and processing demands grow, Big Data systems must be able to scale accordingly. Scalability ensures that storage and computing capacity can expand to meet increasing requirements without substantial loss of performance. Distributed computing frameworks play an important role by enabling large datasets to be processed across multiple computing nodes rather than relying on a single system (Mahmoudian et al., 2023).

Collaboration

Big Data initiatives frequently involve multiple organizations, institutions, and researchers working with shared or interconnected datasets. Effective collaboration is therefore essential for meaningful data sharing and joint analysis. The use of standardized data formats and interoperable platforms can facilitate the exchange and integration of information between different stakeholders and reduce difficulties caused by incompatible systems or data structures (Mahmoudian et al., 2023).

Confidentiality of Information

Because Big Data may contain sensitive or private information, ensuring its security and confidentiality is a major challenge. Protecting such information requires appropriate safeguards, including secure data-management systems, access controls, and other security mechanisms, to reduce the risk of unauthorized access, misuse, or data breaches (Mahmoudian et al., 2023).

Energy Management

Processing and storing Big Data can require substantial computational and infrastructure resources. As data volumes increase, the energy requirements associated with data processing, storage, and supporting infrastructure also become important considerations. Therefore, improving energy efficiency in Big Data infrastructure is necessary for reducing resource consumption and supporting the long-term sustainability of data-intensive systems (Mahmoudian et al., 2023).

CONCLUSION

Big Data is playing a transformative role in modern agriculture by enabling data-driven and precision-based farming practices. The integration of technologies such as IoT sensors, remote sensing, drones, cloud computing, and machine learning allows the collection and analysis of large volumes of agricultural data from multiple sources. These technologies help monitor crop growth, soil conditions, weather patterns, and farm operations in real time.

Through data preprocessing, integration, feature extraction, and advanced analytical techniques, raw agricultural data can be converted into meaningful insights. Tools such as machine learning models, big data platforms, and visualization systems support effective analysis and interpretation of complex datasets. The implementation of Decision Support Systems (DSS) further enables farmers and agricultural researchers to make informed and timely management decisions.

Overall, Big Data technologies contribute to improving crop productivity, optimizing resource use, enhancing pest and disease management, and strengthening agricultural planning and risk management. Although challenges such as data integration, infrastructure limitations, and digital literacy remain, continuous technological advancements are making big data applications more accessible. Therefore, Big Data has great potential to support sustainable agriculture, precision farming, and improved food security in the future.

Acknowledgements

The authors would like to express their deepest gratitude to Tamil Nadu Agricultural University (TNAU) for providing the necessary facilities and support to carry out this work. Special thanks are extended to the Chairman

and advisory committee members for their valuable feedback, guidance, and constructive suggestions that greatly improved the quality of the manuscript.

Authors contributions

KA conceptualized the review framework and prepared the original draft of the manuscript. RM provided overall guidance, refined the research theme, and critically reviewed and approved the final manuscript. GG assisted in structuring the review, improving technical content, and revising the manuscript. KP and AR contributed expert insights and critically evaluated the manuscript. All authors reviewed and approved the final version of the manuscript before submission.

Availability of data and material All information used in this study was obtained from previously published sources, including peer-reviewed journal articles, and established academic databases such as Scopus and Web of Science. This study did not generate or analyze any new primary datasets.

Compliance with ethical standards

Conflict of interest: The authors do not have any conflicts of interest to declare.

Ethical Issues: None

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used Grammarly to improve the language. After using this tool/service, the author(s) reviewed and edited the content as needed and takes full responsibility for the content of the publication

REFERENCES

1. Ahanou, Z., Mrabti, F. & Dhassi, Y. (2026). Big data analytics in precision agriculture: An automated systematic literature review. *Computers and Electronics in Agriculture*, 243: 111401. DOI: 10.1016/j.compag.2025.111401.
2. Alahmad, T., Neményi, M., & Nyéki, A. (2023). Applying IoT sensors and big data to improve precision crop production: a review. *Agronomy*, 13(10), 2603
3. Anandhi, P., G. Leena, P. S. Ganapati, S. Elamathi, K. Subrahmaniyan, and S. Venugopal. "Weather based prediction model for rice yellow stem borer (*Scirpophaga incertulas* Walker) under changing climate scenarios in Cauvery delta zone, Tamil Nadu." *Journal of environmental biology* 46, no. 1 (2025): 95-102.
4. Anitha, R., Ramathilagam, A., Jeyavelan, M., Pandi, C. K., & Karthieswaran, R. A. (2024). A Machine Learning Approach for Recommending Soil Nutrient Levels. 2024 2nd International Conference on Networking and Communications (ICNWC), 1–6. <https://doi.org/10.1109/ICNWC60771.2024.10537577>
5. Ayaz, M., Ammad-Uddin, M., Sharif, Z., Mansour, A., & Aggoune, E.-H. M. (2019). Internet-of-Things (IoT)-Based Smart Agriculture: Toward Making the Fields Talk. *IEEE Access*, 7, 129551–129583. <https://doi.org/10.1109/ACCESS.2019.2932609>. □
6. Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors*, 21(11), 3758. <https://doi.org/10.3390/s21113758>
7. Bhat, S. A., & Huang, N. F. (2021). Big data and AI revolution in precision agriculture: Survey and challenges. *IEEE Access*, 9, 110209–110222.
8. Cao, J., Zhang, Z., Tao, F., Zhang, L., Luo, Y., Zhang, J., Han, J., & Xie, J. (2021). Integrating multi-source data for rice yield prediction across China using machine learning and deep learning approaches. *Agricultural and Forest Meteorology*, 297, 108275.
9. Elijah, O., Rahman, T. A., Orikumhi, I., Leow, C. Y., & Hindia, M. H. D. N. (2018). An overview of Internet of Things (IoT) and data analytics in agriculture: Benefits and challenges. *IEEE Internet of Things Journal*, 5(5), 3758–3773. <https://doi.org/10.1109/JIOT.2018.2844296>
10. FAO Desert Locust Information Service. (2024). Desert Locust Information Service (DLIS). Food and Agriculture Organization of the United Nations. <https://www.fao.org/locust-watch/activities/dlis-home/dlis/en>
11. Folorunso, O., Ojoo, O., Busari, M., Adebayo, M., Adejumbi, J., Folorunso, D., Ayo, F., Alabi, O., & Olabanjo, O. (2025). GeaGrow: A mobile tool for soil nutrient prediction and fertilizer optimization using artificial neural networks. *Frontiers in Sustainable Food Systems*, 9, 1533423. <https://doi.org/10.3389/fsufs.2025.1533423>
12. Fong, T. Y., Huang, Y. F., Chin, R. J., & Koo, C. H. (2025). Enhanced estimation of reference evapotranspiration using hybrid deep learning models and remote sensing variables. *Agricultural Water Management*, 315, 109534.
13. Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144.
14. Geethalakshmi, V., S. Kokilavani, S. P. Ramanathan, G. A. Dheebakaran, N. K. Sathyamoorthy, and N. Maragatham. "Monthly rainfall hind-cast using machine learning algorithms for Coimbatore, Tamil Nadu." *Mausam* 73, no. 1 (2022): 19-26.
15. Ghosh, A., Nashaat, M., Miller, J., Quader, S., & Marston, C. (2018). A comprehensive review of tools for exploratory analysis of tabular industrial datasets. *Visual Informatics*, 2(4), 235-253
16. Hiremath, S., Wittke, S., Palosuo, T., Kaivosoja, J., Tao, F., Prohl, M., et al. (2021). Crop loss identification at field parcel scale using satellite remote sensing and machine learning. *PLOS ONE*.

17. Jalilian, A., Ghasemi, M. M., Ghaznavi, S., Khayat Moghadam, M. S., Ghasemi Mobtaker, H., Dastbala, G., & Kaab, A. (2025). Evaluating the effectiveness of variable rate technology of herbicides and chemical fertilizers on reducing environmental pollutants, economic costs and energy use in wheat production. *Results in Engineering*.
18. Jiang, Z., Liu, C., Ganapathysubramanian, B., Hayes, D. J., & Sarkar, S. (2020). Predicting county-scale maize yields with publicly available data. *Scientific Reports*, 10, 14957.
19. Kamienski, C., Soininen, J.-P., Taumberger, M., Dantas, R., Toscano, A., Salmon Cinotti, T., Filev Maia, R., & Torre Neto, A. (2019). Smart water management platform: IoT-based precision irrigation for agriculture. *Sensors*, 19(2), 276. <https://doi.org/10.3390/s19020276>
20. Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017). A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture*, 143, 23–37. <https://doi.org/10.1016/j.compag.2017.09.037>
21. Kebede, A.S., Muluneh, T.W., & Adege, A.B. (2025). Detection of weeds in teff crops using deep learning and UAV imagery for precision herbicide application. *Scientific Reports*, 15, 30708. <https://doi.org/10.1038/s41598-025-15380-3>
22. Keskar, V., Yadav, J., & Kumar, A. H. (2020). 5V's of Big Data Attributes and their Relevance and Importance across Domains. *International Journal of Innovative Technology and Exploring Engineering*, 9(11), 154-163
23. Khan, H. H., Malik, M. N., Konečná, Z., Chofreh, A. G., Goni, F. A., & Klemeš, J. J. (2022). Blockchain technology for agricultural supply chains during the COVID-19 pandemic: Benefits and cleaner solutions. *Journal of Cleaner Production*, 347, 1–13.
24. Krisnawijaya, A.A.K., et al. (2022). Data type and data sources for agricultural Big Data and machine learning. *Sustainability*, 14(23): 16131.
25. Kumaraperumal, R., Pazhanivelan, S., Geethalakshmi, V., Raj, M. N., Muthumanickam, D., Kaliaperumal, R., Shankar, V., Nair, A. M., Yadav, M. K., & Kshatriya, T. V. T. (2022). Comparison of machine learning-based prediction of qualitative and quantitative digital soil-mapping approaches for eastern districts of Tamil Nadu, India. *Land*, 11(12), 2279. <https://doi.org/10.3390/land11122279>
26. Li, X., Geng, H., Zhang, L., Peng, S., Xin, Q., Huang, J., Li, X., Liu, S., & Wang, Y. (2022). Improving maize yield prediction at the county level from 2002 to 2015 in China using a novel deep learning approach. *Computers and Electronics in Agriculture*, 202, 107356.
27. Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8), 2674. <https://doi.org/10.3390/s18082674>
28. M. Mahmoudian, S. M. Zanjani, H. Shahinzadeh, Y. Kabalci, E. Kabalci and F. Ebrahimi, "An Overview of Big Data Concepts, Methods, and Analytics: Challenges, Issues, and Opportunities," 2023 5th Global Power, Energy and Communication Conference (GPECOM), Nevsehir, Turkiye, 2023, pp. 554-559, doi: 10.1109/GPECOM58364.2023.10175760.
29. Madhavan, Gnanasekaran, Sellaperumal Pazhanivelan, A. P. Sivamurugan, Dhanaraju Muthumanickam, Selvaraj Selvakumar, and Kaliaperumal Ragunath. "Quantitative assessment of cotton evapotranspiration, irrigation requirements and water productivity from weather-based estimations." *Plant Science Today* 12 (2025): 11843.
30. Madhumathi, R., Arumuganathan, T., Shruthi, R., & Raghavendar, S. (2022, March). Soil NPK prediction using multiple linear regression. In 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS) (Vol. 1, pp. 542-546). IEEE.
31. Maffezzoli, F., Ardolino, M., Bacchetti, A., Perona, M. & Renga, F. (2022). Agriculture 4.0: A systematic literature review on the paradigm, technologies and benefits. *Futures*, 142: 102998. DOI: 10.1016/j.futures.2022.102998.
32. Mandapati, R., Gumma, M.K., Metuku, D.R., Bellam, P.K., Panjala, P., Maitra, S., & Maila, N. (2023). Crop Yield Assessment Using Field-Based Data and Crop Models at the Village Level: A Case Study on a Homogeneous Rice Area in Telangana, India. *AgriEngineering*, 5(4), 1909–1924. <https://doi.org/10.3390/agriengineering5040117>
33. Maniam, S., Tee, Y.-K., Memar, E., Wong, H. Y., & Zaman, M. (2026). Smart irrigation management: IoT-based RNN-LSTM model for soil moisture prediction in precision agriculture. *Smart Agricultural Technology*. Advance online publication.
34. Munna, M. A., et al. (2022). Site-specific seeding for maize production using management zone maps delineated with multi-sensors data fusion scheme. *Soil and Tillage Research*, 220, 105377.
35. Munna, M.A., Haesaert, G., Van Meirvenne, M. et al. Multi-sensors data fusion approach for site-specific seeding of consumption and seed potato production. *Precision Agric* 22, 1890–1917 (2021). <https://doi.org/10.1007/s11119-021-09817-8>
36. Murthy, C. S., Poddar, M. K., Choudhary, K. K., Pandey, V., Srikanth, P., Ramasubramanian, S., & Senthil Kumar, G. (2022). Paddy crop insurance using satellite-based composite index of crop performance. *Geomatics, Natural Hazards and Risk*, 13(1), 310–336.
37. NADU, PREDICTION OF TAMIL. "BIG DATA ANALYTICS IN CROP YIELD PREDICTION OF TAMIL NADU (RICE)." (2024).

38. Nancy Jane, Y., Praveen, P., Kavya Swetha, J., & Shanthosh Kumar, E. (2023). Intelligent Shiny Dashboard for District-wise Crop Suggestion in Tamil Nadu. 2023 12th International Conference on Advanced Computing (ICoAC), 1–4. <https://doi.org/10.1109/ICoAC59537.2023.10249727>
39. Ngo, V. M., Le-Khac, N. A., & Kechadi, M. T. (2019, June). Designing and implementing data warehouse for agricultural big data. In International conference on big data (pp. 1-17). Cham: Springer International Publishing
40. Pazhanivelan, S., K. P. Raguath, N. S. Sudarmanian, R. Kumaraperumal, T. Setiyono, and E. D. Quicho. 2019. "Integrating Time-Series SAR Data and ORYZA Crop Growth Model in Rice Area Mapping and Yield Estimation for Crop Insurances." International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-3/W6: 239–243. <https://doi.org/10.5194/isprs-archives-XLII-3-W6-239-2019>.
41. Podnar, I., et al. (2024). Efficient data management in agricultural IoT: Compression, security, and MQTT protocol analysis. *Sensors*, 24(11), 3517. <https://doi.org/10.3390/s24113517>
42. Rahim, A. I. A., & Chokkalingam, L. (2025). Future multi-dimensional drought analysis using machine learning and geospatial approaches: A case study of Tiruchirappalli District, Tamil Nadu, India. *Environmental Science and Pollution Research International*. Advance online publication.
43. Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., & Hughes, D. P. (2017). Deep learning for image-based cassava disease detection. *Frontiers in Plant Science*, 8, Article 1852. <https://doi.org/10.3389/fpls.2017.01852>
44. Saiz-Rubio, V., & Rovira-Más, F. (2020). From smart farming towards agriculture 5.0: A review on crop data management. *Agronomy*, 10(2), 207.
45. Seiche, A.T., Wittstruck, L., & Jarmer, T. (2024). Weed Detection from Unmanned Aerial Vehicle Imagery Using Deep Learning-A Comparison between High-End and Low-Cost Multispectral Sensors. *Sensors*, 24(5), 1544. <https://doi.org/10.3390/s24051544>
46. Shanmugapriya, P., K. R. Latha, S. Pazhanivelan, R. Kumaraperumal, G. Karthikeyan, and N. S. Sudarmanian. "Cotton yield prediction using drone derived LAI and chlorophyll content." *Journal of Agrometeorology* 24, no. 4 (2022): 348-352
47. Sharma, S. S., Mukherjee, J., & Dell'Acqua, F. (2025). Leveraging Sentinel-2 data and machine learning for drought detection in India: The process of ground truth construction and a case study. *Remote Sensing*, 17(18), 3159. <https://doi.org/10.3390/rs17183159>
48. Smaya, H. (2022). The influence of big data analytics in the industry. *Open Access Library Journal*, 9(2), 1-12
49. Sun, Xiuyu, Xiaohui Zhong, Xiaoze Xu, Yuanqing Huang, Hao Li, J. David Neelin, Deliang Chen et al. "FuXi weather: a data-to-forecast machine learning system for global weather." *arXiv preprint arXiv:2408.05472* (2024).
50. Thinakaran, J., Paul, S., Beulah Christalin Latha, C., & Jacob, G. (2023). Blockchain in big data for agriculture supply chain. In S. Namasudra & K. Akkaya (Eds.), *Blockchain and its applications in Industry 4.0* (Studies in Big Data, Vol. 119). Springer. https://doi.org/10.1007/978-981-19-8730-4_9
51. Vrieling, A., Meroni, M., Mude, A.G., Chantarat, S., Ummenhofer, C.C., & de Bie, C.A.J.M. (2016). Early assessment of seasonal forage availability for mitigating the impact of drought on East African pastoralists. *Remote Sensing of Environment*, 174, 44–55. <https://doi.org/10.1016/j.rse.2015.11.030>
52. Wolfert, S., Ge, L., Verdouw, C. & Bogaardt, M.J. (2017). Big Data in Smart Farming – A review. *Agricultural Systems*, 153: 69–80. DOI: 10.1016/j.agsy.2017.01.023.
53. Zhang et al. (2024). Predicting wheat yield from 2001 to 2020 in Hebei Province at county and pixel levels based on synthesized time series images of Landsat and MODIS. *Scientific Reports*, 14.
54. Zhang, Y., Wang, G., Wang, Z., & Yang, L. (2020). Applications of artificial intelligence in agriculture: Status and prospects. *Computers and Electronics in Agriculture*, 170, 105236. <https://doi.org/10.1016/j.compag.2020.105236>
55. Zhou, X., Zheng, H. B., Xu, X. Q., He, J. Y., Ge, X. K., Yao, X., Cheng, T., Zhu, Y., Cao, W. X., & Tian, Y. C. (2017). Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130, 246–255.