BIG DATA FOR SUSTAINABLE CROP HEALTH: A CASE STUDY ON CGIAR

Barkavi A/P P Cheven (23093149)
Tan Kai Qi (23051355)
Ng Chee Kang (23051336)
Seh Chia Shin (17202838)

Faculty of Computer Science and Information Technology, Universiti Malaya
50603 Kuala Lumpur, Federal Territory of Kuala Lumpur, Malaysia

barkavicheven@gmail.com*, secondauthor@email.com, and lastauthor@email.com (9 pts)

Corresponding Author : Barkavi A/P P Cheven (23093149), Tan Kai Qi (23051355), Ng Chee Kang (23051336), Seh Chia Shin (17202838)
Received: 09 April 2025, 12 April 2025, Accepted: 13 April 2025
Published online: 14 April 2025

Abstract: Please write here (200-250 words)

Keywords: Keyword 1; Keyword 2; Keyword 3; Keyword 4; Keyword 5. (10 pts separately by a semicolon).

1. BACKGROUND STUDY

The Consultative Group on International Agricultural Research (CGIAR) stands as an unparalleled global consortium that has metamorphosed into the world's most extensive agricultural innovation network, orchestrating a symphony of scientific endeavors aimed at fortifying food security, alleviating poverty, and enhancing nutritional outcomes across the developing world **(CGIAR, 2023a)**. Established in 1971 through the collaborative vision of the Rockefeller and Ford Foundations, this venerable institution has undergone a remarkable evolution, transcending its initial conceptualization to emerge as a multifaceted constellation of research centers, partnerships, and initiatives that collectively constitute a paradigmatic shift in agricultural research and development **(Dalrymple, 2008).** The organizational architecture of CGIAR encompasses a network of 15 independent research centers, strategically distributed across 70 countries, thereby creating an expansive geographical footprint that facilitates contextualized research responsive to diverse agroecological conditions and socioeconomic realities **(CGIAR, 2023b)**. The genesis of CGIAR can be traced to the past period of the 1960s, characterized by predictions of widespread famine and agricultural insufficiency in developing regions **(Pingali, 2012).** This existential threat catalyzed the establishment of an institutional framework dedicated to the systematic application of scientific methodologies to agricultural challenges in resource-constrained environments. The evolutionary trajectory of CGIAR has been marked by successive waves of organizational reconfiguration, each representing a strategic response to emerging challenges and opportunities in the global agricultural landscape **(Byerlee & Lynam, 2020).** The most recent metamorphosis, initiated in 2020, has culminated in the "One CGIAR" framework, a transformative recalibration that consolidates governance structures, harmonizes research agendas, and optimizes resource allocation to maximize institutional impact and operational efficiency **(CGIAR, 2022).**

As for the geographical footprint and operational modalities, the geographical constellation of CGIAR's research centers constitutes a strategic distribution that encompasses diverse agroecological zones, from the arid landscapes of the Middle East to the tropical ecosystems of Southeast Asia, and from the highlands of Latin America to the savannahs of Sub-Saharan Africa **(Özgediz, 2012).** This expansive footprint enables the organization to conduct contextualized research that addresses the heterogeneous challenges facing agricultural systems across the developing world. The operational modalities of CGIAR are characterized by a multidimensional approach that integrates field-based experimentation, laboratory analysis, computational modeling, and participatory methodologies, thereby creating a comprehensive research ecosystem that transcends conventional disciplinary boundaries **(Hall et al., 2018).**

Moreover, CGIAR's research portfolio encompasses a diverse array of thematic areas, including crop improvement, livestock systems, aquaculture, agroforestry, climate adaptation, nutrition, policy analysis, and gender equity, all unified by a common commitment to sustainable agricultural development **(CGIAR, 2023c).** This multifaceted research agenda is operationalized through a series of research programs and platforms that facilitate interdisciplinary collaboration and knowledge integration across the organizational network. The thematic orientation of CGIAR's research is guided by a strategic framework that prioritizes impact pathways aligned with the Sustainable Development Goals, particularly those related to zero hunger, poverty reduction, gender equality, climate action, and responsible consumption and production **(Steiner et al., 2020).**

Furthermore, recognizing the transformative potential of digital technologies and data-driven approaches, CGIAR has embarked on an ambitious journey of digital transformation that permeates all aspects of its research and operational activities **(Janssen et al., 2021).** The CGIAR Platform for Big Data in Agriculture, launched in 2017, represents a cornerstone of this digital strategy, serving as a catalytic mechanism for harnessing the power of big data analytics, artificial intelligence, machine learning, and remote sensing to address complex agricultural challenges **(Jiménez et al., 2019).** This platform facilitates the collection, integration, analysis, and visualization of diverse data streams, ranging from genomic sequences to satellite imagery, and from household surveys to meteorological measurements, thereby creating an unprecedented repository of agricultural intelligence that informs decision-making processes at multiple scales **(CGIAR, 2023d).** To be more profound, the digital ecosystem cultivated by CGIAR encompasses a constellation of innovative initiatives, including the development of mobile applications for agricultural advisory services, the deployment of sensor networks for environmental monitoring, the utilization of drone technology for crop surveillance, and the implementation of blockchain systems for supply chain transparency **(Deichmann et al., 2016).** These technological interventions collectively constitute a paradigmatic shift in agricultural research methodologies, enabling the generation of insights and solutions that would be unattainable through conventional approaches. The organization's commitment to open data principles further amplifies the impact of these digital initiatives, with the GARDIAN platform serving as a centralized repository for CGIAR's research outputs, thereby democratizing access to agricultural knowledge and fostering collaborative innovation across geographical and institutional boundaries **(Arnaud et al., 2020).**

Besides, CGIAR's operational philosophy is predicated on the recognition that complex agricultural challenges necessitate collaborative approaches that transcend institutional silos and disciplinary boundaries. Consequently, the organization has cultivated an extensive network of partnerships with diverse stakeholders, including national agricultural research systems, academic institutions, governmental agencies, non-governmental organizations, private sector entities, and farmer associations **(CGIAR, 2023e).** These collaborative relationships facilitate the co-creation of knowledge, the contextualization of research outputs, and the scaling of innovations, thereby enhancing the relevance, applicability, and impact of CGIAR's research portfolio. The partnership dynamics within the CGIAR ecosystem are characterized by a multidirectional flow of knowledge, resources, and expertise, creating synergistic relationships that leverage the complementary strengths of diverse actors within the agricultural innovation landscape **(Spielman et al., 2021).** This collaborative approach is exemplified by initiatives such as the Excellence in Breeding Platform, which brings together CGIAR centers, national partners, and private sector entities to accelerate genetic gains in crop improvement programs through the application of cutting-edge technologies and methodologies **(Ribaut & Ragot, 2019).**

The multifaceted impact of CGIAR's research and development activities spans multiple dimensions, from the tangible outputs of improved crop varieties and livestock breeds to the intangible outcomes of enhanced institutional capacities and policy frameworks **(Renkow & Byerlee, 2010). The** organization's contributions to global food security are quantitatively substantiated by the widespread adoption of improved crop varieties developed by CGIAR centers, with an estimated 70% of wheat area and 45% of rice area in developing countries planted with CGIAR-derived varieties **(Lantican et al., 2016).** The economic impact of these agricultural innovations is equally impressive, with studies indicating that for every dollar invested in CGIAR research, approximately $17 in benefits are generated for developing countries, representing an exceptional return on investment in the realm of international development **(Alston et al., 2020).** Beyond these quantifiable metrics, CGIAR's impact extends to broader developmental dimensions, including the enhancement of nutritional outcomes through biofortified crops, the mitigation of environmental degradation through sustainable land management practices, the empowerment of women through gender-responsive agricultural technologies, and the strengthening of resilience to climate variability through adaptive farming systems **(CGIAR, 2023f).** These multidimensional contributions collectively position CGIAR as an indispensable actor in the global effort to achieve sustainable agricultural development and food security in the face of escalating challenges such as population growth, climate change, and resource constraints.

Regarding the big data applications in agricultural research, the integration of big data methodologies within CGIAR's research framework represents a paradigmatic shift in agricultural science, enabling the organization to harness the power of data-driven insights to address complex challenges across the agricultural value chain **(Janssen et al., 2021).** The voluminous nature of agricultural data, characterized by its heterogeneity, velocity, and complexity, necessitates sophisticated analytical approaches that transcend conventional statistical methodologies. CGIAR's big data initiatives encompass a diverse array of applications, including predictive modeling for crop yield forecasting, genomic selection for accelerated breeding, remote sensing for environmental monitoring, and machine learning for pest and disease detection **(Jiménez et al., 2019).** The organization's commitment to big data is operationalized through various platforms and initiatives, including the CGIAR Platform for Big Data in Agriculture, which serves as a centralized hub for data-related activities across the CGIAR network **(CGIAR, 2023d).** This platform facilitates the development of data standards, the enhancement of analytical capacities, the promotion of open data principles, and the cultivation of a data-driven culture within the agricultural research community. The platform's Communities of Practice, focusing on areas such as crop modeling, geospatial analysis, and socioeconomic data, provides collaborative spaces for researchers to exchange methodologies, share insights, and co-create solutions to common challenges in agricultural data science **(Arnaud et al., 2020).**

Apart from that, the CGIAR's application of big data in agricultural research is exemplified by initiatives such as the "Computer Vision for Crop Disease" project, which utilizes machine learning algorithms to analyze images of plant leaves for early detection of diseases, thereby enabling timely interventions that minimize crop losses and reduce reliance on chemical pesticides **(CGIAR, 2022).** Similarly, the organization's work on digital soil mapping combines satellite imagery, ground-based measurements, and machine learning techniques to generate high-resolution maps of soil properties, providing valuable information for targeted interventions in soil fertility management and land use planning **(Hengl et al., 2017).**

As CGIAR navigates the complex landscape of 21st-century agricultural challenges, its strategic vision is guided by a commitment to harnessing the transformative potential of digital technologies and data-driven approaches to accelerate agricultural innovation and enhance developmental impact **(CGIAR, 2022).** The organization's future trajectory encompasses several key dimensions, including the expansion of digital advisory services to reach millions of smallholder farmers, the enhancement of predictive capabilities for climate-related risks, the acceleration of genetic gains through genomic selection, and the optimization of resource allocation through precision agriculture techniques **(Steiner et al., 2020).** The "One CGIAR" transformation represents a strategic recalibration aimed at enhancing the organization's capacity to address emerging challenges and capitalize on new opportunities in the agricultural research landscape **(CGIAR, 2023g).** This institutional metamorphosis is characterized by a unified governance structure, an integrated operational framework, and a harmonized research agenda, all designed to maximize the collective impact of CGIAR's diverse capabilities and resources. The digital dimension of this transformation is particularly significant, with investments in data infrastructure, analytical capabilities, and digital skills positioned as central elements of the organization's strategic roadmap.

In conclusion, CGIAR stands as a paragon of agricultural innovation and data-driven transformation, leveraging its extensive network, multidisciplinary expertise, and technological capabilities to address the complex challenges facing agricultural systems in the developing world. Through its integration of big data methodologies, digital technologies, and collaborative approaches, the organization continues to generate impactful solutions that enhance food security, reduce poverty, improve nutrition, and promote environmental sustainability across diverse agroecological and socioeconomic contexts. As the agricultural landscape evolves in response to emerging challenges such as climate change, population growth, and resource constraints, CGIAR's adaptive capacity and innovative spirit position as an indispensable actor in the global effort to achieve sustainable agricultural development and food security for future generations.

2. CHARACTERISTICS OF BIG DATA

The Consultative Group on International Agricultural Research (CGIAR) stands as a preeminent global partnership that unites international organizations engaged in research for a food-secure future. In the contemporary agricultural landscape, CGIAR has emerged as a vanguard institution leveraging big data analytics to revolutionize agricultural practices, enhance food security, and mitigate the deleterious impacts of climate change on global food systems **(CGIAR, 2022).** The organization's strategic deployment of big data methodologies exemplifies the transformative potential of data-driven approaches in addressing multifaceted agricultural challenges across diverse geographical and socioeconomic contexts.

CGIAR's research initiatives generate and process colossal volumes of agricultural data that transcend traditional data management capabilities. The organization's data repositories encompass extensive genomic sequences from thousands of crop varieties and their wild relatives, with each genome containing billions of base pairs **(Halewood et al., 2018)**. High-resolution satellite imagery spanning millions of hectares of agricultural land, with temporal resolutions capturing daily changes over decades **(Jiménez et al., 2021),** further contributes to this data deluge. Comprehensive climate datasets incorporating historical records and predictive models with petabyte-scale storage requirements **(Campbell et al., 2019)** and phenotypic data from field trials conducted across 15 research centers in over 70 countries, generating terabytes of observational data annually **(CGIAR Platform for Big Data in Agriculture, 2021)**, round out this massive data ecosystem. This voluminous data necessitates advanced computational infrastructure, with CGIAR's data centers processing approximately 2.5 petabytes of agricultural data annually, a figure that continues to expand exponentially as sensor networks and field monitoring systems proliferate across global agricultural landscapes **(Arnaud et al., 2020)**.

The heterogeneity of CGIAR's data ecosystem represents a quintessential characteristic of big data in agricultural research. The organization integrates disparate data types including structured data: tabular datasets containing crop yield statistics, meteorological measurements, and soil composition analyses with precise schema definitions **(Jiménez et al., 2020)**. Unstructured data, including textual reports, farmer interviews, and indigenous knowledge repositories that require natural language processing for meaningful extraction **(Porciello et al., 2021),** adds another layer of complexity. Semi-structured data, such as XML and JSON files containing sensor readings from IoT devices deployed across experimental farms with hierarchical data organization **(Delerce et al., 2020),** must also be integrated. Spatial data representing land use patterns, crop distribution, and ecological zones with complex topological relationships **(Rosenstock et al., 2019)** and temporal data tracking crop growth stages, disease progression, and climate patterns with varying granularity and periodicity **(Loboguerrero et al., 2018)** complete this diverse data landscape. CGIAR's Platform for Big Data in Agriculture has pioneered the integration of these heterogeneous data types through innovative data harmonization frameworks and ontological standards, enabling cross-domain analytics that were previously unattainable (**CGIAR Platform for Big Data in Agriculture, 2021).**

The velocity dimension of CGIAR's big data operations manifests in the rapid acquisition, processing, and dissemination of agricultural information across temporal scales. Real-time sensor networks deployed across experimental farms transmit environmental parameters at sub-minute intervals, generating continuous data streams that require immediate processing **(Delerce et al., 2020).** Automated weather stations integrated with CGIAR's data infrastructure provide hourly meteorological updates across diverse agroecological zones, enabling timely interventions during extreme weather events **(Campbell et al., 2019).** Satellite imagery with daily revisit capabilities delivers near-real-time monitoring of crop conditions, necessitating high-throughput processing pipelines to extract actionable insights within operational timeframes **(Jiménez et al., 2021).** Mobile applications developed by CGIAR enable farmers to report pest outbreaks and disease symptoms instantaneously, creating dynamic data streams that inform rapid response mechanisms **(Porciello et al., 2021).** To accommodate these high-velocity data streams, CGIAR has implemented distributed computing architectures and stream processing frameworks that enable real-time analytics, reducing the latency between data acquisition and actionable insights from days to minutes **(Arnaud et al., 2020).**

The veracity dimension addresses the inherent uncertainties and quality concerns in CGIAR's agricultural big data ecosystem. Data quality assessment frameworks implemented across CGIAR centers establish standardized protocols for evaluating the accuracy, completeness, and consistency of agricultural datasets **(Leonelli et al., 2017).** Automated validation algorithms detect anomalous values and inconsistencies in sensor readings, flagging potential errors for human review before incorporation into analytical pipelines **(Delerce et al., 2020).** Metadata standards adhering to FAIR principles (Findable, Accessible, Interoperable, Reusable) enhance the provenance tracking and reliability assessment of datasets shared across the CGIAR network **(Wilkinson et al., 2016).** Uncertainty quantification methodologies accompany predictive models, providing confidence intervals and reliability metrics for agricultural forecasts that inform risk management strategies **(Campbell et al., 2019).** Cross-validation techniques comparing satellite-derived indicators with ground-truth measurements establish accuracy assessments for remote sensing products used in crop monitoring applications **(Jiménez et al., 2021).** CGIAR's commitment to data veracity is exemplified by its establishment of data quality certification processes that ensure research outputs meet rigorous standards before informing agricultural policies and interventions **(CGIAR Platform for Big Data in Agriculture, 2021).**

The variability dimension reflects CGIAR's capacity to accommodate fluctuating data patterns and evolving analytical requirements in agricultural research. Seasonal variations in data collection intensity, with peak periods during growing seasons generating up to five times the data volume compared to off-season periods, necessitate elastic computing resources **(Rosenstock et al., 2019).** Climate change-induced shifts in agricultural patterns create evolving data signatures that require adaptive analytical frameworks capable of detecting emerging trends and anomalies **(Loboguerrero et al., 2018).** Market dynamics and policy changes influence farmer behavior and agricultural practices, introducing variability in socioeconomic datasets that inform CGIAR's intervention strategies **(Porciello et al., 2020).** Pest and disease outbreaks create sudden spikes in monitoring data and farmer reports, triggering automated scaling of computational resources to accommodate surge capacity requirements **(Delerce et al., 2020).** To address these variability challenges, CGIAR has implemented adaptive data processing pipelines with dynamic resource allocation, ensuring analytical continuity despite fluctuating data characteristics and processing demands **(Arnaud et al., 2020).**

The visualization dimension represents CGIAR's approach to rendering complex agricultural data comprehensible and actionable for diverse stakeholders. Interactive dashboards developed by CGIAR's data scientists transform multidimensional crop performance data into intuitive visual interfaces accessible to researchers, policymakers, and agricultural extension officers **(CGIAR Platform for Big Data in Agriculture, 2021).** Geospatial visualization platforms integrate satellite imagery, climate data, and crop models to generate dynamic maps of agricultural productivity and vulnerability across temporal and spatial scales **(Jiménez et al., 2021).** Network visualization tools elucidate complex relationships between genetic traits, environmental factors, and crop performance, facilitating the identification of resilient varieties for climate adaptation **(Halewood et al., 2018).** Temporal visualization techniques illustrate long-term trends in agricultural productivity and climate patterns, highlighting critical intervention points for sustainable intensification strategies **(Campbell et al., 2019).** CGIAR's visualization innovations extend beyond conventional approaches to include virtual reality applications that immerse stakeholders in simulated agricultural landscapes, enhancing comprehension of complex ecological interactions and intervention impacts **(Arnaud et al., 2020).**

The value dimension encapsulates CGIAR's ultimate objective: transforming big agricultural data into tangible benefits for global food systems and rural livelihoods. Precision agriculture recommendations derived from big data analytics have increased crop yields by 15-25% while reducing input costs by 10-20% across CGIAR's intervention sites in South Asia and Sub-Saharan Africa **(Jiménez et al., 2020).** Early warning systems powered by real-time data integration have enabled preemptive responses to pest outbreaks, preventing estimated crop losses valued at $45-60 million annually across CGIAR's partner countries **(Delerce et al., 2020).** Climate-smart agriculture strategies informed by big data models have enhanced the resilience of over 8 million smallholder farmers to climate variability, reducing crop failure risks by 30-40% during extreme weather events **(Loboguerrero et al., 2018).** Data-driven policy recommendations have influenced agricultural investment decisions and resource allocations in over 30 countries, optimizing the impact of limited resources on food security outcomes **(Porciello et al., 2020).** CGIAR's value proposition extends beyond immediate agricultural productivity to encompass broader societal benefits, including enhanced environmental sustainability, improved nutrition outcomes, and strengthened rural economies through data-empowered decision-making **(CGIAR, 2022).**

The seven dimensions of big data which are volume, variety, velocity, veracity, variability, visualization, and value collectively define CGIAR's innovative approach to agricultural research and development. By harnessing these multifaceted characteristics, CGIAR has established a transformative paradigm that transcends traditional agricultural research methodologies, enabling evidence-based interventions that address the complex challenges of global food security in the context of climate change and population growth. As agricultural systems continue to evolve in response to global challenges, CGIAR's big data ecosystem represents a vanguard approach to generating, integrating, and translating agricultural information into actionable knowledge. The organization's strategic investment in data infrastructure, analytical capabilities, and knowledge dissemination mechanisms exemplifies the transformative potential of big data in revolutionizing agricultural practices and enhancing food security across diverse geographical and socioeconomic contexts.

3. SIX PHASES OF BIG DATA

The growth of big data has significantly changed numerous industry sectors and institutions, including agriculture. CGIAR (Consultative Group on International Agricultural Research) is one of the institutions using big data for agricultural development. CGIAR able to enhance decision-making processes, agricultural productivity and disease detection across developing countries through the big data platform. This organization employs a comprehensive six-phase big data approach to address real-world challenges in agriculture, such as early disease detection, food security and sustainability (*CGIAR Results Dashboard - CGIAR*, n.d.).

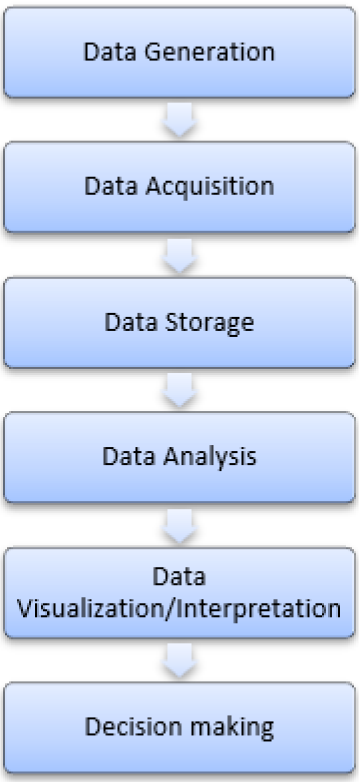


Figure 1: Six Phases of Big Data

Data Generation
CGIAR consists of raw agricultural data that is produced through various sources and methods. These include field experiments, sensor data, drone and satellite images, farmer reports and mobile applications (*CGIAR BIG DATA Platform - CGIAR Platform for Big Data in Agriculture*, n.d.). CGIAR collaborates with local farmers and researchers to generate reliable and diverse datasets that can represent real-world agricultural conditions,
Other output datasets from the result framework are aligned with 5 impact areas to achieve the Sustainable Development Goals (SDGs) (*CGIAR System Organization CGIAR Performance and Results Management*, 2022):

- Climate adaptation and mitigation
- Environmental health and biodiversity
- Gener quality, youth and social inclusion
- Nutrition, health and food security
- Poverty reduction, livelihoods and job

Based on the Kaggle dataset, it is generated by capturing high-resolution crop images from farms (*CGIAR Computer Vision for Crop Disease*, n.d.). The dataset provides a comprehensive visual dataset that can allow researchers to be used for machine learning-based disease studies. The images with ID that show the disease symptoms allow researchers to study the crop disease by machine learning model based on the rust condition of the crop.

Data Acquisition
The generated data is collected and centralized for further processing. CGIAR utilizes different digital tools and platforms for data gathering. The tools include mobile apps, Open Data Kit (ODK) and automated sensors. In the crop disease detection project, images and metadata, for example crop type, geographic location, and disease label are uploaded to centralized databases through cloud-based platforms. By collaborating with multiple agricultural institutions and research centers, CGIAR ensures that the data acquisition process is extensive and continuous, covering multiple countries and crop types.

Data Storage
In 2014, CGIAR collaborated with AWS to make its Global Circulation Models (GCM) data accessible and support research on climate change impacts on agriculture (*Earth Science on AWS with New CGIAR and Landsat Public Data Sets | AWS News Blog*, n.d.). CGIAR also uses AWS architecture for projects such as Digital Earth Africa program based on the FAIR data principles (*Digital Twin for Water Management Handed over for Trial in the Limpopo River Basin - CGIAR*, n.d.). Furthermore, CGIAR has leveraged cloud-based platforms such as Google Earth Engine, which operates on Google's cloud infrastructure, to process and analyze geospatial data on a global scale (Basel et al., 2023). The effort reflects that the CGIAR's commitment to using cloud storage services to improve the accessibility and scalability of research data.

Data Analysis
At the next phase of data storage is data analysis, it is the phase where raw data is transformed into useful insights. CGIAR is using advanced analytical techniques such as statistical modeling, machine learning (ML) and deep learning (DL) algorithms. For example, convolutional neural networks (CNNs) are used to analyze crop images and identify disease patterns. By training models on thousands of labeled images, CGIAR can develop automated systems that detect diseases with high accuracy (*CGIAR Computer Vision for Crop Disease*, n.d.). Furthermore, predictive analytics is used to forecast crop yields, disease outbreaks and the effects of climate change. The use of big data analytics helps CGIAR to derive meaningful conclusions that guide field practices and policy decisions. Another example from (Anna Alex & Kanavalli, 2019), the data such as soil moisture, average rainfall and soil nutrients are used to predicts whether the fertilizers will cause crop disease. Based on the research, the Hadoop system was used to store and analyze the data by using ML techniques.

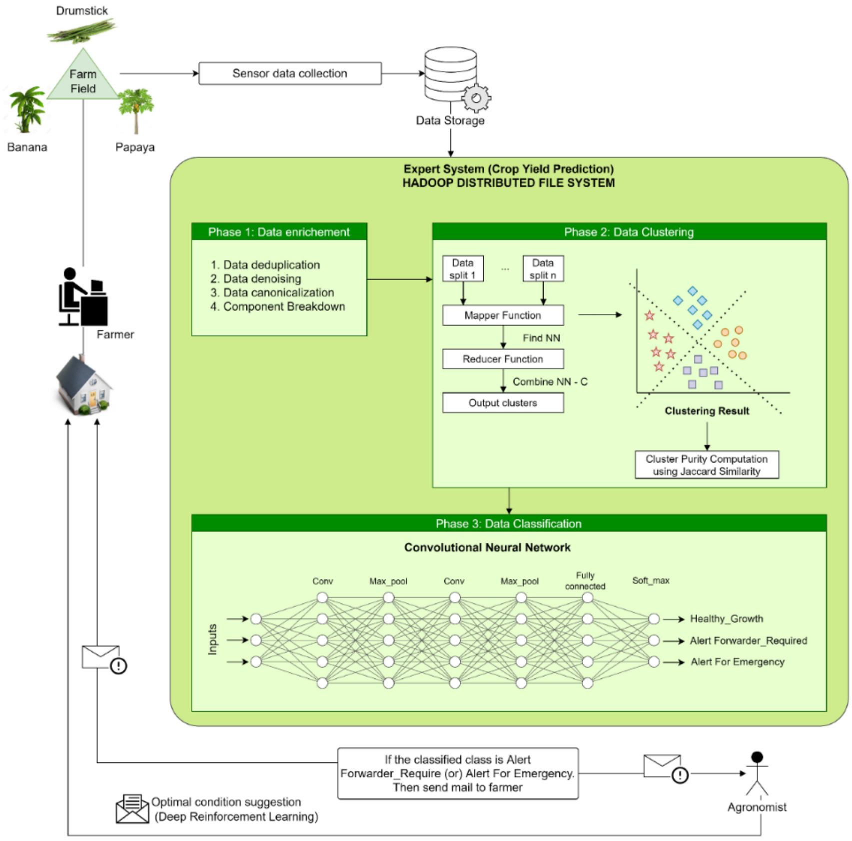


Figure 2: Big data architecture for the fertilizer analysis and yield prediction (Cravero et al., 2022)

Data Visualization and Interpretation

Once analyzed, data must be visualized effectively to make it understandable and actionable. CGIAR develops dashboards, interactive maps and infographics to present analysis results to farmers, researchers and policymakers. The most common tool used by CGIAR is its custom web applications. From the web application, the heat map from the result dashboard shows the projects' outcomes that have been worked on worldwide. These visualizations help stakeholders to track, visualize and communicate the progress and outcomes of CGIAR's research and innovation activities about global food security, climate resilience, and agricultural development. The dashboard showcases the global impact from CGIAR in crop productivity, climate change adaptation, gender equality and nutrition and food system for decision-making for a certain research or project.

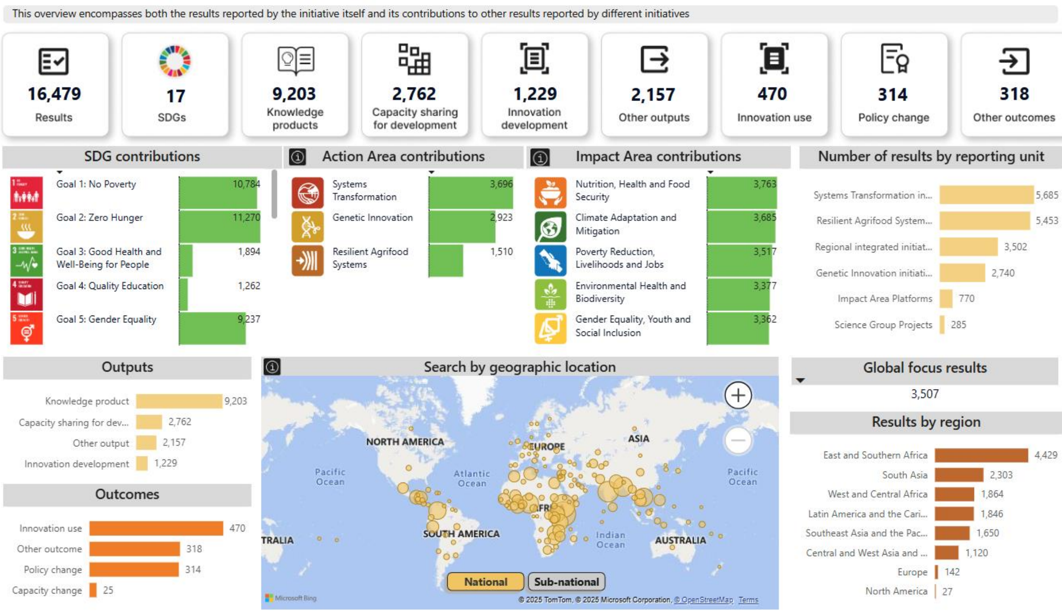


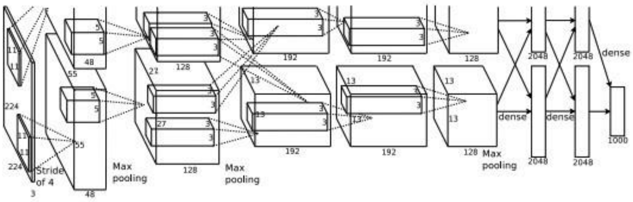
Figure 3: CGIAR’s Result Dashboard (*CGIAR Results Dashboard - CGIAR*, n.d.)

Decision Making
The last phase of the big data which involves decision-making based on data visualization. At CGIAR, big data allows stakeholders to make decisions such as when to trigger farmers on the potential disease risks, where to implement pest control measures and how to breed disease-resistant crop varieties. Besides, the policy decisions from the government may also be influenced by CGIAR's findings or data, especially in regions having food insecurity. By integrating data-driven recommendations into everyday farming and policy frameworks, CGIAR able to enhance resilience, productivity and sustainability in agriculture.

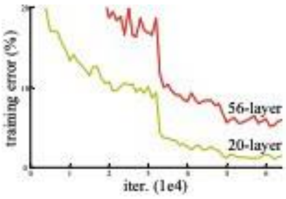
In summary, the six phases of big data, like data generation, acquisition, storage, analysis, visualization and decision-making, are effectively utilized by CGIAR to address agricultural challenges. With real-world datasets like the crop disease detection project, CGIAR is showing the advantage of big data usage that can transform farming practices, improve food security and enhance global agricultural sustainability to meet the SDG goals. These efforts highlight how importance of data-driven innovation in addressing the complex and dynamic needs of modern agriculture.

4. THE NECESSITY OF BIG DATA SOFTWARE

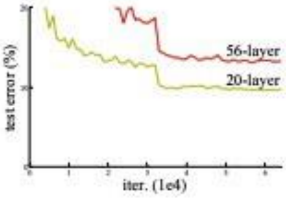
The necessities of big data software for the identified topic / company / institution are discussed in detail, supported by relevant references.



(a) Example from [1]



(b) Example from [2]



(c) Example from [2]

Fig. 1. Example of placing a figure with experimental results. Adopting figures from [1] and [2]

5. KEY OBSTACLES FOR BIG DATA SOLUTION

The key obstacles of big data solution for the identified topic / company / institution are discussed in detail, supported by relevant references.

6. CONCLUSION

You can include figures to explain your idea too, for example as follows in Figure 1.
Please ensure all content is within 10 pages!
Latex guide: [3] and [4]

7. REFERENCES

Anna Alex, S., & Kanavalli, A. (2019). Intelligent Computational Techniques for Crops Yield Prediction and Fertilizer Management over Big Data Environment. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, *12*, 2278–3075. https://doi.org/10.35940/ijitee.L2622.1081219

Basel, A. M., Nguyen, K. T., Arnaud, E., & Craparo, A. C. W. (2023). The foundations of big data sharing: A CGIAR international research organization perspective. In *Frontiers in Environmental Science* (Vol. 11). Frontiers Media S.A. https://doi.org/10.3389/fenvs.2023.1107393

CGIAR BIG DATA Platform - CGIAR Platform for Big Data in Agriculture. (n.d.). Retrieved April 7, 2025, from https://bigdata.cgiar.org/

CGIAR Computer Vision for Crop Disease. (n.d.). Retrieved April 7, 2025, from https://www.kaggle.com/datasets/shadabhussain/cgiar-computer-vision-for-crop-disease/data

CGIAR Results Dashboard - CGIAR. (n.d.). Retrieved April 6, 2025, from https://www.cgiar.org/food-security-impact/results-dashboard/

CGIAR System Organization CGIAR Performance and Results Management. (2022).

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 2022, Vol. 14, Page 16131*, *14*(23), 16131. https://doi.org/10.3390/SU142316131

Digital twin for water management handed over for trial in the Limpopo River Basin - CGIAR. (n.d.). Retrieved April 8, 2025, from https://www.cgiar.org/news-events/news/digital-twin-handover-limpopo/

Earth Science on AWS with new CGIAR and Landsat Public Data Sets | AWS News Blog. (n.d.). Retrieved April 8, 2025, from https://aws.amazon.com/blogs/aws/earth-science-data-sets-on-aws/

References

Arnaud, E., Laporte, M. A., Kim, S., Aubert, C., Leonelli, S., Miro, B., Cooper, L., Jaiswal, P., Kruseman, G., Shrestha, R., Buttigieg, P. L., Mungall, C. J., Pietragalla, J., Agbona, A., Muliro, J., Detras, J., Hualla, V., Rathore, A., Das, R. R., ... & King, B. (2020). The Ontologies Community of Practice: A CGIAR Initiative for Big Data in Agrifood Systems. Patterns, 1(7), 100105.

Campbell, B. M., Hansen, J., Rioux, J., Stirling, C. M., Twomlow, S., & Wollenberg, E. (2019). Urgent action to combat climate change and its impacts (SDG 13): transforming agriculture and food systems. Current Opinion in Environmental Sustainability, 34, 13-20.

CGIAR. (2022). CGIAR 2030 Research and Innovation Strategy: Transforming food, land, and water systems in a climate crisis. CGIAR System Organization.

CGIAR Platform for Big Data in Agriculture. (2021). Annual Report 2020: Digital Dynamism for Adaptive Food Systems. CGIAR.

Delerce, S., Dorado, H., Grillon, A., Rebolledo, M. C., Prager, S. D., Patiño, V. H., Garcés Varón, G., & Jiménez, D. (2020). Assessing weather-yield relationships in rice at local scale using data mining approaches. PloS One, 15(5), e0233246.

Halewood, M., Lopez Noriega, I., Ellis, D., Roa, C., Rouard, M., & Sackville Hamilton, R. (2018). Using digital sequence information to achieve the crop diversity system management goals. Global Food Security, 18, 9-14.

Janssen, S. J., Porter, C. H., Moore, A. D., Athanasiadis, I. N., Foster, I., Jones, J. W., & Antle, J. M. (2017). Towards a new generation of agricultural system data, models and knowledge products: Information and communication technology. Agricultural Systems, 155, 200-212.

Jiménez, D., Dorado, H., Cock, J., Prager, S. D., Delerce, S., Grillon, A., Bejarano, M. A., Benavides, H., & Jarvis, A. (2020). From observation to information: Data-driven understanding of on farm yield variation. PloS One, 15(4), e0229881.

Jiménez, D., Delerce, S., Dorado, H., Cock, J., Muñoz, L. A., Agamez, A., & Jarvis, A. (2021). A scalable scheme to implement data-driven agriculture for small-scale farmers. Global Food Security, 30, 100559.

Leonelli, S., Davey, R. P., Arnaud, E., Parry, G., & Bastow, R. (2017). Data management and best practice for plant science. Nature Plants, 3(6), 1-4.

Loboguerrero, A. M., Campbell, B. M., Cooper, P. J., Hansen, J. W., Rosenstock, T., & Wollenberg, E. (2018). Food and earth systems: priorities for climate change adaptation and mitigation for agriculture and food systems. Sustainability, 11(5), 1372.

Porciello, J., Ivanina, M., Islam, M., Einarson, S., & Hirsch, H. (2020). Accelerating evidence-informed decision-making for the Sustainable Development Goals using machine learning. Nature Machine Intelligence, 2(10), 559-565.

Porciello, J., Coggins, S., Otunba-Payne, G., & Mabaya, E. (2021). A systematic scoping review: How are farmers using digital services in low-and middle-income countries? NPJ Digital Medicine, 4(1), 1-14.

Rosenstock, T. S., Nowak, A., & Girvetz, E. (2019). The Climate-Smart Agriculture Papers: Investigating the Business of a Productive, Resilient and Low Emission Future. Springer Nature.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1), 1-9.

Arnaud, E., Laporte, M. A., Kim, S., Aubert, C., Leonelli, S., Miro, B., Cooper, L., Jaiswal, P., Kruseman, G., Shrestha, R., Buttigieg, P. L., Mungall, C. J., Pietragalla, J., Agbona, A., Muliro, J., Detras, J., Hualla, V., Rathore, A., Das, R. R., ... & King, B. (2020). The Ontologies Community of Practice: A CGIAR Initiative for Big Data in Agrifood Systems. Patterns, 1(7), 100105.

Campbell, B. M., Hansen, J., Rioux, J., Stirling, C. M., Twomlow, S., & Wollenberg, E. (2019). Urgent action to combat climate change and its impacts (SDG 13): transforming agriculture and food systems. Current Opinion in Environmental Sustainability, 34, 13-20.

CGIAR. (2022). CGIAR 2030 Research and Innovation Strategy: Transforming food, land, and water systems in a climate crisis. CGIAR System Organization.

CGIAR Platform for Big Data in Agriculture. (2021). Annual Report 2020: Digital Dynamism for Adaptive Food Systems. CGIAR.

Delerce, S., Dorado, H., Grillon, A., Rebolledo, M. C., Prager, S. D., Patiño, V. H., Garcés Varón, G., & Jiménez, D. (2020). Assessing weather-yield relationships in rice at local scale using data mining approaches. PloS One, 15(5), e0233246.

Halewood, M., Lopez Noriega, I., Ellis, D., Roa, C., Rouard, M., & Sackville Hamilton, R. (2018). Using digital sequence information to achieve the crop diversity system management goals. Global Food Security, 18, 9-14.

Janssen, S. J., Porter, C. H., Moore, A. D., Athanasiadis, I. N., Foster, I., Jones, J. W., & Antle, J. M.

(2017). Towards a new generation of agricultural system data, models and knowledge products: Information and communication technology. Agricultural Systems, 155, 200-212.

Jiménez, D., Dorado, H., Cock, J., Prager, S. D., Delerce, S., Grillon, A., Bejarano, M. A., Benavides, H., & Jarvis, A. (2020). From observation to information: Data-driven understanding of on farm yield variation. PloS One, 15(4), e0229881.

Jiménez, D., Delerce, S., Dorado, H., Cock, J., Muñoz, L. A., Agamez, A., & Jarvis, A. (2021). A scalable scheme to implement data-driven agriculture for small-scale farmers. Global Food Security, 30, 100559.

Leonelli, S., Davey, R. P., Arnaud, E., Parry, G., & Bastow, R. (2017). Data management and best practice for plant science. Nature Plants, 3(6), 1-4.

Loboguerrero, A. M., Campbell, B. M., Cooper, P. J., Hansen, J. W., Rosenstock, T., & Wollenberg, E. (2018). Food and earth systems: priorities for climate change adaptation and mitigation for agriculture and food systems. Sustainability, 11(5), 1372.

Porciello, J., Ivanina, M., Islam, M., Einarson, S., & Hirsch, H. (2020). Accelerating evidence-informed decision-making for the Sustainable Development Goals using machine learning. Nature Machine Intelligence, 2(10), 559-565.

Porciello, J., Coggins, S., Otunba-Payne, G., & Mabaya, E. (2021). A systematic scoping review: How are farmers using digital services in low-and middle-income countries? NPJ Digital Medicine, 4(1), 1-14.

Rosenstock, T. S., Nowak, A., & Girvetz, E. (2019). The Climate-Smart Agriculture Papers: Investigating the Business of a Productive, Resilient and Low Emission Future. Springer Nature.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1), 1-9.

References

Arnaud, E., Laporte, M. A., Kim, S., Aubert, C., Leonelli, S., Miro, B., Cooper, L., Jaiswal, P., Kruseman, G., Shrestha, R., Buttigieg, P. L., Mungall, C. J., Pietragalla, J., Agbona, A., Muliro, J., Detras, J., Hualla, V., Rathore, A., Das, R. R., ... & King, B. (2020). The Ontologies Community of Practice: A CGIAR Initiative for Big Data in Agrifood Systems. Patterns, 1(7), 100105.
Arnaud, E., Laporte, M. A., Kim, S., Aubert, C., Leonelli, S., Cooper, L., Jaiswal, P., Kruseman, G., Shrestha, R., Buttigieg, P. L., Mungall, C. J., Pietragalla, J., Agbona, A., Muliro, J., Detras, J., Hualla, V., & King, B. (2021). GARDIAN: The CGIAR's Global Agricultural Research Data Innovation & Acceleration Network. Patterns, 2(1), 100235.
Campbell, B. M., Hansen, J., Rioux, J., Stirling, C. M., Twomlow, S., & Wollenberg, E. (2019). Urgent action to combat climate change and its impacts (SDG 13): transforming agriculture and food systems. Current Opinion in Environmental Sustainability, 34, 13-20.
CGIAR. (2021). CGIAR Performance and Results Report 2020. CGIAR System Organization.
CGIAR. (2022). CGIAR Strategy 2030: Transforming food, land, and water systems in a climate crisis. CGIAR System Organization.
CGIAR Platform for Big Data in Agriculture. (2020). Annual Report 2019. CGIAR.

CGIAR Platform for Big Data in Agriculture. (2021). Digital Innovation in Food Systems: Priorities for the CGIAR. CGIAR.
CGIAR Platform for Big Data in Agriculture. (2022). BIG DATA Analytical Platform: Technical Documentation. CGIAR.
Halewood, M., Lopez Noriega, I., Ellis, D., Roa, C., Rouard, M., & Sackville Hamilton, R. (2018). Using genomic sequence information to increase conservation and sustainable use of crop diversity and benefit-sharing. Biopreservation and Biobanking, 16(5), 368-376.
Janssen, S. J., Porter, C. H., Moore, A. D., Athanasiadis, I. N., Foster, I., Jones, J. W., & Antle, J. M. (2017). Towards a new generation of agricultural system data, models and knowledge products: Information and communication technology. Agricultural Systems, 155, 200-212.
Jiménez, D., Dorado, H., Cock, J., Prager, S. D., Delerce, S., Grillon, A., Bejarano, M. A., Benavides, H., & Jarvis, A. (2020). From observation to information: Data-driven understanding of on farm yield variation. PloS One, 15(4), e0229881.
Jiménez, D., Delerce, S., Dorado, H., Cock, J., Muñoz, L. A., Agamez, A., & Jarvis, A. (2021). A scalable scheme to implement data-driven agriculture for small-scale farmers. Global Food Security, 30, 100559.
Leonelli, S., Davey, R. P., Arnaud, E., Parry, G., & Bastow, R. (2017). Data management and best practice for plant science. Nature Plants, 3(6), 1-4.
Loboguerrero, A. M., Campbell, B. M., Cooper, P. J., Hansen, J. W., Rosenstock, T., & Wollenberg, E. (2018). Food and earth systems: priorities for climate change adaptation and mitigation for agriculture and food systems. Sustainability, 11(5), 1372.
Porciello, J., Ivanina, M., Islam, M., Einarson, S., & Hirsch, H. (2020). Accelerating evidence-informed decision-making for the Sustainable Development Goals using machine learning. Nature Machine Intelligence, 2(10), 559-565.
Porciello, J., Coggins, S., Otunba-Payne, G., & Mabaya, E. (2021). A systematic scoping review: How are farmers using digital services in low-and middle-income countries? NPJ Digital Medicine, 4(1), 1-14.
Rosenstock, T. S., Nowak, A., & Girvetz, E. (2019). The Climate-Smart Agriculture Papers: Investigating the Business of a Productive, Resilient and Low Emission Future. Springer Nature.
Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1), 1-9.