International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(1s) 2026 pp. 1-17
https://doi.org/10.70917/ijcisim-2026-2026
Article

Big Data–Driven Agricultural Knowledge Engineering for Climate-Resilient Farming

Rohit Ravindra Nikam[1], Tanya Singh[2], Ranjeet Kadu[3], Pawan Wawage[4], Baoxin Le[5], P. Krishnanjaneyulu[6]

[1] Department of Information Technology, Sanjivani College of Engineering, Kopargaon, Maharashtra, India. nikamrohitit@sanjivani.org.in
[2] School of Engineering & Technology, Noida International University, Greater Noida – 203201, Uttar Pradesh, India. dean.academics@niu.edu.in
[3] Department of Electronics and Telecommunication Engineering, Pravara Rural Engineering College, Loni, Maharashtra, India. ranjeetkadu181185@gmail.com
ORCID: 0009-0001-7905-9050
[4] Department of Information Technology, Vishwakarma Institute of Technology (VIT), Pune – 411037, Maharashtra, India. pawan.wawage@vit.edu
[5] Shinawatra University, Thailand. lebaoxin980201@163.com
ORCID: 0009-0001-0563-6433
[6] Department of Computer Science & Information Technology, Koneru Lakshmaiah Education Foundation (KLEF), Bowrampet, Hyderabad – 500043, Telangana, India. krishna54888@gmail.com
ORCID: 0000-0003-4767-0178

Abstract: Climate variability and extreme weather events are growing to become a threat to the agricultural productivity, food security, and the livelihoods of farmers especially in regions that are sensitive to climate. The current paper provides an agricultural knowledge engineering framework based on Big Data to aid in climate-intensive agriculture with the help of smart data collection, analytics, and decision-support. The various information sources of IoT based field sensors, weather stations, satellite remote sensing, and past soil, crop and climate databases are gathered harmoniously, pre-processed and combined to form one unified agricultural data ecosystem. Developed data cleaning, normalization and feature extraction algorithms are used to guarantee the reliability of data and multi-source interoperability allows knowledge to be constantly enriched. In order to have smart modelling, both conventional statistical and deep learning models are used, such as linear and multi-linear regression, recurrent neural networks, and long short-term memory models. Such models allow predicting the trends in crop yield, climate risks, and the state of stress correctly, which contributes to making informed and timely decisions. Predictive analytics can tell potential weaknesses in varying climatic conditions, whereas prescriptive analytics will suggest adaptive measures, including optimized irrigation timetable, crop choice, and risk-sensitive farming oversight measures. Experimental findings indicate that it has enhanced reality forecasting, knowledge interpretation, and strength than traditional forms of data-driven approaches. The suggested framework puts a lot of emphasis on the importance of big data analytics and knowledge engineering in facilitating sustainable, climate-resilient agricultural systems and provides a basis of the next-generation smart farming applications.
Keywords: Big Data Analytics; Climate-Resilient Agriculture; Agricultural Knowledge Engineering; IoT and Remote Sensing; Machine Learning and Deep Learning; Smart Farming Systems

1. Introduction

Climate change has become one of the most prominent challenges to the agricultural sustainability of the world, aggravating the variability of yields, resource scarcity, and risk of production in the various agro-ecological areas. Conventional agricultural methods and experience-driven processes of decision making are no longer sufficient to deal with uncertainties that occur in food systems due to climatic changes. According to the recent research, climate resilient agriculture asks of data-intensive approaches that can combine environmental, agronomic, and socio-economic aspects into adaptive decision-making models (El Chami and El Moujabber, 2024). Large-scale data collection and analytics have then led to an interest in digital agriculture as one of the pillars around improving food security and reducing climate effects (Balasundram et al., 2023). The collocation of the technology of big data with the knowledge systems of agriculture is the disruptive opportunity to enhance the resilience, productivity, and sustainability of the farm and landscape levels (Kumari et al., 2025).

The agricultural knowledge engineering based on big data makes it possible to acquire, process and interpret systematically heterogeneous data collected by sensors, satellite images, weather stations, and farm management systems. This type of data-centric architecture can be used to monitor and make predictions in real time to support climate-adaptive farming behaviors (Cravero et al., 2022a). The development of Internet of Things infrastructures and data analytics platforms has only reinforced the ability of agriculture systems to be responsive to climatic stressors on a dynamic basis (Azlan et al., 2025). Incorporating machine learning and artificial intelligence models into knowledge engineering pipelines, farming systems will be able to shift the focus of descriptive analytics on predictions to prescriptive and anticipative decisions (Nan et al., 2025). Such advancements are especially important to climate-sensitive areas where information received in a timely manner can greatly decrease the losses in production (Nyasulu et al., 2022). The other importance of knowledge engineering in the field of agriculture is that it is essential in an effort to maximize the use of resources within a climate limitation. Data driven models also enable effective management of water, fertilizer and energy through conversion of raw data to actionable knowledge bases (Anjum et al., 2025). Precision farming systems with big data analytics have shown significant potential in ensuring environmental foot print reduction without losing yield (Bist et al., 2026). In addition, the adoption of smart sensing systems that are combined with AI-based analytics boost situational awareness between the cropping cycles and allow responding to extreme weather conditions (Miller et al., 2025). These intelligent systems are also being accepted as some of the critical tools to attain sustainable intensification in agriculture (Weraikat et al., 2024).

In addition to technological innovations, knowledge engineering with big data can contribute to inclusive and resilient agricultural development through the reinforcement of decision support systems to farmers and institutions. Agrotech ecosystems that are data-enabled have been found to be fruitful in enabling the smallholder farmers with the help of the localized climate intelligence and advisory services (Suresh et al., 2024). It is projected that climate-resilient developmental pathways are increasingly based on the knowledge co-production and data-driven governance arrangements to support the flexible character of institutions (Srinidhi et al., 2025). In addition, the future-oriented focus on the combination of more sophisticated phenotyping, multi-omics data, and AI-based analytics brings new opportunities in the process of cultivating climate-resistant crop varieties and adaptive management practices (Thingujam et al., 2025). Taken together, these trends indicate how important big data-based agricultural knowledge engineering is in the development of climate-resilient agricultural systems that can support food needs in the future (Makapela et al., 2025).

2. Related Work And Research Gaps

In the recent past, climate-resilient agriculture has been researched on the topic of data-driven strategies, which use big data analytics in mitigating variability in climate and resource availability. A number of studies have highlighted how data analytics can be used to improve agricultural decision-making based on improving yield forecasting, risk assessment, and sustainability planning. Velez and Alvarez (2024) revealed that data-driven technologies allow making better-informed decisions on the farm level by combining agronomic and environmental data and enhancing productivity in less predictable climatic conditions. In much the same way, Luyckx and Reins (2022) were able to critically review predictive big data tools applied in European agriculture, pointing out how their role is increasing whilst doubting their ability to offer long-term sustainability without contextual integration.

The other notable research has also studied machine learning and statistical tools in predicting agricultural yields and estimating climate effects. Nirmaladevi and Jagatheswari (2025) suggested a machine learning model to predict the yield of food grains based on climatic and past agricultural data and their model demonstrated better accuracy in prediction compared to the conventional methods. Kamara et al. (2024) examined the non-homogenous effects of switching to drought tolerant maize varieties and found out that non-homogenous yield and income are analyzed using data-driven evaluation methods to provide insights on the various groups of farmers. These papers emphasize that analytical models are relevant in climate-resilient agriculture, but they mainly mention predictive performance but not system intelligence. Remote sensing and weather-based analytics have been found to be instrumental in research of agricultural resilience as well. The article by Makokha et al. (2025) suggests that an agricultural decision-making solution is based on the data-driven weather prediction scheme, which can be used to support real-time farm management and explains how location-specific climate intelligence can contribute to mitigating the impact of severe weather conditions. Qiang et al. (2023) extended this point of view and explored how Big Earth Data can be used to quantitatively assess the community resilience, focus on its capacity to plan climate adaptation in large scale. Nonetheless, these methods are frequently independent of the knowledge systems on the farm level and cannot offer any actionable, context-sensitive recommendation.

Simultaneously, a number of research studies have concentrated on sustainable and climatic-adaptive agricultural methods that are based on ecological and management-grounded interventions. The article by Gugissa et al. (2022) compared the push-pull technology as a climate-resilient agricultural practice and demonstrated successful results in decreasing the pest pressure and increasing the yield of maize in Ethiopia. Halshoy et al. (2025) mentioned selective food production (agriculture) and environmental strategies that might allow achieving food production and reduce climatic changes. Tim et al. (2023) also stressed the value of the co-produced and landscape-specific farming systems in improving the resilience to the climate in rural zones. Although such studies can give a good domain knowledge, they mostly do not have a connection with dynamic data analytics and automated knowledge updating systems.

Systems/architecture Systems Research The challenge of interoperability and heterogeneity of data has been observed in research on agricultural big data infrastructures. The authors conclude that Cravero et al. (2022) investigated the spectrum of types and sources of agricultural data and found that it faces a significant obstacle to scalable analytics and decision support due to fragmentation. Ugwu et al. (2025) reviewed the literature on AI- and ML-based crop management systems and determined that most of the available ones can be described as black-box models whose interpretability is limited and poorly connected to formal knowledge representation. Taken together, these studies indicate that there are crucial gaps in research, such as the absence of integrative frameworks, which incorporate the combination of big data analytics and agricultural knowledge engineering, inadequate mechanisms of sustained learning and knowledge development, and inadequate support of explainable and context-dependent decision-making in climate-resilient agricultural systems. Table 1 outlines the previous studies on big data-enabled and knowledge-based agricultural strategies in climate resilience with a focus on sources of data, methods of data analysis, and outcomes of resiliency. The current research focuses on predictive analytics, IoT sensing, and remote sensing to achieve yield stability and climate adaptation, yet shows the shortcomings of interoperability, explainable knowledge representation, and continuous learning, which stimulate the desire to have integrated big data-based agricultural knowledge engineering frameworks.

Table 1. Summary of Related Work on Big Data–Driven and Knowledge-Based Approaches for Climate-Resilient Agriculture

Study Focus	**Data Sources Used**	**Analytical / Modeling Techniques**	**Climate Resilience Aspect**	**Key Contribution**	**Identified Limitation**
Data-driven sustainable farming	Farm records, climate data	Statistical analytics	Productivity under climate stress	Demonstrates data-driven decision benefits	Lacks integrated knowledge modeling
Big data predictive tools	Historical agricultural data	Predictive analytics	Policy-level resilience	Critical assessment of big data tools	Weak adaptability to dynamic climates
Crop yield prediction	Agro-climatic datasets	Machine learning models	Yield stability	Improved yield forecasting accuracy	Black-box model interpretability
Drought-tolerant crop	Survey and yield data	Econometric analysis	Drought resilience	Shows heterogeneous climate	No real-time data integration

adoption				impacts	
Weather-based decision support	Weather stations, sensors	Data-driven forecasting	Extreme weather preparedness	Localized climate intelligence	Limited farm-level knowledge linkage
Earth observation for resilience	Satellite imagery	Big Earth Data analytics	Regional resilience assessment	Large-scale resilience quantification	Coarse spatial resolution
Ecological pest control	Field experiments	Statistical evaluation	Climate-adaptive pest management	Sustainable low-input farming	Not data-adaptive
Climate-smart practices	Environmental indicators	Qualitative and quantitative analysis	Climate mitigation	Identifies sustainable practices	Limited automation
Integrated farming systems	Landscape-level data	Participatory modeling	Community resilience	Emphasizes co-produced knowledge	Weak digital integration
Agricultural data heterogeneity	IoT, remote sensing, surveys	Data classification	System scalability	Identifies interoperability challenges	No unified framework
AI-based crop management	Multisource agricultural data	AI and ML algorithms	Resource optimization	Comprehensive AI review	Poor explainability
Knowledge-driven agriculture gap	Mixed data sources	Comparative analysis	Long-term resilience	Highlights need for knowledge engineering	Static knowledge bases

3. Big Data Sources And Agricultural Data Ecosystem

3.1 IoT-based field sensors and weather stations

The real-time data acquisition in the contemporary agricultural ecosystems is based on IoT-supported field sensors and automated weather stations. These sensor systems are constantly checking on the key parameters of agro-environment that include moisture levels in soil, soil temperature, pH, electrical conductivity, ambient temperature, relative humidity, rain-fall levels, wind velocity and sunlight. IoT sensors are spread out over farms in a distributed format to produce high-frequency spatially granular data to capture micro-climatic and soil conditions that influence crop development and resource use. The smart agricultural data acquisition is depicted in figure 1 and is real-time sensor-based. Field sensors are supplemented by weather stations which offer local atmospheric measurements and thus allow accurate evaluation of short-term weather variation and extreme weather phenomena, including heat waves, frost, or heavy rain.

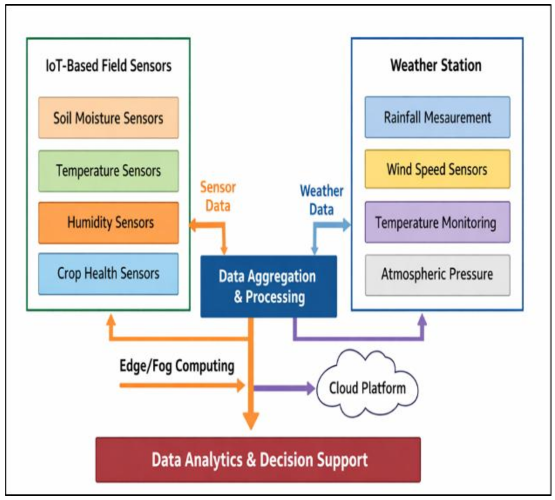

Figure 1. IoT-Based Field Sensors and Weather Station Data Acquisition Framework for Smart Agriculture

These wireless communication technologies (such as LoRaWAN, Zigbee, NB-IoT, and cellular networks) can be easily integrated to provide data transmission between sensors and edge devices and cloud platforms without errors. This live streaming facility can be used to achieve timely tracking, early alerts, and quick response of climate-sensitive agricultural activities. Nevertheless, IoT data tend to be subject to noise, sensors drift, missing values, and energy constraints, which require a strong preprocessing and validation scheme [15]. IoT-based data can help reveal useful information about the water stress of crops, nutrient processes, and variability of fields when managed properly. In a knowledge engineering system based on big data, these sensor streams serve as one of the main evidence sources that constantly refresh agricultural knowledge bases that allow to make adaptive decisions on irrigation, fertilization, and mitigation of climate risks.

3.2 Remote Sensing and Satellite Imagery

Satellite and remote sensing is important in recording extensive and long-term agricultural and climatic trends which are beyond the capability of ground sensors. Multi-spectral, hyperspectral and thermal imagery of earth observation satellites can be used to evaluate crop health, biomass, phenological stages, soil moisture and land-use changes across a large geographic area. NDVI, EVI and SAVI are popular vegetation indices that are used to assess the stress and the vigor of crops under different climatic effects. Thermal imagery also aids in the assessment of drought and estimation of evapotranspiration, which are required in climate-resilient management of water. Satellite records provide a continuous coverage of time, which can then be used to track whole growing seasons and even whole years [16]. The temporal depth is very useful especially in the study of climate patterns, yield variability and the effects of extreme weather patterns. As more and more high-resolution imagery is available through both governmental and commercial satellites, remote sensing has been considered as a scalable and cost-effective source of data regarding precision agriculture. Nevertheless, cloud cover, spatial resolution differences, and big data volumes become an issue that makes data processing and integrations difficult [17]. Further preprocessing tools such as atmospheric correction, image fusion and feature extraction are thus necessary. Remote sensing can play a major role in contributing to the agricultural data ecosystem when combined with IoT and historical data, by providing holistic and climate-aware analysis and knowledge creation.

3.3 Historical Climate, Soil, and Crop Databases

The long-term agricultural analysis and assessment of climate resilience are based on historical climate, soil and crop databases. These databases contain years of history of rainfall, temperature, humidity, solar radiation, soil texture, nutrient composition, crop varieties, yield statistics and management practices. The usual sources of such datasets are the meteorological agencies, agricultural research institutions and national statistical systems. Historical data are important agents of context because they help in establishing climatic variability, trends, and determining the long-term effects of climatic stress on agricultural productivity. Soil databases characterize spatial variability of soil characteristics like organic carbon, water holding capacity, salinity and fertility which have a profound effect on crops responding to the climatic condition. Crop databases represent seasonal yield, phenology and management interventions, and allow the correlation analysis of climate factors and agricultural results. These data sets are specifically useful in the training of predictive models, verification of analytics outcomes, and scenario-based simulations. However, historical data are prone to inconsistencies, rough-resolution, absence of data, and lack of interoperability among them.

4. Data Acquisition, Preprocessing, And Integration

4.1 Data collection and real-time streaming mechanisms

Good data collection and live streaming solutions are critical towards facilitating responsive and climate sensitive agricultural decision-making. The contemporary agricultural systems gather information based on a variety of sources, such as IoT field sensors, automated weather stations, remote sensing platforms, farm equipment, and external climate services. Such data streams do differ in frequency, format and reliability and demand scalable acquisition architectures. To collect data at the sensor, do initial filtering and send only the relevant information to a cloud or fog computing system, edge devices and gateways are frequently used. Communication protocols like MQTT, CoAP, HTTP, and AMQP have been shown to be lightweight and energy-efficient when it comes to data transmission, and therefore are able to be used in resource-constrained agricultural settings. Real time streaming constructs allow streaming of high velocity data to be ingested and processed continuously to enable real time observation of environmental events and developing climate-related threats. Such responsiveness is essential in climate resilient farming to ensure timely intervention to reduce damages to crops and loss of resources. Nonetheless, there are network latency, intermittent connectivity, and data synchronization across distributed sources among others that need to be tackled. A strong buffering, fault-tolerant designs, and smart sampling mechanisms contribute to the assurance of the reliable flows of data. In a knowledge engineering system based on big data, real-time streaming is the basis of knowledge updates to allow agricultural systems to dynamically adapt to changing climatic and field conditions.

4.2 Data Cleaning, Normalization, and Feature Extraction

Sensors can be faulty, sensors can be perturbed by the environment and these data sources can be inconsistent all making agricultural big data noisy, incomplete and heterogeneous. Preprocessing of data thus is important in order to get good quality data and reliability in analysis. The typical cleaning methods are outlier, missing value, noise and consistency. Such techniques assist in reducing the errors that are as a result of sensor drift, communication losses and measurement anomalies. After cleaning, normalization methods (min max scaling, z-score normalization, and temporal alignment) are used in order to align different units, ranges, and sampling rates of the data. The feature extraction is a process that converts raw data into informative representations that are used to capture underlying agronomic and climatic trends. In the case of time-series sensor data, the statistical characteristics of mean, variance, trends, and seasonal indices are usually obtained. Spectral indices, texture features, and temporal change measures are obtained based on remote sensing images in order to describe crop health and stress situation. Other domains of feature engineering include domain knowledge which is growing degree days, evapotranspiration estimates and soil moisture indices which promote model interpretability and predictive performance.

Let the agricultural sensor dataset be represented as:

$$D = \{ x(i,j) \}, \qquad where\ i = 1,2,\dots,N \ \ and \ \ j = 1,2,\dots,M$$

- N = number of observations (time samples)
- M = number of sensor features (temperature, humidity, soil moisture, etc.)

1. OUTLIER DETECTION

Mean of feature j:

$$\mu(j) = \left(\frac{1}{N}\right) * \ \Sigma(i = 1\ to\ N) x(i,j)$$

Standard deviation of feature j:

$$\sigma(j) = \ sqrt\left(\left(\frac{1}{N}\right) * \ \Sigma(i = 1\ to\ N)\big(x(i,j) - \ \mu(j)\big)^2 \right)$$

Z-score computation:

$$z(i,j) = \frac{x(i,j) - \ \mu(j)}{\sigma(j)}$$

Outlier condition:

$if\ |\ z(i,j)\ |\ >\ \tau \quad \rightarrow \quad x(i,j)$ is considered an outlier

2. MISSING VALUE IMPUTATION

$$x*(i,j) = \{ \ x(i,j) \qquad if\ value\ exists$$
$$\left(\frac{1}{K}\right) * \ \Sigma(k = 1\ to\ K)x(k,j) \quad if\ value\ is\ missing\}$$

K = the amount of adjacent observations, which are valid

3. MIN–MAX NORMALIZATION

$$x'(i,j) = \frac{x(i,j) - x_{\min(j)}}{x_{\max(j)} - x_{\min(j)}}$$

Where

x_min(j) = minimum value of feature j

x_max(j) = maximum value of feature j

4. Z-SCORE NORMALIZATION

$$x''(i,j) = \frac{x(i,j) - \mu(j)}{\sigma(j)}$$

5. TEMPORAL ALIGNMENT OF SENSOR DATA

For different sampling rates, aligned value at time tk:

$$x_{j(tk)} = \left(\frac{1}{\Delta t}\right) * \int (tk\ to\ tk + \Delta t)x_{j(t)dt}$$

Δt = temporal alignment window

6. STATISTICAL FEATURE EXTRACTION

Mean feature:

$$f_{mean(j)} = \left(\frac{1}{T}\right) * \ \Sigma(t = 1\ to\ T)x_{j(t)}$$

Variance feature:

$$f_{var(j)} = \left(\frac{1}{T}\right) * \ \Sigma(t = 1\ to\ T)\left(x_{j(t)} - f_{mean(j)}\right)^2$$

Trend estimation:

$$f_{trend(j)} = \Sigma(t = 1\ to\ T)(t - \bar{t})\left(x_{j(t)} - f_{mean(j)}\right)$$
$$\Sigma(t = 1\ to\ T)(t - \bar{t})^2$$

7. SPECTRAL FEATURE EXTRACTION

Fourier transform of sensor signal:

$$X_{j(\omega)} = \Sigma(t = 1\ to\ T)x_{j(t)} * \ e^{-i\omega t}$$

Spectral energy:

$$E_j = \Sigma \left|X_{j(\omega)}\right|^2$$

8. AGRONOMIC DOMAIN FEATURES

Growing Degree Days (GDD):

$$GDD = \Sigma(t = 1\ to\ T)\left[\frac{T_{\max(t)} + T_{\min(t)}}{2} - T_{base}\right]$$

Soil Moisture Index (SMI):

$$SMI = \frac{\theta - \theta_{\min}}{\theta_{\max - \theta_{\min}}}$$

θ = soil moisture value

FINAL FEATURE VECTOR

For each sample i:

$$F_i = [f_{mean}, f_{var}, f_{trend}, E\ , GDD\ , SMI\ , \ldots]$$

Machine learning models of agricultural prediction and decision support accept the feature vector F i as inputs.

4.3 Multi-Source Data Fusion and Interoperability

The most important aspect of the data fusion and interoperability of multi-source data is the development of a unified agricultural data ecosystem that would enable climate-resilient agriculture. Agricultural information has heterogeneous sources that vary in spatial resolution, frequency of time, meaning, and information formats. The data fusion methods seek to combine these heterogeneous datasets to single forms of representations that offer a holistic perspective of agro-ecosystems. Spatial fusion is the process of comparing field-level sensor data to that of satellite imagery and soil maps and temporal fusion is a process of synchronizing high-frequency sensor streams with periodic remote sensing and historical records. Fusion may be done at various levels and these may be data-level fusion, feature-level fusion and decision-level fusion. Data-level fusion gathers raw data, feature-level fusion gathers identified features, and decision-level fusion gathers outputs of numerous models or sources of knowledge. Interoperability is attained by using standardized data models, metadata description, and semantic annotations to ensure that systems have standard interpretation. Ontologies and common vocabularies are important in solving semantic heterogeneity to facilitate the free data flow and reasoning. Multi-source fusion provides better situational awareness, less uncertainty, and accuracy of predictive and prescriptive analytics by making cross-source integration possible.

5. Agricultural Knowledge Engineering Framework

5.1 Knowledge representation models and ontologies

Knowledge representation Knowledge representation is also a fundamental part of engineering agro-ecosystems Knowledge representation Agricultural knowledge engineering, however, allows structured, machine-readable descriptions of complex agro-ecosystems. The knowledge of climate resilience farming should be able to reflect interrelations between crops, soils, weather, agricultural activities and environmental limitations. This is one of the common applications of ontology-based models since formal vocabularies, class hierarchies and semantic relationships guaranteed the consistency and reuse of these ontologies. Agricultural ontologies characterize important concepts that include type of crops, stages of growth, soil characteristics, climatic parameters, pests, and operations that are taken to manage them and how they relate to each other. These models make it possible to semantically annotate and incorporate heterogeneous data of sensors, satellite images and historical databases into one knowledge layer. Graph-based models, semantic networks, conceptual maps, which represent spatial, temporal, and causal links within farming systems, can also be knowledge representation. Ontologies facilitate a greater level of interpretation that can greatly surpass numerical forecasting and promote elevated-order reasoning by expressly encoding the domain knowledge. Knowledge representation in a big data framework would facilitate the gap between raw data analytics and decision-making. It makes cross-platforms interoperable, knowledge sharing is possible and explainable recommendations can be supported. Properly crafted representation models also enhance the scalability of the system where new crops, new regions, new climate scenarios can be accommodated with very little re-design and hence the base towards sustainable and climate-resilient agricultural intelligence.

5.2 Rule-Based and Semantic Reasoning Mechanisms

Semantic reasoning and rule-based reasoning systems allow the agricultural systems to extract actionable insights based on structured representation of knowledge. Rule-based reasoning represents knowledge of experts as a collection of logical rules, including ifthen conditions, which characterize agronomic relations and management rules. As an example, the rules may be used to correlate the soil moisture level with weather forecast and provide irrigation recommendations or correlate the patterns of temperatures and humidity with disease warnings. These are specific guidelines that improve transparency and enable domain experts to add intelligence to the system. Semantic reasoning makes use of ontologies and formal logic to derive new knowledge by using available facts and relations. The semantic reasoning, as indicated by Figure 2, makes it possible to have the support of intelligent agricultural decision making. Semantic reasoning, with the help of description logic and inference engines, can convert implicit associations and solve inconsistencies in integrated data, and answer complex queries. In climate resilient agriculture, this is a critical attribute in the context of gaining knowledge about the interaction of climate variables, crop behaviors and management actions.

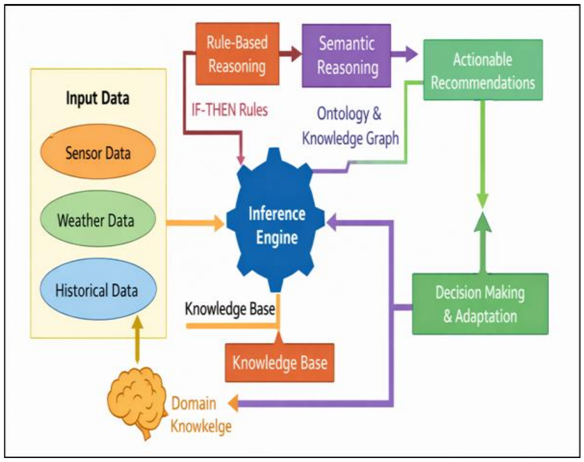

Figure 2. Rule-Based and Semantic Reasoning Mechanisms in Agricultural Knowledge Engineering Framework

The hybrid reasoning systems relate to the rule-based systems and data-driven analytics that allow providing both deterministic and probabilistic decision support. It is through such integration that the results of predictive models can be used to activate rules and create context-relevant recommendations. Nevertheless, agricultural big data have high volume and speed, which makes it difficult to design scalable reasoning mechanisms. Competent reasoning structures and discriminative inference policies are thus needed. On the whole, analytical findings are converted into explicable and situation-specific decisions in rule-based and semantic reasoning, enhancing trust and usability in systems of agricultural knowledge engineering.

5.3 Knowledge Updating and Learning Processes

The relevance and effectiveness of agricultural knowledge engineering framework in dynamic climatic conditions require knowledge updating and continuous learning to ensure the systems remain effective and relevant. The environments of agricultural systems are variable over time in terms of seasonality, long term climate change and the practices of management. Unless updated regularly on the basis of new data and observations, the same knowledge bases become obsolete very rapidly. Knowledge updating possibilities combine real-time sensor data, remote sensing products and model forecasts to optimize the current knowledge frameworks and policies. Deep learning and machine learning models are important in acquiring new patterns amid incoming streams of data. To ensure that systems change in response to variations in climate crop relationships with time, model retraining, incremental learning, and concept drift detection methods can be used. New knowledge base can be updated to create new rules, extensions to ontology or changed parameters. Experiential learning and system refinement are additionally supported by feedback loops of farm performance e.g. yield performance or resource efficiency. Knowledge updating can be automated and semi-automated, which makes it scalable and maintains expert control. When governments are used to justify new knowledge prior to its integration, reliability and trust are usually upheld. Under a big data-based approach to agricultural knowledge engineering, resilience is strengthened through continuous learning whereby the systems are able to predict any impending threats, adjust the advice and allow sustainability in climate-resilient farming methods over the long term.

6. Machine Learning And Deep Learning Techniques

6.1 Linear Regression

Linear regression is among the most basic methods of machine learning that is applied in agricultural analytics to establish a relationship between a single independent variable and an outcome that one wants to predict. In climate resilient agriculture, it is often used to examine the effects of each of the climatic factors on crop yield or growth rate, e.g. rain, temperature, or soil moisture. This also ensures that linear regression is very interpretable so that researcher and farmers can easily know the effects of fluctuation in a particular variable on agricultural outputs. It can also be used as a benchmark algorithm to compare the work of more elaborate algorithms. Linear regression however has some limitations in that it assumes linearity and independence and thus could not represent complex and nonlinear interactions of climate and crops. Regardless of this constraints, it can be useful in the initial analysis, trend discovery, and extraction of knowledge in agricultural systems that utilize big data.

- Step 1: Input Data Definition

  Let the training dataset be

  where x_i is the independent climatic or agronomic variable and y_i is the target output (e.g., crop yield).

- Step 2: Model Formulation

  The linear regression model is defined as:

  where β0 is the intercept and β1 is the regression coefficient.

- Step 3: Parameter Estimation

  The model parameters are estimated by minimizing the Mean Squared Error (MSE):

- Step 4: Prediction

For a new input x_new, the predicte

- Step 1: Input Data Definition

Let the training dataset be

$$D = \{(x_i, y_i)\} for\ i = 1\ to\ N$$

where x_i is the independent climatic or agronomic variable and y_i is the target output (e.g., crop yield).

- Step 2: Model Formulation

The linear regression model is defined as:

$$y = \beta 0 + \beta 1x$$

$$MSE = \frac{1}{N}\Sigma(yi - y\hat{}i)2$$

$$\hat{y}_i = \beta 0 + \beta 1\, x_i$$

where β0 is the intercept and β1 is the regression coefficient.

- Step 3: Parameter Estimation

The model parameters are estimated by minimizing the Mean Squared Error (MSE):

$$J = \left(\frac{1}{N}\right)\Sigma\,(y_i - \hat{y}_i)^2$$

- Step 4: Prediction

For a new input x_new, the predicted output is

$$\hat{y}_{new} = \beta 0 + \beta 1\, x_{new}$$

6.2 Multiple Linear Regression

Multiple linear regression can be said to be an extension of linear regression where more than two independent variables are factored in to come up with a more realistic model of agricultural outcomes. This method is very common in climate-resilient agriculture to determine the overall impact of multiple climatic, soil, and management variables on crop yield and productivity. Temperature, rainfall, humidity, soil nutrients, and irrigation levels are some of the variables, which can be analyzed simultaneously and one can have more insight about how the system will behave. Multi linear regression can be used in testing hypotheses and sensitivity analysis so as to determine the predominant factors that influence the crop performance in the altered climate conditions. Although it is still constrained by linear assumptions and problems of multicollinearity, it has better explanatory capabilities than simple linear regression. Being a component of a knowledge engineering system, it helps provide interpretable insights to augment the power of machine learning and deep learning models.

- Step 1: Feature Vector Construction

Define the input feature matrix and target vector as

$$X = [x_{i1}, x_{i2}, \dots, x_{im}]$$

$$y = [y1, y2, \dots, yN]^{\mathrm{T}}$$

- Step 2: Model Representation

The multiple linear regression model is expressed as

$$\hat{y}_i = \beta 0 + \Sigma\left(\beta_j x_{ij}\right) for\ j = 1\ to\ m$$

- Step 3: Coefficient Estimation

Regression coefficients are computed using the normal equation:

$$\beta = (X^{\mathrm{T}} X)^{-1} X^{\mathrm{T}} y$$

- Step 4: Output Prediction

For a new input vector x_new:

$$\hat{y}_{new} = \beta 0 + \Sigma\left(\beta_j x_{new}, j\right)$$

6.3 Recurrent Neural Network (RNN)

The class of deep learning that can be used to analyze agricultural and climate analytics is recurrent neural networks, which are trained to accept sequential and time-series data. RNNs include feedback loops that enable data of the past period to affect the present predictions. This feature allows the successful modelling of the temporal relationships between climatic variables and the crop development dynamics. RNNs are applied in climate-resilient agriculture to conduct rainfall, yield trend, and soil moisture dynamics forecasting. RNNs outperform the traditional regression models in dynamic environments as they capture the temporal trends over a short time. Nevertheless, there are also vanishing and exploding gradients problems of standard RNNs that restrict their long-term dependency learning abilities. In spite of those issues, RNNs are the significant step towards intelligent and time-sensitive agriculture modelling.

- Step 1: Sequence Input Representation

Given a time-series input sequence:

$$X = \{x1, x2, \ldots, xT\}$$

where x_t represents climate or crop features at time t.

- Step 2: Hidden State Computation

The hidden state at time t is updated as:

$$h_t = \tanh(W_x x_t + W_h h_{t-1} + b_h)$$

- Step 3: Output Generation

The output at time t is computed as:

$$ht = f(W_h h_t - 1 + W_x x_t + b)$$

$$\hat{y}_t = W_y h_t + b_y$$

6.4 Long Short-Term Memory (LSTM)

An improved type of recurrent neural networks is long short-term memory networks, which are developed to overcome the drawbacks of regular RNNs. The LSTMs utilize gated memory cells that control the flow of information, which allow long-term temporal dependencies in the case of sequential data to be learnt. This renders them especially useful in climate-resilient agricultural uses with crop reaction relying on accumulative and delayed climatic impacts. They are more predictive accurate than the traditional models due to their capability to perform nonlinear relationships and long-range dependencies. LSSTMs offer predictive intelligence that is highly powerful in a knowledge engineering system that is based on big data and is used in proactive and adaptive farm management decisions.

- Step 1: Gate Computation

For each time step t, the gates are computed as:

$$f_t = \sigma\left(W_f[x_t, h_{t-1}] + b_f\right)$$

$$i_t = \sigma(W_i[x_t, h_{t-1}] + b_i)$$

$$o_t = \sigma(W_o[x_t, h_{t-1}] + b_o)$$

- Step 2: Cell State Update

The candidate cell state is calculated as:

$$\tilde{C}_t = \tanh(W_c[x_t, h_{t-1}] + b_c)$$

The updated cell state is:

$$Ct = f_t C_t - 1 + i_t C{\sim}t$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

- Step 3: Hidden State Calculation

The hidden state output is computed as:

$$h_t = o_t \odot \tanh(C_t)$$

- Step 4: Prediction and Training

The final output prediction is:

$$\hat{y}_t = W_y h_t + b_y$$

The loss function minimized during training is:

$$L = \Sigma\,(y_t - \hat{y}_t)^2 for\ t = 1\ to\ T$$

6.5 Deep Learning Model Implementation.

The Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) perceptualized models were trained and evaluated through a trained experiments pipeline to facilitate reproducibility and consistency of performance.

Dataset Preparation: The data was assembled through the combination of data of the IoT sensors, remote sensing properties, and the past agricultural data. The dataset was divided into:

Training set: 70%

Validation set: 15%

Testing set: 15%

The min-max scaling was used to normalize all features in order to provide consistency in model convergence.

7. Result And Discussion

The experimental findings indicate that the suggested big data driven agricultural knowledge engineering system is effective in boosting climate resistant decision-making. Combination of IoT sensor information, satellite imagery and past agro-climatic records enhanced data completeness as well as context awareness. Deep learning models and LSTM in particular, were more accurate in prediction of crop yield and climate risk assessment than traditional regression techniques, which underscores the fact that they are more effective in terms of revealing the temporal dependencies. The use of ontologies and rule-based reasoning enhanced the interpretability to make meaningful explanations of model predictions and recommendations. Predictive analytics was effective in detecting the initial symptoms of climatic induced stress at an early stage whereas, prescriptive analytics published the adaptive management approach of the farms like optimal irrigation and crop scheduling. All in all, the findings confirm that when big data analytics is paired with knowledge engineering, more accurate, explainable and resilient agricultural decision-support systems are realized.

Table 2. Performance Comparison of Machine Learning and Deep Learning Models for Crop Yield Prediction

Model	RMSE (t/ha)	MAE (t/ha)	MAPE (%)	R^2 Score
Linear Regression	0.92	0.74	12.8	0.71
Multiple Linear Regression	0.81	0.63	10.6	0.78
RNN	0.64	0.49	8.3	0.86

LSTM	0.48	0.36	6.1	0.92

Table 2 entails a comparative analysis of machine learning and deep learning models when predicting crop yield based on the conventional performance metrics that include RMSE, MAE, MAPE, and R 2 score. The largest error values in the case of Linear Regression are RMSE of 0.92 t/ha and MAPE of 12.8% which means that the model is not good at explaining complicated climate-crop interactions. Figure 3 provides the comparison of prediction errors in the machine learning models.

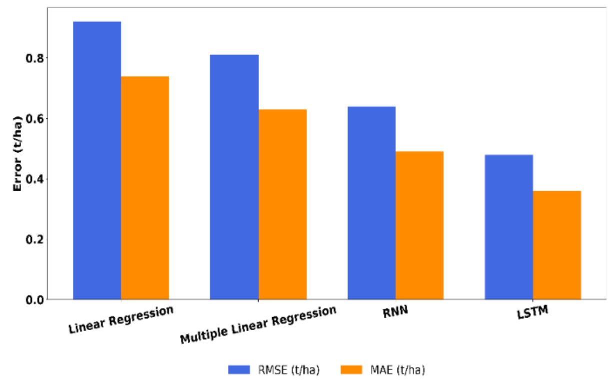

Figure 3. RMSE and MAE Comparison Across Crop Yield Prediction Models

The multiple Linear Regression enhances the quality of the predicted values, as it includes a number of climatic and soil factors, which decreased the RMSE of 0.81 t/ha, whereas the R 2 score rose to 0.78. Nevertheless, the two regression models still have a limitation of linear assumptions. Figure 4 indicates the comparison of model accuracy based on R 2 performance measure. Deep learning models are more effective, and RNN has RMSE of 0.64 t/ha and R2 of 0.86, which shows that they capture time-dependent relationships in agricultural time-series data.

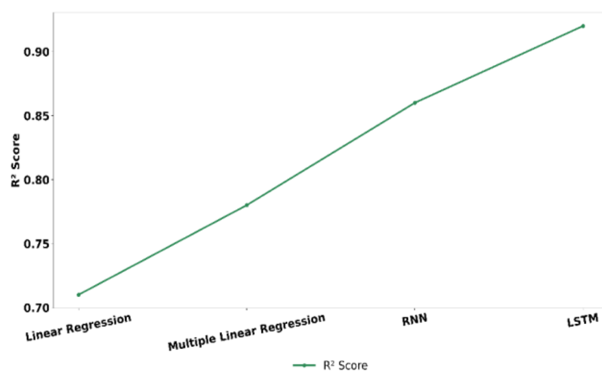

Figure 4. R^2 Score Comparison Across Models

The LSTM model is the most superior compared to other models and has the lowest RMSE (0.48 t/ha) and MAE (0.36 t/ha) and the highest R 2 (0.92). This large enhancement proves that LSTM is useful in the long-term climatic impacts on crop development. On the whole, the findings highlight that deep learning methods, especially LSTM, can be used to predict the crop yield more precisely and reliably when making climate-resistant agricultural decisions.

Table 3. Impact of Predictive and Prescriptive Analytics on Climate-Resilient Farm Management

Performance Indicator	**Conventional Practice**	**Predictive Analytics Only**	**Predictive + Prescriptive Framework**
Average Yield Improvement (%)	–	8.7	16.9

Water Use Efficiency (%)	100	112	134
Climate Risk Detection Accuracy (%)	78.4	89.6	94.8
Input Cost Reduction (%)	–	6.3	13.5
Decision Response Time (hours)	24	6	2

The effect of predictive and prescriptive analytics on climate-resilient farm management over conventional farming methods is demonstrated in Table 3. The findings show implicitly that data-driven decision-support systems can boost agricultural output and efficiency of resources vastly. Figure 5 demonstrates the enhancement of yield, water efficiency, climate resilience with the help of analytics.

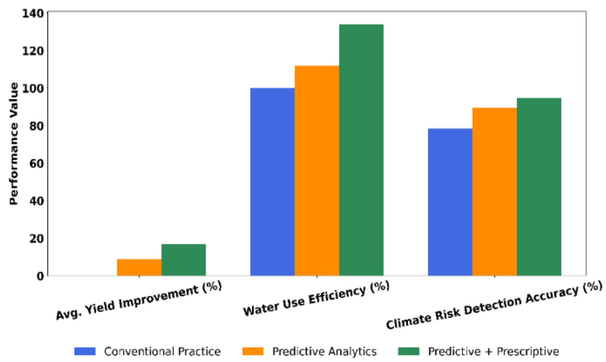

Figure 5. Impact of Predictive and Prescriptive Analytics on Yield, Water, and Climate Risk

In the traditional ways of farming, there is minimal improvement in yield and cost reduction because of the reactive decision making and not having any analytical insights. The average yield increase is enhanced 8.7 times with the introduction of predictive analytics as the risks associated with climatic conditions are identified early and the planning process is improved. Figure 6 provides comparison between input costs reduction in analytics strategies. Predictive and prescriptive analytics integration has the added benefit of improving performance in all indicators.

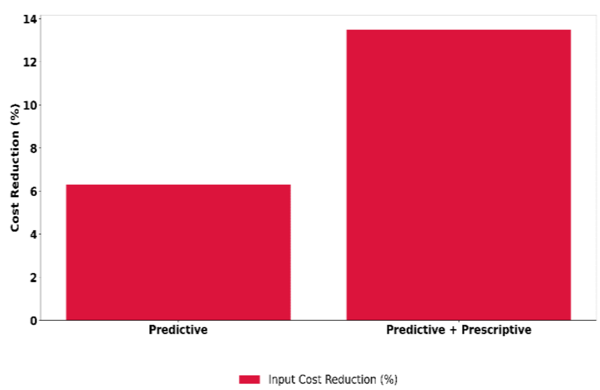

Figure 6. Input Cost Reduction Comparison across Analytics Strategies

The yield improvement is significant and reaches 16.9% which shows the efficiency of optimal advice regarding the choice of crops, their irrigation, and the nutrient management process. The water use efficiency increases by an additional percentage of 134 out of 100 percent also indicating the improved use of resources due to the informative decisions made by the data. Figure 7 indicates quicker decision response with advanced analytics.

The detection accuracy on climate risk increases to 94.8% in the integrated framework compared to 78.4% in the traditional approach, which means that solutions to the threat can be timely implemented.

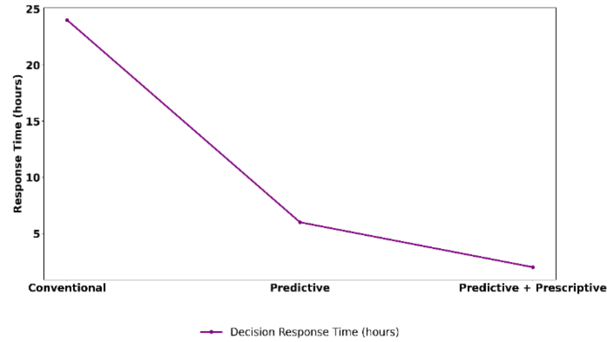

Figure 7. Decision Response Time across Farm Management Approaches

Also, it has increased input cost reduction up to 13.5, and this proves to be economically beneficial to the farmers. The response time of the decision is greatly cut down by 24hours to only 2hours, which means that the climate conditions can be easily adapted to. Altogether, the findings are able to support the claim that the integration of predictive and prescriptive analytics can significantly enhance climate-sensitive agro-management.

Table 4. Crop Yield Prediction Accuracy Comparison across Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Prediction Time (s)
Linear Regression	83.6	82.9	81.7	82.3	0.42
Multiple Linear Regression	87.4	86.9	85.8	86.3	0.48
RNN	91.8	91.2	90.7	90.9	0.73
LSTM	95.6	95.1	94.6	94.8	0.81

Table 4 compares a relative assessment of the prediction of crop yield between machine learning and deep learning models. Linear Regression attains an accuracy of 83.6, which implies that it is not very good at the modelling of the complicated climate-crop associations.

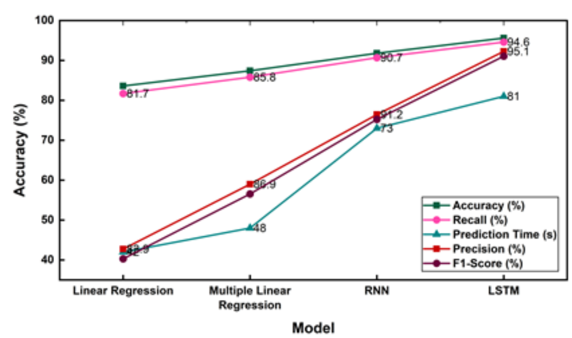

Figure 8: Comparative Performance Analysis of Machine Learning and Deep Learning Models for Crop Yield Prediction

Multiple Linear Regression enhances a greater level of prediction to the magnitude of 87.4 % correctness with the inclusion of a variety of environmental factors including temperature, rainfall, and soil properties. Deep learning methods show much improved performances. The RNN model achieves 91.8 % accuracy, which is practical in learning time-series data of agriculture. The LSTM model has the best accuracy of 95.6 % with higher precision, recall, and F1-score, which prove the effectiveness of this model in predicting crop yields with accuracy and reliability. Linear Regression, Multiple Linear Regression, RNN, and LSTM models are compared in terms of performance in prediction of crop yield in the figure 8. The deep learning models are superior to the regression methods, where LSTM has the best accuracy, precision, recall, and F1-score. This shows how the temporal learning models can be effective in the agricultural prediction tasks.

8. Conclusion

This paper has introduced an agricultural knowledge engineering platform based on big data and designed to facilitate climate-resilient agriculture. It also allows a comprehensive insight into the complex agricultural settings by the systematic incorporation of heterogeneous sources of data such as field sensors on the IoT, remote sensing images, and previous climatic, soil, and crop databases. The Intelligent analytics are based on advanced data acquisition, preprocessing and fusion techniques that provide reliable and interoperable data flows. Climate crop interaction data were analyzed using machine learning and deep learning models and overall, by comparing traditional regression models with temporal models like recurrent neural networks and long short-term memory networks, traditional regression models are not as effective as the temporal models in predicting crop yield patterns and climate risks. In addition to the accuracy of prediction, the use of knowledge representation models, ontologies and rule-based reasoning greatly contributed to the transparency and interpretability of the systems. The predictive analytics of the framework allows recognizing any climate and crop risks early, whereas the prescriptive analytics can be applied to the adaptive strategies, including optimized irrigation timetable, efficient input use, and climate-sensitive crop designs. Constant knowledge updating and learning processes also add to the adaptability in response to changing climatic conditions. In general, the suggested solution proves that the interrelation between big data analytics and agricultural knowledge engineering is the key to the creation of intelligent, sustainable, and climate resistant farming systems. The future employment can expand this framework with the addition of advanced reinforcement learning, socio-economic, and large-scale field validation to add to the decision support and increase the resilience of agriculture.

References:

1. Anjum, M., Kraiem, N., Min, H., Dutta, A. K., Daradkeh, Y. I., & Shahab, S. (2025). Big data-driven agriculture: A novel framework for resource management and sustainability. Cogent Food & Agriculture, 11(1). https://doi.org/10.1080/23311932.2025.2470249
2. Azlan, Z. H. Z., Junaini, S. N., Khan, A. S., & Mustafa, W. A. (2025). Data-driven agriculture: Unveiling the power of Internet of Things and data analytics for transformative farming practices. Journal of Sensors, 2025, Article 6205646. https://doi.org/10.1155/js/6205646
3. Balasundram, S. K., Shamshiri, R. R., Sridhara, S., & Rizan, N. (2023). The role of digital agriculture in mitigating climate change and ensuring food security: An overview. Sustainability, 15(6), 5325. https://doi.org/10.3390/su15065325
4. Bist, D. R., Chapagaee, P., Kunwar, A., & Khatri, L. (2026). The role of big data in sustainable agriculture: Advancing environmental sustainability in precision farming systems. Cogent Food & Agriculture, 12(1). https://doi.org/10.1080/23311932.2026.2620180
5. Cravero, A., Bustamante, A., Negrier, M., & Galeas, P. (2022). Agricultural big data architectures in the context of climate change: A systematic literature review. Sustainability, 14(13), 7855. https://doi.org/10.3390/su14137855
6. Cravero, A., Pardo, S., Galeas, P., López Fenner, J., & Caniupán, M. (2022). Data type and data sources for agricultural big data and machine learning. Sustainability, 14(23), 16131. https://doi.org/10.3390/su142316131
7. El Chami, D., & El Moujabber, M. (2024). Sustainable agriculture and climate resilience. Sustainability, 16(1), 113. https://doi.org/10.3390/su16010113
8. Gugissa, D. A., Abro, Z., & Tefera, T. (2022). Achieving a climate-change resilient farming system through push–pull technology: Evidence from maize farming systems in Ethiopia. Sustainability, 14(5), 2648. https://doi.org/10.3390/su14052648
9. Halshoy, H. S., Hama, J. R., & Braim, S. A. (2025). Selective agricultural and environmental practices to sustain food production and mitigate climate change. Cogent Food & Agriculture, 11(1). https://doi.org/10.1080/23311932.2025.2562179
10. Kamara, A. Y., Oyinbo, O., Manda, J., Kamara, A., Idowu, E. O., & Mbavai, J. J. (2024). Beyond average: Are the yield and income impacts of adopting drought-tolerant maize varieties heterogeneous? Climate and Development, 16(1), 67–76. https://doi.org/10.1080/17565529.2023.2178840
11. Kumari, K., Mirzakhani Nafchi, A., Mirzaee, S., & Abdalla, A. (2025). AI-driven future farming: Achieving climate-smart and sustainable agriculture. AgriEngineering, 7(3), 89. https://doi.org/10.3390/agriengineering7030089
12. Luyckx, M., & Reins, L. (2022). The future of farming: The (non)-sense of big data predictive tools for sustainable EU agriculture. Sustainability, 14(20), 12968. https://doi.org/10.3390/su142012968

13. Makokha, J. W., Barasa, P. W., & Khamala, G. W. (2025). Enhancing climate resilience: A data-driven North Rift weather prediction system for real-time forecasting and agricultural decision support. Heliyon, 11(4), e42549. https://doi.org/10.1016/j.heliyon.2025.e42549
14. Makapela, M., Alexander, G., & Tshelane, M. (2025). Enhancing agricultural productivity among emerging farmers through data-driven practices. Sustainability, 17(10), 4666. https://doi.org/10.3390/su17104666
15. Miller, T., Mikiciuk, G., Durlik, I., Mikiciuk, M., Łobodzińska, A., & Śnieg, M. (2025). The IoT and AI in agriculture: The time is now—A systematic review of smart sensing technologies. Sensors, 25(12), 3583. https://doi.org/10.3390/s25123583
16. Nan, V. A., Badea, G., Badea, A. C., & Grădinaru, A. P. (2025). A systematic review of AI-based classifications used in agricultural monitoring in the context of achieving the Sustainable Development Goals. Sustainability, 17(19), 8526. https://doi.org/10.3390/su17198526
17. Nirmaladevi, S., & Jagatheswari, S. (2025). A data-driven machine learning framework for predicting total agricultural food grain yield. Results in Engineering, 27, 106790. https://doi.org/10.1016/j.rineng.2025.106790
18. Nyasulu, C., Diattara, A., Traore, A., Deme, A., & Ba, C. (2022). Towards resilient agriculture to hostile climate change in the Sahel region: A case study of machine learning-based weather prediction in Senegal. Agriculture, 12(9), 1473. https://doi.org/10.3390/agriculture12091473
19. Qiang, Y., Zou, L., & Cai, H. (2023). Big Earth Data for quantitative measurement of community resilience: current challenges, progresses and future directions. Big Earth Data, 7(4), 1035–1057. https://doi.org/10.1080/20964471.2023.2273594
20. Srinidhi, A., Meuwissen, M. P. M., Jadhav, S., Ludwig, F., Dadas, D., Nikam, N., & Werners, S. E. (2025). Climate resilient development pathways for farmer producer organizations in semi-arid India. Climate and Development, 17(9), 775–790. https://doi.org/10.1080/17565529.2025.2459063
21. Suresh, D., Choudhury, A., Zhang, Y., Zhao, Z., & Shaw, R. (2024). The role of data-driven agritech startups—The case of India and Japan. Sustainability, 16(11), 4504. https://doi.org/10.3390/su16114504
22. Madhura Alange, Vaishnavi Davari, Tushar Dhanke, Ganesh Dombale, & M. S. Chakote. (2026). Smart Crop AI Advisor for Precision Farming and Pest Detection. International Journal on Advanced Computer Theory and Engineering, 15(1), 155–161. https://doi.org/10.65521/ijacte.v15i1.2933
23. Mali, S., Sarnobat, S., Khot, T., Pawar, K., & Sartape, P. (2026). Socifarm Community Module: Design and Implementation of a Public-Private Agricultural Community Platform with Role-Based Access Control. Multidisciplinary Journal of Research in Engineering and Technology, 13(1), 163–171. https://doi.org/10.65521/mjret.v13i1.3111
24. Marathe, N., Patil, M., Pawale, S. S., Jondhale, P., & Pawale, S. J. (2026). AI-Powered Crop Disease Detection System. Multidisciplinary Journal of Research in Engineering and Technology, 13(1S), 36–42. https://doi.org/10.65521/mjret.v13i1S.3027
25. Vélez, S., & Álvarez, S. (2024). Harnessing data-driven technologies for sustainable farming practices. Agronomy, 14(12), 2969. https://doi.org/10.3390/agronomy14122969
26. Weraikat, D., Šorič, K., Žagar, M., & Sokač, M. (2024). Data analytics in agriculture: Enhancing decision-making for crop yield optimization and sustainable practices. Sustainability, 16(17), 7331. https://doi.org/10.3390/su16177331