Available at www.sciencedirect.com

INFORMATION PROCESSING IN AGRICULTURE 7 (2020) 355–368

journal homepage: www.elsevier.com/locate/inpa

Paradigm change in Indian agricultural practices using Big Data: Challenges and opportunities from field to plate

Vinay Kellengere Shankarnarayan [a,*], *Hombaliah Ramakrishna* [b]

[a] *Department of Industrial Engineering & Management, Dayananda Sagar College of Engineering, Bengaluru, Karnataka 560078, India*
[b] *Department of Mechanical Engineering, Sapthagiri College of Engineering, Bengaluru, Karnataka 560057, India*

ARTICLE INFO

Article history:
Received 11 August 2019
Received in revised form
8 January 2020
Accepted 13 January 2020
Available online 16 January 2020

Keywords:
Agriculture
Data
Governance
Precision agriculture
Smart farming

ABSTRACT

Agriculture is the backbone of the Indian Economy. However, statistics show that the rural population and arable land per person is declining. This is an ominous development for a country with a population of more than one billion, with over sixty-six percent living in rural areas. This paper aims to review current studies and research in agriculture, employing the recent practice of Big Data analysis, to address various problems in this sector. To execute this review, this article outline a framework for Big Data analytics in agriculture and present ways in which they can be applied to solve problems in the present agricultural domain. Another goal of this review is to gain insight into state-of-the-art Big Data applications in agriculture and to use a structural approach to identify challenges to be addressed in this area. This review of Big Data applications in the agricultural sector has also revealed several collection and analytics tools that may have implications for the power relationships between farmers and large corporations.

Contents

* Corresponding author at: Department of Industrial Engineering & Management, Dayananda Sagar College of Engineering, Kumaraswamy Layout, Bengaluru, Karnataka 560078, India.
E-mail addresses: vinayks-iem@dayanandasagar.edu (V. Kellengere Shankarnarayan), principal@sapthagiri.edu.in (H. Ramakrishna).
Peer review under responsibility of China Agricultural University.
https://doi.org/10.1016/j.inpa.2020.01.001

1. Introduction

Agriculture is facing significant problems, especially in developing countries where people cannot afford modern agricultural technologies. Krishna Kumar, et al. [1] noted that the substantial variability in agricultural production is mainly due to variabilities in weather year-to-year, which is the primary cause of fluctuations in yields. In addition, they shed light on rural poverty and agricultural performance in India. Ahluwalia [2] offered proof that, while agriculture is not the only source of income in rural areas, it is the dominant source. Moreover, according to an Executive Summary published by FAO, the scale of non-agricultural income-generating activity in rural areas almost certainly depends on the level of agricultural production in those areas [3]. The world population is predicted to reach 9.1 billion by 2050, approximately 34 percent higher than the present population. Urbanization will grow at a rapid rate, with about 70% of the population becoming urban and income levels increasing significantly from their current values. Thus, it is essential to balance supply and demand, which play important roles in determining the market prices of commodities.

The concept of Big Data is a relatively new one, so it is expected that knowledge about its applications and their implications for research and development are not widely spread. In reviewing the literature, attention is paid to both technical and economic issues. However, technology is changing rapidly in this domain, and these concepts will probably be outdated soon. Therefore, my analysis has focused primarily on the economic impact that Big Data can have on agricultural management and on the network around it, as this is expected to have a longer-lasting effect. From that perspective, the research questions to be addressed in this review are as follows:

i. What role does Big Data play in agriculture?
ii. What are the opportunities for leveraging existing technological trends?
iii. What changes in agriculture are expected from Big Data developments?
iv. What methodological improvements and specific challenges need to be addressed?

The latter question can be considered as a research agenda for the future.

To answer these questions and to structure the review process, a conceptual framework for analysis has been developed, which is expected to be useful also for future analyses of developments in Big Data in agricultural practices. In the remainder of this paper, the methodology used in reviewing the literature (Section 3) and the conceptual framework will be described (Section 5). The main results from different case studies are summarized in Section 6. Section 9 concludes the review and provides recommendations for further research and action.

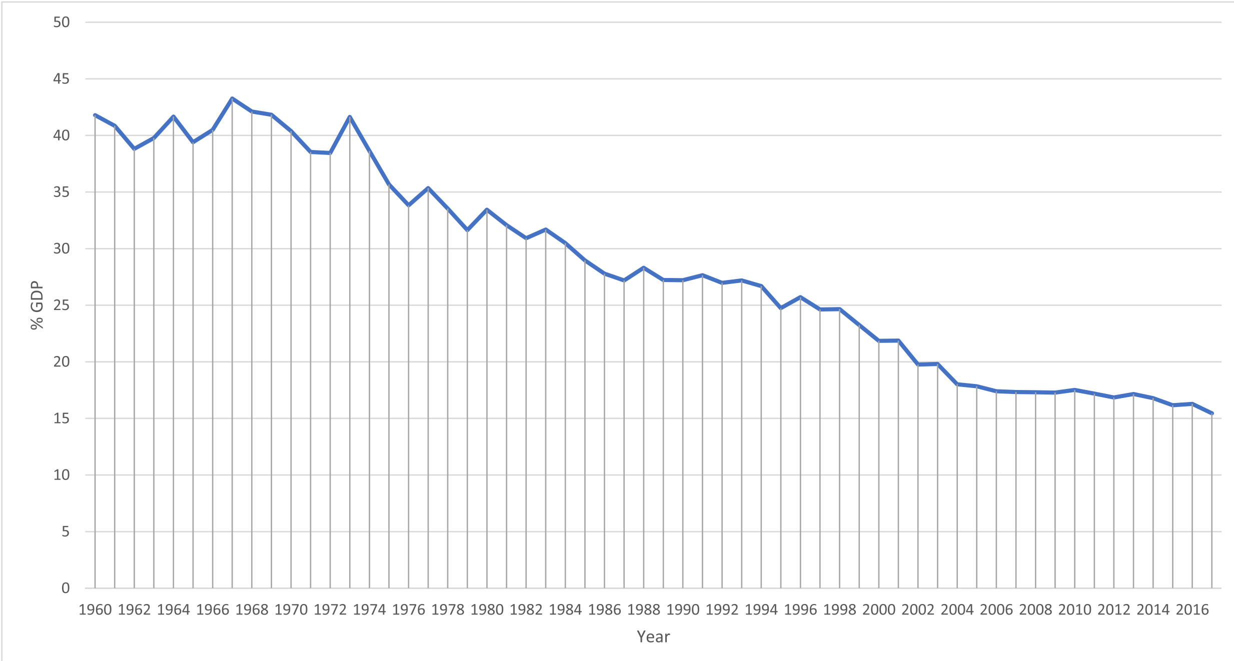

Fig. 1 – Contribution of agriculture toward the national economic growth.

2. The rise of data analytics in agriculture

Agriculture is the major industry for over fifty percent of the country's workforce, and it is more in need of institutional support than any other area. With the largest farm output in the world, Indian agriculture accounts for twenty-three percent of the country's GDP. Given the dramatic slowdown in rural population growth, however, feeding billions of people by 2050 is likely to present significant challenges if the status quo is maintained. As Fig. 1 clearly shows, the contribution that agriculture is making toward the nation's economic growth is declining, primarily due to growth in other sectors of the economy. Nevertheless, this does not affect the importance of the agricultural sector to the Indian economy.

There has been growing interest in large-scale data analytics in recent years. With advances in processing speeds, the availability of cheap cloud storage, and log data, data can be extracted and analysed to obtain accurate results quickly. Similarly, in the agricultural field, to satisfy increasing agricultural-production demands, researchers had started a large-scale data-analytics initiative even before 1999. Basso et al. [4] presented a spatial simulation of agricultural land using advanced technologies such as remote sensing and the Global Positioning System (GPS). Aqeel-Ur-Rehman, et al. [5] investigated the selection of sensors and their effective utilization to solve problems in the agricultural field. Lokers et al. [6] recommended that agro-environmental industries should concentrate on issues of variety and veracity. No two businesses are the same, and there is now an increasing need for location-specific information, which provides solutions tailored to farmers in specific locations. Initially, data management and curation were matters of concern, because researchers were working with small data sets. However, persistent concerns have been raised since the year 2000 about the future of agriculture and food security. These concerns can be addressed by harnessing the potential of Big Data, and researchers have begun to develop models in the agriculture domain [7], including farm models [8], crop models [9], ecological models [10], and land use models [11]. According to a survey conducted by Search Technologies [12], a smart self-driving tractor can collect 30 MB of data per day. This raw data from farm areas can be extracted and analysed using data-driven models referred to as Extract, Transform, and Load which convert the collected data into a standard format. Leon [12] has provided a brief overview of data canonicalization, which deals with character sequences that are case-insensitive in nature, thereby making them easy to integrate into canonicalization algorithms.

3. Review of methodology

The systematic literature review technique is adopted to examine various applications of Big Data in agriculture. The review process followed guidelines proposed by Tranfield et al. [13], which involves three stages:

i. Planning the review
ii. Conducting the review
iii. Findings and dissemination

3.1. *Planning the review*

The primary objective of the present study is to identify the effects of data analytics on agriculture and the challenges presented by digitization in agriculture. In this step, a keyword-based search was performed from the Scopus journal database and conference articles on IEEE Explore, as well

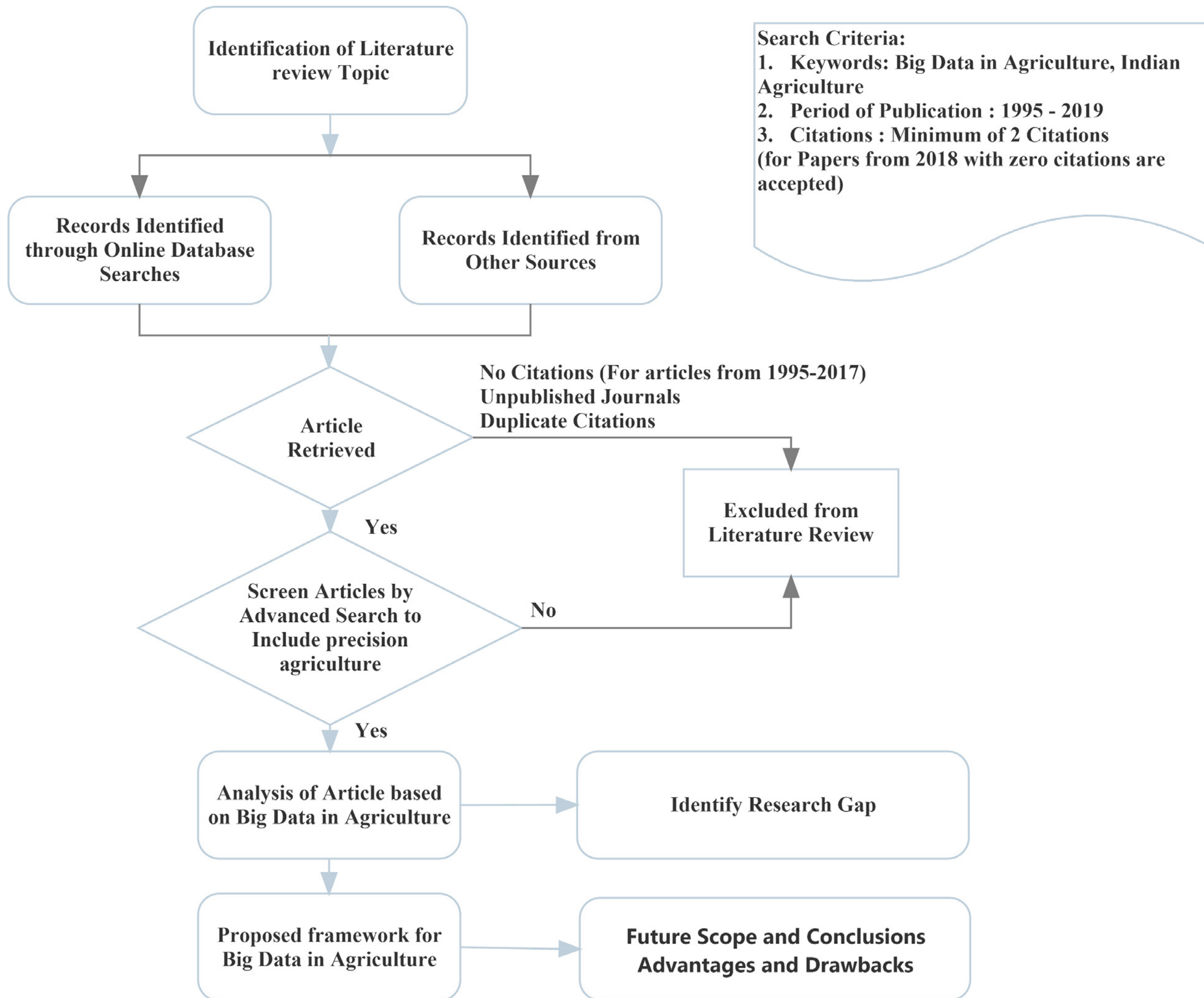

Fig. 2 – Methodology adopted in the review process.

as of scientific indexing databases such as the Web of Science and Google Scholar. The search keywords included "Big Data," "precision agriculture," "smart farming," and "agriculture." Since Big Data is a recent phenomenon, we surveyed research from 1995 to 2018.

3.2. *Conducting the review*

Conducting the review involved the following phases: identification of research, selection of studies, quality assessment, and data synthesis.

3.2.1. *Identification of research*

We selected research papers from the Scopus database, White Papers, and conference proceedings related to the research. We searched for papers referring to Big Data and then filtered out papers referring only to Big Data but which are not applied to agriculture or farming.

3.2.2. *Quality assessment and selection criteria*

In the next step, the selected papers were checked to ensure that they actually applied Big Data to agriculture and used the five V's in agriculture volume, velocity, variety, veracity, and value which are the keys to making Big Data a huge business. I analysed the problems addressed, the results obtained, the tools and techniques used, and the data sources cited and determined which of the five V's were satisfied.

Demchenko, et al. [14] has characterized Big Data into the following five categories:

Volume (1): The agricultural data collected may be structured data or unstructured. The data set required for analysis increases with the growth in population; hence, large amounts of data will be generated. Here, volume deals with the size of data collected for analysis.

Velocity (2): Not all the data collected may be useful. Data that is relevant for analysis must be recorded. For example, Kalil [15] described how President Obama's administration announced the Big Data Research and Development initiative, which opened the gates for addressing different problems faced by the government.

Variety (3): Data does not come from a single source and the agricultural field is no exception. Big agricultural players such as John Deere spend heavily on Big Data to analyse soil type, seed variety, and changes in weather [16]. This use of Big

Table 1 – Big data projects for different development areas in agriculture.

Project	Reference	Type	Open Data	Dimension					Description
				V1	V2	V3	V4	V5	
FarmERP	[21]	*Web/Mobile Application*	No	*L*	*L*	*H*	*L*	*H*	*A PC / LAN based multi user agriculture pathology services, advisory services management software suite for Agriculture Pathology Laboratories, Farm Advisory Centers and companies.*
Gold Farm	[22]	*Online Service*	No	*M*	*L*	*M*	*L*	*M*	*Helps farmer to Hire or Purchase Farm Equipment's and in addition to that Gold Farm built a Complete IoT based application which can track the heath of Farm Equipment in Real Time*
Gigaom	[23]	*Online Analysis*	No	*M*	*H*	*H*	*M*	*M*	*Used to implement Data Analytics in Agriculture. This Includes Benchmarking, Sensor Deployment and Uses better model to analyze crop failure risk.*
Farmers Portal	[24]	*Government Web Service*	Yes	*H*	*L*	*L*	*L*	*H*	*Through this portal a Farmer can get all relevant information related to his Village/Block/City/State though Text Message/Mail. Also available in 15 Different Languages.*
Kheyti	[25]	*Non-Profit*	Yes	*L*	*M*	*H*	*M*	*H*	*This team design, adopt, and implement low-cost farming solutions that help farmers to increase their yield and increase profits.*
Unimart	[26]	*Advisory/Training*	No	*H*	*M*	*H*	*M*	*M*	*Caters Agri-Input Products and services, Soil Testing, Water Conservation, Weather Information, Agronomic Advises, Smart Agri Technologies.*
JFarmservices	[27]	*Mobile Application Advisory Services*	Yes	*L*	*L*	*M*	*L*	*L*	*Free Access to reliable weather Information, Crop Prices, Rental Platform, Latest Agriculture news, Agriculture advisory.*
Department of Agriculture Cooperation and Farmer Welfare	[28]	*Government Web Service,*	Yes	*M*	*M*	*M*	*M*	*M*	*Agriculture Informatics Division of Government of India. Offers nearly 36 + exclusive Schemes and services for farmers, from Agriculture Market Platform to Digital Agriculture*
Indian Agriculture University	[29]	*Government Web Service*	Yes	*M*	*M*	*L*	*L*	*M*	*Has Vast Data Base of Top Agriculture Universities in India and Around the world, Latest Researches in Agriculture, Books, Journals, Websites all related to Agriculture in its Vault.*
Agriculture and Processed food products Export development authority	[30]	*Government Web/Mobile Application*	Yes	*L*	*M*	*L*	*L*	*L*	*Facilitates Export of Food Products, conducts surveys and Feasibility studies, helps in online farm registration, Processing and Approval of Farmer applications and Sample Collection for testing.*
Janmanrega	[31]	*Android Mobile Application*	Yes	*L*	*L*	*L*	*L*	*L*	*A citizen centric mobile app provides an interface to improve quality of public services under Mahatma Gandhi National Rural Employment Guarantee Act Scheme.*
Indian Society for Agri Business Professionals	[32]	*Non-Profit*	No	*M*	*L*	*L*	*M*	*L*	*Empower Farm Sector through Sustainable Farming and Integrated Farming Techniques*

(Continued on next page)

Table 1 – (continued)

Project	Reference	Type	Open Data	Dimension V1	V2	V3	V4	V5	Description
India.gov.in Agriculture Portal	[33]	*Government Web based service*	*yes*	*M*	*M*	*L*	*L*	*L*	*Has a vast data base related to Agriculture Produces, Machineries, Research, Also portal facilitates very detailed information on Various government policies, Schemes, Market prices, Animal Husbandry, Fisheries, Horticulture, Sericulture, Loans, Credit, etc.,*
Reuters Market Light (RML)	[34]	*Web/Mobile Portal*	*NO*	*H*	*H*	*M*	*H*	*M*	*Award winning Private Business firm provides various analytical services. communicates unbiased and highly personalized Agri information which covers every stage of the crop cycle*
Government of India, Department of Fertilizers	[35]	*Government Online Portal*	*Yes*	*M*	*M*	*L*	*L*	*L*	*Access Fertilizer Monitoring System which is designed and developed by Department of Fertilizers Government of India. It provides comprehensive on fertilizers in a dynamic timeframe.*
Agriquest	[36]	*Non-Profit*	*Yes*	*H*	*M*	*H*	*M*	*H*	*Provides information on Seed Marketing Companies, Agriculture subsidiary companies, Food Processing Industries, Agri Ads which helps farmers. Along with this it also maintains a database of Agriculture Periodicals and Journals.*
AgVue	[37]	*Online/Offline Services*	*No*	*H*	*H*	*H*	*H*	*M*	*Helps in Detailed analysis of Crop, Uses Geographical Imaging System (GIS) for analysis by collecting Satellite Images of Field Crops. Analysis include vegetative indices, Biomass Analysis, Crop Counting and Digital Surface mapping*
Precision AG	[38]	*Software Application*	*No*	*M*	*M*	*H*	*H*	*H*	*It is a Independent Enterprise serving Globally using Precision Agriculture Techniques like Adjusting Production Inputs based on Variability, and also uses GIS and GPS systems for analysis.*
Ag-Analytics	[39]	*Software Application*	*No*	*M*	*M*	*H*	*H*	*H*	*Has its own API, Use Predictive Analytics platform to collect data and analyze Soil, Crop Data Layer Service, CLU Boundaries Service, Federal Crop Insurance service etc.*
Tools and Services for Data Analysis									
Rapid Miner	[40]	*Software Tool*	*Not Applicable*	*H*	*M*	*H*	*H*	*L*	*An Opensource Predictive Analysis platform which is used to evaluate data health, Completeness and Quality in Real Time.*
Open Data Kit	[41]	*Software Tool*	*Not Applicable*	*M*	*M*	*H*	*H*	*L*	*Free and Opensource Software tool to collect, Manage and using data resource in constrained environments.*
Weka 3	[42]	*Software Tool*	*Not Applicable*	*M*	*M*	*H*	*H*	*L*	*Weka is a Free and Opensource collection of machine learning algorithms, Includes tools for data preparation, classification, regression, clustering, and visualization.*
Amazon EMR	[43]	*Online Data Processing Tool*	*Not Applicable*	*H*	*H*	*M*	*H*	*H*	*Data processing for large instances.*

(Continued on next page)

Table 1 – (continued)

Project	Reference	Type	Open Data	Dimension					Description
				V1	V2	V3	V4	V5	
Open Source and Paid Online Data Base for Analysis									
Data.GOV	[44]	*Government Service*	*Yes*	*H*	*H*	*M*	*M*	*H*	*Collection of Open Data can be used for Research and Analysis*
Statista	[45]	*Website*	*No*	*H*	*H*	*H*	*H*	*H*	*Access to over more than one million Data Sets, Statistics and Facts*
The World Bank	[46]	*Non-Profit*	*Yes*	*H*	*H*	*M*	*H*	*M*	*Maintains millions of Database Analysis which contains collection of time series data on a wide variety of topics.*
Open Government Data Platform India	[47]	*Government*	*Yes*	*M*	*M*	*M*	*M*	*M*	*Stored around 2 Million data resources, 10 K API's and over 1000 Data Visualizations*
International Data Corporation	[48]	Advisory	No	M	H	H	M	M	Provides Advisory services and Market Intelligence for Information technology, telecommunications and Consumer Markets.

Data not only helps farmers in reducing capital, but also significantly increases crop yield.

Veracity (4): Veracity reflects uncertainties in the data and addresses the integrity and accuracy of the data. Ali et al. [17] highlighted the challenges and issues encountered in establishing the veracity of data sets and in ensuring encryption and privacy in Big Data employed for development.

Value (5): Collected data is useless unless it is processed to give it value. Jin et al. [18] described the grand challenges associated with value of Big Data, namely data complexity, computational complexity, and system complexity.

3.3. *Findings and dissemination*

The results of the review will be presented and illustrated in a descriptive form. Based on these results, I then propose a framework for Big Data in agriculture and discuss the future scope of this domain. Fig. 2 illustrates the methodology I adopted for the review process.

4. Big data for developmental areas

Kshetri [19] presented the factors necessary to implement Big Data techniques for agricultural development. In this paper, I review different sources available for the development of farming, as illustrated in Table 1. Some of the problems encountered by farmers in the areas of procurement, natural disaster, soil condition, supply chain, biotech, and financial data management will be addressed. In addition, security and encryption play major roles, as farmers are appropriately concerned about the misuse of information generated from their farming activities. Jacob Bunge [20] noted that big multinational agricultural firms like Monsanto are promoting and influencing farmers to buy their products.

5. Conceptual framework

Porter [48] first introduced the value-chain concept. The idea is that a firm creates value in a product by acquiring raw materials and using them to produce something useful. The more value an organization creates, the more profitable it is likely to become. Moreover, when a firm provides more value to its customers, it builds a competitive advantage. Similarly, Bhadani and Jothimani [49] argued that the data-value chain refers to a framework that deals with a set of activities to create value from an available data set. As shown in Fig. 3, I have divided this framework into four phases: data acquisition, information extraction, knowledge management, and business insights.

5.1. *Data acquisition*

Data collection is the primary step involved in data analytics. It is designed to capture a wide range of structured and unstructured data such as weather reports, soil conditions, images captured from satellites, and a variety of other sources. It is thus necessary to program filters to capture only the required data sets and to nullify unnecessary data [50]. In addition, it is necessary to generate suitable metadata for every data set, to provide a clear way to further analyse and

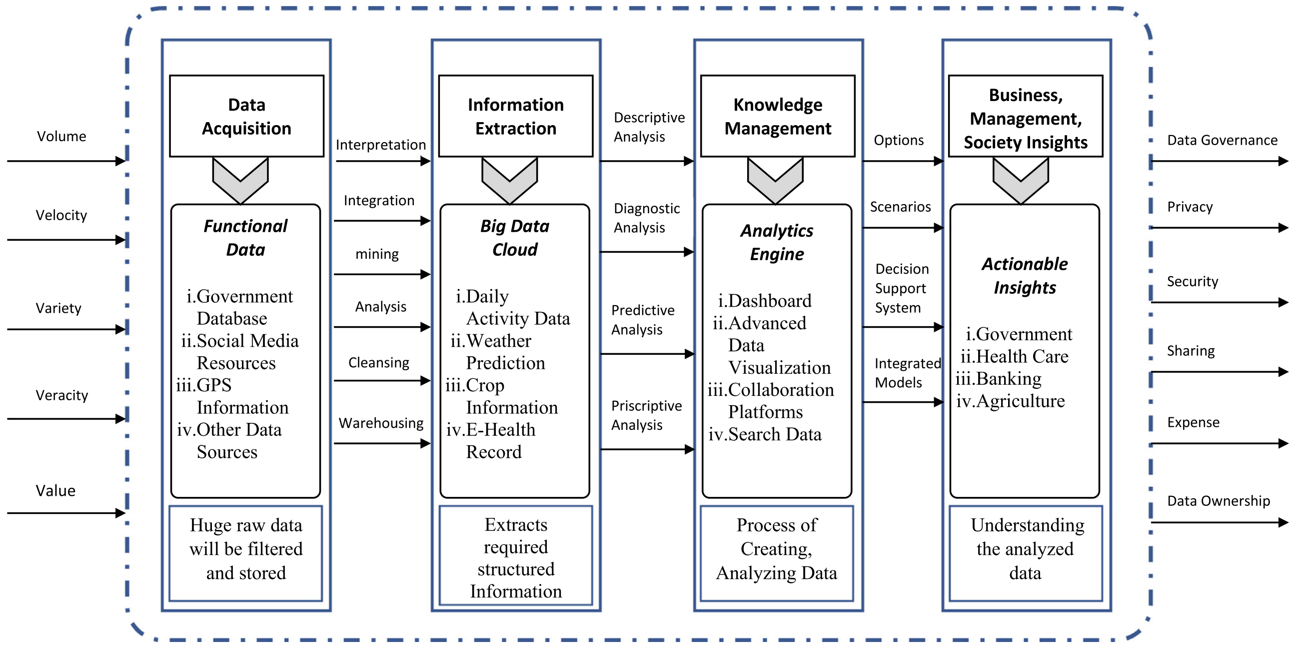

Fig. 3 – Conceptual framework for Big Data analytics in agriculture.

render the data [51]. For example, weather report metadata for the year 2019 can be generated with "date," "time," "filename," and "geolocation."

5.2. *Information extraction*

The data acquired is usually not ready for analysis. The information-extraction phase pulls out only the required information for analysis. Prathap [52] demonstrated how information extraction is performed in different cases. Information is extracted by interpreting, integrating, mining, analysing, cleansing, and warehousing the data.

5.3. *Knowledge management*

Data analytics are performed to improve decision-making. Russom [53] pointed out that Big Data analytics involves using advanced analytic techniques like Map Reduce and Online Analytics Processing (OLAP) to operate on Big Data. Big Data provides huge statistical samples, which enhance the results of the analysis. In other words, the larger the data sample, the more accurate the analysis. The analysis may involve descriptive, diagnostic, predictive, or prescriptive analytics.

5.3.1. *Descriptive analytics*

Descriptive analytics deals with answering the question, "what happened?" For instance, Laude [54] provided a better understanding of how agricultural practices and policies are relevant to data science, to clarify the main features of an automated system based on prescriptive analytics for agriculture. Descriptive analytics indicate whether the result is right or wrong.

5.3.2. *Diagnostic analytics*

Diagnostic analysis answers the question, "what is the reason?" For example, Makin et al. [55] explained the reasons for the presence of pesticides in the atmosphere across Canadian agricultural regions. This type of analytics involves measuring the risk of a dataset.

5.3.3. *Predictive analytics*

Predictive analytics predicts "what is going to happen." Cleary et al. [56] reported preliminary findings of an outbreak of foodborne Salmonella Bareilly in the United Kingdom, clarifying how the consumption of bean sprouts was significantly associated with the disease.

5.3.4. *Prescriptive analysis*

Prescriptive analysis prescribes actions that must be taken to avoid future problems. Krishna Kumar et al. [1] showed the influence of monsoon rainfall on crop production and demonstrated the correlation between monsoon and post-monsoon rainfall, explaining how seeds must be planted according to rainfall patterns.

5.4. *Business insights*

This is the final phase, where the raw data is turned into actionable insights. Here, the data captured from sensors on the farm is used to improve the management of crop-failure risk and boost feed efficiency in real time. Big Data is used to provide predictive insights for future outcomes in farming [57].

6. Opportunities for future Farming: Big data

There is a saying that "data is the new oil, but we need resources to refine it," similar to the manner in which crude oil is extracted, refined, and directed to destinations where it is needed. The same is true in the case of agriculture. The agricultural field has been quite slow in adopting new technologies, especially those that require a larger scale of operations and where the initial investment costs required are high. The implementation of new technologies has a significant impact on the prospects of farmers. For example, Virtus Nutrition and Dairy.com jointly introduced "my dairy dashboard" [58], a cloud-based platform for data analysis, supply-chain monitoring, and agricultural production performance optimization tool for dairy farmers. These advanced technologies help dairies stay focused on what really matters, and they save consultants' time, enabling them to focus on analysis and to help dairies take actions to improve their bottom lines. With visualisation and analysis tools, more producers, farmers, and consultants will be able to connect and simplify their data management at a lower price point [59]. This shift toward new technologies is made possible by research and development in both hardware and software services. In recent years, agricultural innovation has resulted in promising new ways to boost productivity. For farmers in rural India, gaining access to quality products, accurate information, and fair prices often presents a challenge. Jaruzelski et al. [60] noted that it is essential to have increased investment in agricultural research and development to feed the world's growing population.

6.1. *Case studies*

Some case studies demonstrate how data analytics is revolutionizing the agricultural sector.

Case 1. Data Analytics for Precision Agriculture:

Satellite-derived management zones (SAMZs) are one of several products of continuing efforts to realize the concept of precision agriculture [61]. SAMZ denotes a methodology and software for implementing the methodology to create management zones. It helps in deriving management zones from images of fields acquired by instruments aboard Earth-orbiting satellites. In this approach, the main source of data is archived satellite imagery. The advantage of this approach is that the archived data can be reused for precision agriculture at low cost. The archives contain information on all sorts of variability within a field, including weather, crop types, crop management, soil types, and water drainage patterns. It is also possible to obtain periodic observations of germination percentage, plant emergence, physiological maturity, pest and disease incidence, and yield. The SAMZ methodology involves establishing a web-based interface and using an algorithm that automatically and rapidly generates management zones from archived satellite image data in response to requests from farmers. From this research, a farmer can make a request either by uploading data describing a field boundary to the website or by drawing the boundary on a reference image. Hence, a farmer can engage in precision farming shortly after gaining access to the website, without incurring the high costs of conventional precision-agriculture data-collection practices such as collecting soil samples, mapping the electrical conductivity of the soil, and compiling multiyear crop-yield data.

Recent advances in remote sensing technology have made the SAMZ approach a handy tool for precision agriculture. To analyse the accuracy of this tool, we chose agricultural area belonging to Yadgir District, Karnataka, India, as the study area. In this study, we used free cloud ISRO IRS P6 LISS-III data from 28 October 2009 and NDVI data sets from January to May and from November and December 2011 from the Bhuvan online portal of the National Remote Sensing Centre (NRSC), ISRO, Hyderabad. Delineation of management zones is one of the prerequisites for precision agriculture. The authors used Management Zone Analyst software to divide the agricultural area into six management zones to develop a decision-support system for data analytics. Management Zone Analyst calculates descriptive statistics, performs the unsupervised fuzzy classification procedure for a range of cluster numbers, and provides the user with two performance indices.

Case 2. Data Analytics for Innovation and Yield:

To study how farmers can utilize data effectively, John Deere developed an online portal [62] that aggregates data from soil probes and sensors outfitted on John Deere agricultural equipment to gather information as the equipment and machinery move about the agricultural field. The firm also included data from outside sources in the portal, including nearby farms, thereby offering insights into productivity under a wide range of conditions. The portal also provides vital information that can help farmers optimize yields and planning and to better manage available resources. The final payoff can be tremendous. According to a study published in Nature, the combined actions of curbing nutrient overuse and employing sustainable intensification could increase the production of maize, wheat, and rice by 30% [63].

Case 3. Big Data for Soil Health Management:

The government of India introduced the Multipurpose National ID Card in 2015 [64], intending to give every farmer a health card for their soils in order to inform them of the status of nutrients in the soil and consequently to guide them in fertilizer applications to maximize yield. The entire governmental agricultural extension and research system galvanized itself, collected samples, analysed them for 12 soil chemicals, and recommended fertilizer dosages which they printed on Soil Health Cards (SHCs) that were distributed to farmers. The scheme delivered on the basic promise, as currently 25.3 million samples have been collected, and SHCs have been distributed to 107.4 million farmers.

However, this scheme has three major drawbacks:

First, operational challenges plague the system. Soil samples are collected from every 2-square-hectare parcel of land in irrigated areas and are then transported together for analysis in a dated network of wet-chemistry labs. This puts a tremendous strain on the system and impacts the quality of soil analysis. Subsequent studies have shown a low correlation between the results generated by the SHC scheme and those generated by gold-standard labs.

Second, the current design of the scheme oversimplifies nutrient recommendations. Increasingly, research is showing that a crop's "yield response" to a nutrient is very complex. It is determined not only by deficiencies of nutrients, but also by other variables such as rainfall, production practices, the presence of other nutrients, soil acidity, and temperature, to mention but a few. The correct yield response can be predicted only from a model that uses accurate data for the above parameters, a system which the current scheme does not employ.

Finally, the scheme underestimates its own potential, because the large-scale collection of soil data finds little use outside of filling out a physical card. This vast repository of data, painstakingly aggregated from millions of samples, remains largely isolated from researchers, start-ups, and even state governments. Access to this information could help start-ups to combine soil-health-card data with rainfall and irrigation data to deliver precision irrigation advisories to farmers on their mobile phones.

These shortcomings, however, present a remarkable opportunity for Indian agriculture. The system could move to a sampling-based soil-information system that reduces the need for tens of millions of samples which strains the available lab capacity and produces better results four times faster the current rate at half the cost. A recommendation has also been made to encourage the use of a custom fertilizer blend in addition to asking the farmer to reduce sowing depth. Fertilizer companies, building upon such a platform, could leverage soil data, weather data, and farmer-demand patterns to shape the distribution of fertilizer blends in different districts. Such a platform could catalyse a wave of innovations in agriculture.

6.2. *Key factors*

The PwC network [65] has listed the key factors that will drive future agribusiness markets, as presented in Table 2. The three driving factors include: increasing yields, productivity improvements, and sustainability enhancement. It is believed that by adopting all three driving factors, farmers may be able to expand their acreage successfully and improve their agricultural businesses.

6.2.1. *Increasing yields*

The advent of technology gives farmers control over spatial variables. Stafford [66] stated that even though this concept is “technology-driven,” further technology development is essential in this domain in order to provide spatial data. Technologies such as GIS play a vital role in managing natural resources, environmental protection, and in regional and urban planning [67].

6.2.2. *Productivity improvement*

Agriculture 4.0 or the “industrialization of agriculture” is not a new concept, but it has been expanded heavily over the last few decades. Verdouw et al. [68] suggested that farms not only must be efficient enough to produce quality food but also must take into account high-quality environmental complexities and be adaptable to changing market demands. Sørensen et al. [69] argued that current agricultural information systems do not meet the current requirements adequately. The current transformation of the agricultural sector, and the need for better analysis and transformation of the collected data, place additional demands on the precision and integration of the planning and control functions. In order to analyse complex and soft systems, it will be necessary to develop an effective Farm Management Information System that effectively meets farmers’ changing needs. The boundaries and scope of such a Farm Management Information System can be described in terms of entities, interfacing with managers, software, databases and functionalities. Davenport [70] believes that a software package is essential for compiling different functions and integrating them into a single system.

6.2.3. *Sustainability enhancement*

From the methodological point of view, sustainability-enhancement tools can provide a decision-making framework for both farmers and policymakers. With the continued depletion of non-renewable resources and soil degradation, several reform movements have evolved around the world in response to concerns for agriculture and food safety. With the introduction of new agricultural tools [71], there has been a tremendous increase in inputs, such as seeds with high yields per acre, chemicals, fertilizers, pesticides, and insecticides [72]. These initiatives support farmers who are willing to adopt new technology.

7. Challenges and barriers

More than 90% of the data that is currently available was created in the last 5 years. However, the term “Big Data” has been in use since 2005. Smart systems and data science can modernize and produce growth in the agricultural sector. This is changing agricultural practices, from cultivating to warehousing of products. Big Data is still not a popular term in the agricultural sector, mainly due to challenges surrounding its implementation. Some of the challenges include data governance, privacy, security, sharing, expense, and data ownership, as highlighted above.

7.1. *Data governance*

An analysis conducted by Perritt [73] shows that it is crucial to remove technological barriers that hinder the communication and transfer of data sets. They argue that even though Big Data can produce a huge payback, it is faced with many challenges from the five V’s. Capturing accurate and actionable data requires a methodical skillset for handling

Table 2 – The Three driving factors of Agribusiness.

Increasing yields	Productivity improvement	Sustainability enhancement
Precision Agriculture Solutions	Agro-Enterprise Resource Planning	Land Use Mapping
Genomics	Telematics	Total Productive Management
Agriculture 4.0	Autonomous Equipment	Waste Management
Data Analytics	Drones	Energy Management
GIS Image Management	Predictive Analytics	Performance Analysis
Access to seasonal Trends		Harvest Tracking

the data-mining system [19]. Based on their research in India, Sekhar et al. [74] concluded that people depend on cultivation because of illiteracy, which limits them to the current methods of cultivation.

7.2. Privacy

Moorthy et al. [75] discussed privacy issues in Big Data. They believe that it is essential to develop new ways of handling privacy issues. They noted that consumer privacy, collection of personal data, and its utilization are the areas in which marketers are interested. Carbonell [76] raised concerns regarding privacy in data analytics, noting that agricultural equipment manufacturers like DuPont, John Deere, and Monsanto are clearly heading toward becoming information brokers, and they pointed out how the abuse of patents is evolving toward an unparalleled collection of farming data.

7.3. Security

Privacy and the security of data go hand in hand. From a public perspective, food security is viewed as the main driver for any further technological advancements [77–79]. Tong et al. [80] stated that besides the farmers' perspective, consumers are now even more concerned about the nutritional aspects of food, safety, health, and well-being. Wolfert et al. [81] termed precision agriculture the "Holy Grail" that will address the imbalance in food supply and demand.

7.4. Sharing

Data analytics serves as an integral tool for managing and sharing potentially sensitive data in a secure manner [82]. The authors above noted that it is necessary to consider the implications of technological changes and the confidentiality of the data while developing a platform.

7.5. Expense

Does Big Data require big money? Though Big Data is a trending technology in India, it is still costly. To be more precise, storing or transferring 1 GB of data in the cloud is expensive either on the Amazon AWS or Google Cloud suite. In addition, offline computation software packages and hardware require additional investments.

7.6. Data ownership

In any organization, centralization and standardization fail mainly due to the lack of power and freedom [83]. The lack of autonomy of an organization affects its cost and benefits structures, including the role of data ownership. To overcome this limitation, Pruscino & Shores [84] developed a technology with which one can assign and change the ownership for a particular set of data in a database.

In addition to the above challenges, sentimental feelings play a role in adopting modern technology. In India especially, most farmers change their farming practices based on recommendations from friends or relatives. Budget also matters in adopting change; as Big Data or smart technology is still at the inception stage in India, the cost of adopting it will be very high. Finally, Indians are very sentimental toward their ancestral properties, and most of them would not replace or change the farming devices.

8. Policy and legal implications

Big Data has the potential to expand, and it promises better evaluation of farm-level decisions. It can also expand the ability to evaluate the effects of policy interventions in the agricultural economy. In the long term, the growth of Big Data may give rise to new models for the evaluation of policy shocks to economic systems. However, in the near term, the availability of larger and potentially more robust datasets may increase the accuracy of existing model outputs. No business remains the same, and there is now an increasing need for information generated in a location-specific manner to provide solutions that fit what each farmer needs. Through Big Data and connected devices, every one of the goals concerning profitability, efficiency, and cost management not only is achievable but also is completely realistic.

Exploring the ethical implications of Big Data in food and agriculture builds on critical data studies, which have historically focused on the effects of technologies on farmers and food systems. A survey by the American Farm Bureau in 2014 [76] found that farmers feared government officials gaining access to their private information and using it for market speculation. Gilpin [85] noted that farmers are increasingly required to reveal their most personal farm details in order to gain access to the benefits of technology, while those who turn the data into useful information reveal little or nothing about the back-end processes or about how or where the information will be kept or used. Despite alarming trends in and potential issues with data-driven farming, this technology is not biased negatively, and it could be put to ground-breaking use by farmers.

9. Conclusions and the future of agriculture

In this study used, we have performed a systematic literature review to investigate the current state of research in the area of Big Data analytics in the agricultural and farming domain. We find that in adopting an emerging technology like Big Data, there are common issues that every industry must address in order to realize the benefits of digital transformation. For the agricultural sector, it is valuable to learn from other sectors that are further along in the adoption of Big Data analytics. A heavy investment in erudite agricultural equipment does not address the food-scarcity issue, and this is where the Big Data concept is required. At present, it is essential for every farmer in India to adopt new and modern tools and techniques to balance food demand and supply concerns. Current environmental conditions are not favourable, making it harder for farmers to predict rainfall, soil conditions, and even the groundwater level. My systematic literature review indicates that the agricultural sector can benefit significantly from increased access to Big Data.

Our study also showed that there is unprecedented opportunity for access to huge volumes of satellite data. If this technology is made easily available and user-friendly, it can support millions of farmers across India. In farming and agriculture today, outside of more traditional scenarios, a farmer is often beholden to a particular supplier or partner. For example, they may find it necessary to send a certain amount of their most recent harvest to a local grocer or department chain. Regardless of the nature of such agricultural partnerships, it is not always possible to know precisely how much of and when a particular crop is going to be ready. This, coupled with the changing demands on the consumer side can lead to severe supply issues.

Many of the technological challenges and barriers described in this paper must be addressed in order to exploit the full potential of Big Data. There is still a need for dedicated research in the field of Big Data, especially in agriculture, and unrestricted access to data sources is necessary to inspire more researchers in this domain to practice, innovate, and create other platforms to provide insights for farmers across the globe. Moreover, robots can play an important role in control, and increasing machine assistance in analysis and planning may create a virtually autonomous cyber-physical cycle. Concerns regarding ownership of data generated from farms, access to it and intended uses were also addressed. Finally, companies must find a balance between data that is confidential to farms and data that can be used to enhance their work. These factors will shape both Big Data and agricultural space in the years to come.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] Krishna Kumar K, Rupa Kumar K, Ashrit RG, Deshpande NR, Hansen JW. Climate impacts on indian agriculture. Int J Climatol 2004;24(11):1375–93.

[2] Ahluwalia MS. Rural poverty and agricultural performance in India. J Dev Stud 1978;14(3):298–323.

[3] Bellù LG. The future of food and agriculture: alternative pathways to 2050, 2nd ed. The Food and Agriculture Organization (FAO). Link: http://www.fao.org/3/CA1553EN/ca1553en.pdf. 2018 (Published)/2019 [Retrieved].

[4] Basso B, Ritchie JT, Pierce FJ, Braga RP, Jones JW. Spatial validation of crop models for precision agriculture. Agric Syst 2001;68(2):97–112.

[5] Aqeel-ur-Rehman, Abbasi AZ, Islam N, Shaikh ZA. A review of wireless sensors and networks' applications in agriculture. Computer Standards and Interfaces, vol. 36, no. 2; 2014. p. 263–70

[6] Lokers R, Knapen R, Janssen S, van Randen Y, Jansen J. Analysis of big data technologies for use in agro-environmental science. Environ Modell Software 2016;84:494–504.

[7] Agboola, J. Big Data is the future. But where are the farmers? CGIAR platform for big data in agriculture. CGIAR-platform for big data in agriculture; 2018. https://bigdata.cgiar.org/big-data-is-the-future-but-where-are-the-farmers.

[8] Janssen S, Van Ittersum MK. Assessing farm innovations and responses to policies: a review of bio-economic farm models. Agric Syst 2007;94(3):622–36.

[9] Van Ittersum MK, Ewert F, Heckelei T, Wery J, Alkan Olsson J, et al. Integrated assessment of agricultural systems–a component-based framework for the European Union (SEAMLESS). Agric Syst 2008;96(1–3):150–65.

[10] Schmolke A, Thorbek P, Deangelis DL, Grimm V. Ecological models supporting environmental decision making: a strategy for the future. Trends Ecol Evol 2010;25(8):479–86.

[11] Verburg PH, Schot PP, Dijst MJ, Veldkamp A. Land use change modelling: current practice and research priorities. GeoJ: Spatially Integrated Soc Sci Hum 2004(1):309–24.

[12] Leon R. (search. technologies). Precision agriculture: boosting productivity with big data. Search technologies; 2019. https://www.searchtechnologies.com/blog/precision-agriculture-big-data.

[13] Tranfield D, Denyer D, Smart P. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. Br J Manag 2003;14(3):207–22.

[14] Demchenko Y, Ngo C, Membrey P. Architecture framework and components for the big data ecosystem. Syst Netw Eng 2013;1:1–19.

[15] Kalil, T. Big data is a big deal. The White House; 2012. https://obamawhitehouse.archives.gov/blog/2012/03/29/big-data-big-deal.

[16] Faulkner A, Cebul K. Agriculture gets smart: the rise of data and robotics. Cleantech Group; 2014. https://www.cleantech.com/wp-content/uploads/2014/07/Agriculture-Gets-Smart-Report.pdf (Published)/2019 [Retrieved].

[17] Ali A, Qadir J, Rasool R ur, Sathiaseelan A, Zwitter A, Crowcroft J. Big data for development: applications and techniques. Big Data Anal 2016;1(1):2.

[18] Jin X, Wah BW, Cheng X, Wang Y. Significance and challenges of big data research. Big Data Res 2015;2(2):59–64.

[19] Kshetri N. The emerging role of Big Data in key development issues: opportunities, challenges, and concerns. Big Data Soc 2014;1(2). 205395171456422.

[20] Jacob Bunge. Big data comes to the farm, sowing mistrust. Wall Street J; 2014. https://www.wsj.com/articles/no-headline-available-1393372266.

[21] Smart Farm Management Software. Agriculture Software; 2019. https://www.farmerp.com/.

[22] Gold Farm. India's emerging agri platform; 2019. http://www.goldfarm.in/home/.

[23] GigaOm. Cloud data platform for research; 2019. https://gigaom.com.

[24] Farmer portal: one stop shop for farmers; 2019. https://farmer.gov.in/.

[25] Kheyti–Small Farmer. Smart farmer; 2019. https://kheyti.com/.

[26] Unimart – Farmer Advisory. Training & retail; 2019. https://in.uplonline.com/services/unimart.

[27] JFS. JFarm Services. Farmer-to-farmer tractor rental. https://www.jfarmservices.in/.

[28] Department of Agriculture Cooperation & Farmers Welfare. Mo A&FW . Government of India. http://agricoop.nic.in/.

[29] Indian Council of Agricultural Research. Ministry of Agriculture and Farmers Welfare. https://icar.org.in/.

[30] Agriculture and Processed food products Export development authority. Government of India. http://apeda.gov.in/apedawebsite/.

[31] MGNREGA (Mahatma Gandhi National Rural Employment Gurantee Act). Ministry of Rural Development. GoI. https://nrega.nic.in/netnrega/Janmanrega.htm.

[32] Indian Society of Agribusiness Professionals (ISAP). Technology business Incubator. https://www.isapindia.org/.
[33] Agriculture. National portal of India. https://www.india.gov.in/topics/agriculture.
[34] RML Agtech. Formerly reuters market light. https://rmlagtech.com/.
[35] Government of India. Department of fertilizers. Ministry of Chemicals and Fertilizers. http://fert.nic.in/.
[36] Agriquest. Seed certification and development. http://www.agriquest.info/.
[37] Agriculture Drone Services and Crop Analytics. AgVue Technologies. http://www.agvue.tech/.
[38] PrecisionAg. Advancing global digital agriculture. https://www.precisionag.com/.
[39] Ag-Analytics. Farm management platform. https://ag-analytics.org/.
[40] Lightning Fast Data Science Platform for Teams. RapidMiner©. https://rapidminer.com/.
[41] Open Data Kit. A free and open source data collection. https://opendatakit.org/.
[42] Weka 3 – Data Mining with open source machine learning software in java. https://www.cs.waikato.ac.nz/~ml/weka/.
[43] Amazon EMR – Amazon Web Services. https://aws.amazon.com/emr/.
[44] Data.gov. The home of the U.S. Government's open data. Link: https://www.data.gov/.
[45] Statista – The statistics portal for market data, market research and market studies. https://www.statista.com/.
[46] World Bank Open Data. Open Access to global development data. https://data.worldbank.org/.
[47] data.gov.in. Open Government Data [OGD] platform India. https://data.gov.in/.
[48] Porter ME. Competitive strategy: techniques for analyzing industries and competitors. The free press. New York: Simon & Schuster; 1997.
[49] Bhadani A, Jothimani D. Big data: challenges, opportunities and realities. Effective big data management and opportunities for implementation; 2017. p. 1–24.
[50] Kacfah Emani C, Cullot N, Nicolle C. Understandable big data: a survey. Comput Sci Rev 2015;17:70–81.
[51] Agrawal D, Bernstein P, Bertino E, Davidson S, Dayal U, Franklin M, et al. Challenges and opportunities with big data – a community white paper developed by leading researchers across the United States. Computer Research Association; 2012.
[52] Prathap kumar A, Sasank V, Satish kumar K. Information extraction using by big data. Int J Pure Appl Math 2018;118(24):1–21.
[53] Russom P. Big data analytics – TDWI best practices report introduction to big data analytics. Tdwi Res 2011;1:3–5.
[54] Laude H. France's governmental big data analytics: from predictive to prescriptive using R. From predictive to prescriptive using R. Federal data science: transforming government and agricultural policy using. Artif Intell 2018:81–94.
[55] Makin IW, Bastiaanssen WG, Molden DJ. Remote sensing for irrigated agriculture: examples from research and possible applications. Agric Water Manag 2000;46(2):137–55.
[56] Cleary P, on behalf of the outbreak control t, C., Syed Q, Lane C, Quigley C, Kearney J, Fox A. A foodborne outbreak of Salmonella Bareilly in the United Kingdom, 2010. Eurosurveillance 2017;15(48):5–7.
[57] Faulkner A, Cebul K. Agriculture gets smart: the rise of data and robotics. Cleantech Group 2014:1–6.
[58] Cornell University, INSEAD, and the W. I., & Property Organization. The Global Innovation Index 2017 innovation feeding the world. In the global innovation index 2017: Innovation feeding the world; 2017.
[59] Garnett T, Appleby MC, Balmford A, Bateman IJ, Benton TG, et al. Sustainable intensification in agriculture: premises and policies. Science 2013;341(6141):33–4.
[60] Jaruzelski B, Staack V, Johnson T. Innovation from Farm to Table. strategy+business; 2017. https://www.strategy-business.com/article/Innovation-from-Farm-to-Table.
[61] Lepoutre D, Layrol L. Satellite-derived management zones – tech briefs. NASA Tech Briefs Media group; 2005. https://www.techbriefs.com/component/content/article/tb/techbriefs/bio-medical/264.
[62] Marr B. The incredible ways John Deere is using artificial intelligence to transform farming. Forbes; 2018. https://www.forbes.com/sites/bernardmarr/2018/03/09/the-incredible-ways-john-deere-is-using-artificial-intelligence-to-transform-farming/#4cfdbc5b330d.
[63] Zong Y, Wang Y, Li C, Zhang R, Chen K, Ran Y, et al. Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. Nat Biotechnol 2017;35(5):438–40.
[64] Venkatesh A, Bhandari V, Pande V. Big data for farmers. The Indian Express; 2018. https://indianexpress.com/article/opinion/columns/indian-farmers-soil-health-cards-narendra-modi-govt-indian-agriculture-5333347
[65] Geissbauer R, Evelyn Lübben, Schrauf S, Pillsbury S. Global digital operations study 2018: digital champions. PwC Network, 64; 2018. https://www.strategyand.pwc.com/gx/en/insights/industry4-0/global-digital-operations-study-digital-champions.pdf.
[66] Stafford JV. Implementing precision agriculture in the 21st century. J Agric Eng Res 2000;76(3):267–75.
[67] Malczewski J. GIS-based multicriteria decision analysis: a survey of the literature. Int J Geograph Inform Sci 2006;20(7):703–26.
[68] Verdouw CN, Robbemond RM, Wolfert J. ERP in agriculture: lessons learned from the Dutch horticulture. Comput Electron Agric 2015;114:125–33.
[69] Sørensen CG, Fountas S, Nash E, Pesonen L, Bochtis D, Pedersen SM, et al. Conceptual model of a future farm management information system. Comput Electron Agric 2010;72(1):37–47.
[70] Davenport TH, Davenport. Mission critical: realizing the promise of enterprise systems. USA: Harvard Business School Press; 2000.
[71] Hansen JW. Is agricultural sustainability a useful concept?. Agric Syst 1996;50(2):117–43.
[72] Maheswarappa HP, Srinivasan V, Lal R. Carbon footprint and sustainability of agricultural production systems in India. J Crop Improvement 2011;25(4):303–22.
[73] Perritt HH. Open government. Gov Inform Quart 1997;14(4):397–406.
[74] Sekhar C, Kumar JU, Kumar BK, Sekhar C. Big data analytics on indian crop planning to increase agricultural production. Adv Sci Technol Lett 2018;147:211–6.
[75] Moorthy J, Lahiri R, Biswas N, Sanyal D, Ranjan J, Nanath K, et al. Big data: prospects and challenges. Vikalpa 2015;40(1):74–96.
[76] Carbonell IM. The ethics of big data in big agriculture. Internet Policy Review. 2016;5(1):1–13.
[77] Gilpin L. How big data is going to help feed nine billion people by 2050 – TechRepublic. https://www.techrepublic.com/article/how-big-data-is-going-to-help-feed-9-billion-people-by-2050. 2014.
[78] Lesser A. Big data and big agriculture – Gigaom; 2014. https://gigaom.com/report/big-data-and-big-agriculture.
[79] Poppe K, Wolfert S, Verdouw C, Renwick A. A european perspective on the economics of big data. Farm Policy J 2015;12(1):11–9.
[80] Tong L, Hong T, JingHua Z. Research on the Big Data-based government decision and public information service model

of food safety and nutrition industry. J Food Safety Quality 2015;6:366–71.
[81] Wolfert S, Ge L, Verdouw C, Bogaardt MJ. Big data in smart farming – a review. Agric Syst 2017;153:69–80.
[82] Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Hung Byers A. Big data: the next frontier for innovation, competition, and productivity. https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation. 2011/2019 [Retrieved].
[83] Van Alstyne M, Brynjolfsson E, Madnick S. Why not one big database? Principles for data ownership. Decis Support Syst 1995;15(4):267–84.
[84] Bamford Roger J, Chandrasekaran Sashikanth, Pruscino Angelo. Dynamic reassignment of data ownership. Oracle International Corp.; 2004. https://patentimages.storage.googleapis.com/3e/41/94/b43035fd33b8d8/US7277897B2.pdf [Published]/2019[Retrieved].
[85] Gilpin L. How big data is going to help feed nine billion people by 2050 – TechRepublic; 2014. Retrieved from https://www.techrepublic.com/article/how-big-data-is-going-to-help-feed-9-billion-people-by-2050/.