

Big Data Challenges and Opportunities in Agriculture

Maya Gopal P.S., VIT University, Chennai, India

Bhargavi Renta Chintala, School of Computing Science and Engineering, VIT University, Chennai, India

ABSTRACT

This article reviews various aspects of research concerning the background and state-of-the-art of big data in agriculture. This article focuses on data generation, storage, analysis and visualization in big data. In every phase, technical challenges and the latest advancement are discussed, and these discussions aim to provide a comprehensive overview and complete picture of this exciting area. This survey is concluded with a discussion on the application of big data in precision agriculture and its future directions.

KEYWORDS

Agriculture, Big Data, Data Acquisition, Data Analytics, Data Storage, Machine Learning

1. INTRODUCTION

The United Nations estimates that it is the need to increase the food production to 50 percent by the middle of the current century (FAO, 2009). Agricultural production tripled during the last decades as the world's population doubled (Kitzes et al., 2008). Agriculture remains essentially a primary source of food for the population and raw material for a large number of industries (da Silva et al., 2009). Population growth, climate change and bio-energy crops are worldwide trends that are increasing the importance of using science to improve agriculture (Tilman et al., 2011). With the need to produce more food using fewer inputs, agriculture is seeking new products, practices and technologies. Research activities centering on genomics, bioinformatics and computational biology of plants and animals enable the scientists and organizations to better feed the world and improve the quality of food and animal crops. Progress in agricultural growth can serve as a critical position for designing successful strategies to transform the economy and meet sustainable development (Christiaensen et al., 2010) and the investments in agricultural research play a key role to agricultural growth. Farmers will have the tools to get the most from every acre. The future of farming depends largely on adoption of cognitive solutions. While large scale research is still in progress and some applications are already available in the market, the industry is still highly underserved. When it comes to handling realistic challenges faced by farmers and using autonomous decision making and predictive solutions to solve them, farming is still at a budding stage (Jones et al., 2017). Research on new generation agricultural design models shows that the data is most important parameter for

DOI: 10.4018/IJAEIS.2020010103

This article, originally published under IGI Global's copyright on January 1, 2020 will proceed with publication as an Open Access article starting on February 2, 2021 in the gold Open Access journal, International Journal of Agricultural and Environmental Information Systems (converted to gold Open Access January 1, 2021), and will be distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

on-farm decision support, research investment and policy decision making. The agricultural industry will be transformed by data science and artificial intelligence. Collecting reliable agriculture data for farm management decision making is important scenario. The developments in the concept of smart farming make agriculture more efficient and effective with the help of high-precision algorithms (Baseca et al., 2019). The mechanism used in smart farming is machine learning (ML), the scientific field that gives machines the ability to learn without much programming. It has emerged together with big data technologies and high-performance computing to create new opportunities to ease, quantify and understand data intensive processes in agricultural operational environments. The developments indicate that agriculture can benefit from machine learning at every stage like spices management, field management, crop management and livestock management. The artificial intelligence (AI) and machine learning are used in a number of agricultural applications today include the yield prediction algorithms based on weather and historical yield data, image recognition algorithms to detect pest and diseases in plants and robotics to harvest different types of specialty crops (Tibbetts 2018). This aspect needs an adaptive method to control the data sources and decision-making systems for better production and marketing with less waste of resource. Agriculture big data is playing important role by incorporating the AI and ML. The farmers are using data to calculate harvest yields, fertilizer demands, cost savings and even to identify optimization strategies for future crops as smart machines and sensors on farms and farm data grow in quantity and scope, farming processes will become increasingly data driven and data enabled.

In order to obtain better productivity, the people are using precision agriculture (De Rango et al., 2019), Somayeh et al. (2018) and Maes and Steppe (2019), automated irrigation scheduling (Li et al., 2018; Soulis & Elmaloglou, 2018), optimization of plant growth, farmland monitoring, greenhouse gases monitoring, production process management and security in crops (Yuan et al., 2018; Huang et al., 2018, Groenveld et al., 2019).

2. BIG DATA TECHNOLOGIES AND TOOLS

Recent technological development led to automation of several processes in various domains like agriculture, health care, fraud detection, etc., which in turn led to the generation of humungous data. McKinsey & Co. (Manyika et al., 2011)) foresees that the society is now facing a tremendous wave of innovation, productivity, and growth as well as new modes of competition and value capture - all driven by Big Data. The term Big Data is mainly used to describe massive, often unstructured, and heterogeneous digital content which is difficult to store and process using traditional data management tools and techniques Talia (2013), Stephen Kaisler et al. (2015), Rob Lokers et al. (2016)). Once these started gaining attention, the data analyst developed it further and currently big data can be described using the 10V model. Borne (2014) has listed the V's as challenges in deploying Big Data into any application. These V-based characterizations represent ten different challenges associated with the main tasks involving big data like- capture, cleaning, curation, integration, storage, processing, indexing, search, sharing, transfer, mining, analysis, and visualization. Big data can be described by the following 10 characteristics which are illustrated in Figure 1 and Table 1.

These covered most of the challenges in big data including collecting, storing, transferring, analyzing, and visualizing. Big data is primarily defined by the volume of a data set. The data sets are generally huge measuring tens of terabytes and some cases crossing the petabytes. The term big data was preceded by very large databases which were managed using database management systems. Currently, big data falls under three categories of data sets namely structured, unstructured and semi-structured. The structured data sets comprise of data which can be used in its original form to derive results. The unstructured data sets comprise of data are without proper formatting and alignment (Khan et al., 2014). Semi-Structured data sets are a combination of both structured and unstructured data. Big data processing requires a particular setup of physical and virtual machines to derive results. The processing is done simultaneously to achieve results as quickly as possible.

Figure 1. The 10 V's of big data

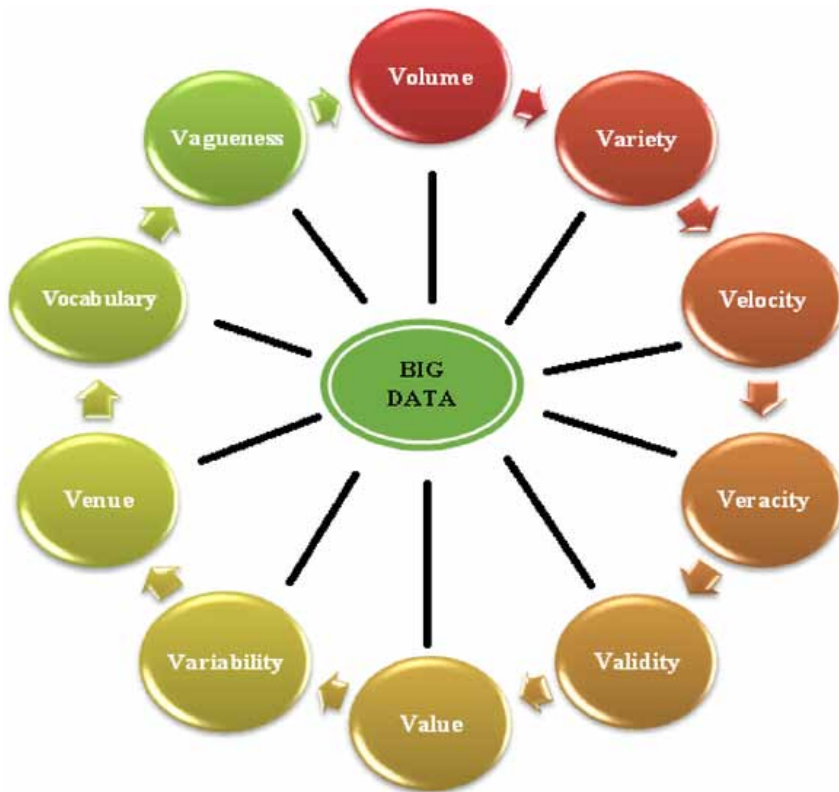


Table 1. The 10 V's of big data

| | |
|-------------|---|
| Volume | There is a large amount of data is generated from various sources. |
| Variety | Data has complex structures, different data types, and formats depending on the data sources. |
| Velocity | In real time processing, the rate of data flowing into and out of the systems is high. |
| Veracity | Due to the different data sources are often of many different origins data quality is not all verifiable. |
| Validity | Validity is based on veracity, the data should have quality, governance, master data management (MDM) on massive, diverse, distributed and heterogeneous. |
| Value | To derive significant value from high volumes of data with a low value density is not straight forward. |
| Variability | The data source should be dynamic, evolving, spatio-temporal data, time series, seasonal, and any other type of non-static behavior, customers, objects of study, etc. |
| Venue | Data sources are distributed, heterogeneous data from multiple platforms, from different owners' systems, with different access and formatting requirements, private vs. public cloud. |
| Vocabulary | It relates new concepts, schema, data models, semantics, ontologies, taxonomies, and other content- and context-based metadata that describe the data's structure, syntax, content, and provenance. |
| Vagueness | It relates to the confusion over the meaning of big data and overall developments around big data. |

These days, big data processing techniques also include Cloud computing and artificial intelligence. These technologies help in reducing manual inputs and oversight by automating many processes and tasks. Data sets are consigned the big data status based on technologies and tools required for their processing. In agriculture big data, various data models and algorithms are needed at different stages starting from data collection, storage (pre-processing), analytics (processing), visualization (post-processing) to decision-making. The development of big data computing technologies and tools such as Hadoop, HDFS, Bigtables, HBase, HIVE, CASSANDRA, ZOOKEEPER, MAHOUT, PIG, Map Reduce, NoSQL, SPARK, Storm etc., have eased the process of exploring the Big Data for various applications (Moorthy et al., 2015). Simulation models play an important role in informing farmer practices by online crop production model like Yield Prophet, breeding strategies by genetic algorithm and government policies by integrated assessment models like Seamless (Hochman et al., 2009, Cooper et al., 2009, Bezlepina et al., 2010).

3. CONCEPTUAL FRAMEWORK OF THIS WORK

In order to analyze the challenges and opportunities in agriculture industry using big data a conceptual framework was developed to provide a systematic classification of issues and concepts (Wolfert et al., 2019). A major complexity of such applications is that they require adoption between the different stakeholders having different roles in the data value chain. There are two main components of the architecture defined and are the data platform for data storage and processing and the operation platform directly dealing with on-site and providing a connection to the agricultural environment.

The data platform is connected to the operation platform using the big data storage and analytics platform, development tools, semantic modeling tools and simulation and resource optimization framework. The architecture of the data platform is shown in Figure 2. The interface allows the uploading of a large batch of historical data collected on the site to the cloud data storage, real-time asynchronous communication between the data and operational platform and export of predictive functions built on the data platform and their deployment on the operation platform for scoring of the operational data.

The subsections provide a detail description of the subcomponent of the data platform and operation platform.

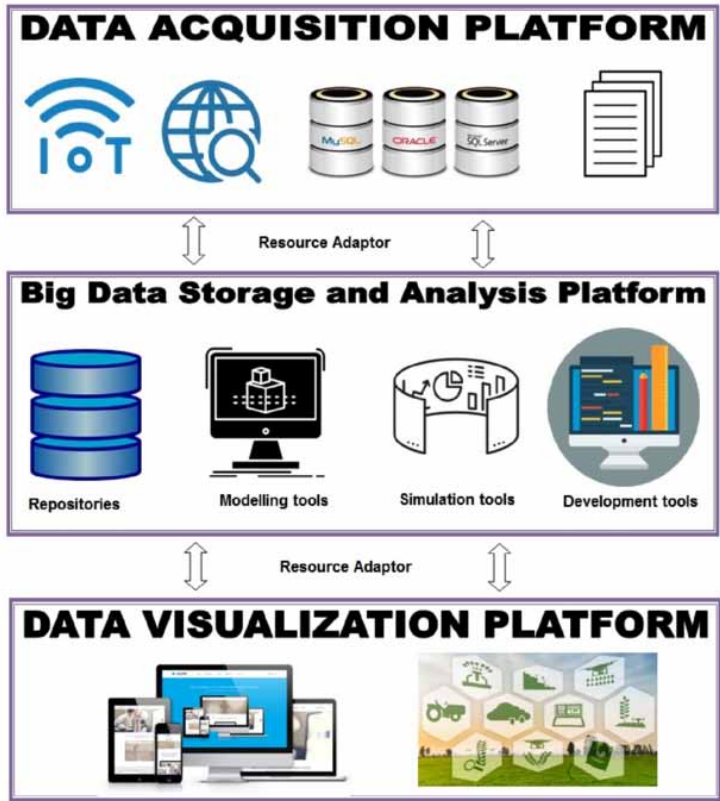
3.1. Big Data Storage and Analytics Platform

The Big Data storage and analytics platform provide resources and functionalities for storage as well as for batch and real-time processing of the Big Data. It provides main integration interfaces between the site operational platform and the Cloud data platform and the programming interfaces for the implementation of the data mining processes.

Data acquisition in big data has two components: identification and collection of Big Data. Identification of Big Data is done by analyzing the two natural formats of data generated by digital and analogue. Digital data is the information which has been captured through a digital medium, e.g. a computer or smart phone app, etc. This type of data has an ever-expanding range since systems keep on collecting different kinds of information from users. Digital data is traceable and can provide both personal and demographic business insights. Data in the form of pictures, videos and other such formats which relate to physical elements of our world are termed as analogue data. Analogue data requires conversion of data into digital format by using sensors, such as cameras, voice recording, digital assistants, etc. The increasing reach of technology has also raised the rate at which traditionally analogue data is being converted or captured through digital mediums.

In agriculture domain and its applications, data collection from various sources plays an essential role. They are classified into two categories named public and private. Public data comprises of records that are collected, maintained and analyzed through publicly funded sources, specifically by government agencies. The important public agriculture data generating sources are weather records by

Figure 2. Conceptual framework



meteorological information and satellite imagery, longer term climate, national soil database, digital elevation models, markets info, regional and national inventories etc. Private data comprises of data records generated at the production level and originate with the farmer, who collect data as part of their normal day-to-day activities such as yield, soil analysis, irrigation levels, livestock movement and grazing rates. In agriculture and horticulture, a well-established range of instruments are available for measuring parameters such as mass, volume, temperature, relative humidity, gas and fluid flow. GPS-based applications in precision farming are being used for farm planning, field mapping, soil sampling, tractor guidance, crop scouting, variable rate applications, and yield mapping (Benoit et al., 2012). Crop sensors play a vital role to ease the quality crop by automatic identification and cropping (Benoit et al., 2012; Rosell et al., 2012). Close-range sensing using tethered balloons, small radio-controlled aircrafts has been employed to gather photographic or multi-spectral information on crops and soils. Chemical sensors are also used to register the presence and amount of specific chemicals or chemical groups (Larisa Lvova et al., 2017)). The advancements of sensors in agriculture data collections contributed to enormous development in the farming methods (Bochtis et al., 2019). The developments in precision agriculture, automated irrigation systems, optimization of plant growth, production monitoring are a few potential applications. This important data for industry is created and captured at high cost. It needs bulk storage devices for experimental data bases, array storage for large-scale scientific computations, and large output files.

For Smart Farming, the potential of unmanned aerial vehicles (UAVs) has been well-recognized (Faulkner and Cebul, 2014; Holmes, 2014). Drones with infrared cameras, GPS technology, are transforming agriculture with their support for better decision making, risk management (Zilberman

et al., 2018). In livestock farming, smart dairy farms are replacing labour with robots in activities like feeding cows, cleaning the barn, and milking the cows (Scott et al., 2015). On arable farms, precision technology is increasingly used for managing information about each plant in the field (Pivoto et al., 2018). With these new technologies, data is not in traditional tables only, but can also appear in other formats like sounds or images (Sonka, 2015). The advances in data analysis techniques trigger the use of data in images or other formats. In Big Data and Smart Farming, human-sourced data have rarely been discussed except in relation to the marketing aspects (Verhoosel et al., 2016). Limited capacity with regard to the collection of relevant social media data and semantic integration of these data from a diversity of sources is considered to be a major challenge (Bennett, 2015).

Data are generated through precision agriculture. Micro-irrigation techniques are cost-effective and water-usage efficient and it need to adopt the parameters from environmental and soil information (Ge et al., 2018, Kamilaris et al., 2017, Sharma et al., 2018). The controlled usage of fertilizers, pesticides need to monitor the probability and occurrence of pests in crops and the surrounding climate information (Shrivastava et al., 2019). Cattle moment tracking and asset tracking are important for agriculture and these data are generated by RFID sensors and monitored (Kwong et al., 2012; Misra et al., 2012).

To better prepare for disruptions in food supply and global crop market price fluctuations of the types witnessed over the last 10 years, timely and accurate information on current and forecasted global food production is needed (Fritz et al., 2019). Improved monitoring will enable more accurate forecasting of commodity prices and a better understanding of the key risks in food supply, helping to reduce global food insecurity (Godfray et al., 2010).

The rainfall estimation and the Normalized Difference Vegetation Index (NDVI), GIEWS use the Agricultural Stress Index (ASI) for early identification of agricultural areas that may be affected by dry spells or droughts (Rojas et al., 2011). The agrometeorological models based on data obtained from national meteorological networks will show hotspots where crops are affected by water stress during the growing period and also it can use for verifying data from public institutions (Rojas, 2015). Early warning systems network will provide decision support to food assistance programs and relief agencies (Funk and Verdin, 2010).

A decision support system, the MARS crop yield forecasting system provides independent and evidence-based information on the status of annual crops in the EU and neighboring countries by monitoring crop growth and forecasting crop yields (Supit et al., 1994; Baruth et al., 2008; Gallego et al., 2010; Boogaard et al., 2013; López-Lozano et al., 2015).

Agricultural Big Data will have no real value without Big Data analytics (Sun et al., 2013). To obtain Big Data analytics, data from different sources need to be integrated. In this process, data quality issues are likely to arise due to errors and duplications in data. The second step in the data acquisition process is collection and storage of data sets identified as big data since the old DBMS techniques were inadequate for managing big data. The process is called MAD – magnetic, agile and deep is used for collecting and storing big data. Since, managing big data requires a significant amount of processing and storage capacity, creating such systems is out-of-reach for most entities which rely on big data analytics. Thus, the most common solutions for big data processing today are based on two principles – distributed storage and Massive Parallel Processing (MPP). The high-end Hadoop platforms and specialty appliances use MPP configurations in their system. JavaScript Object Notation (JSON) is the preferred protocol for saving big data nowadays. Using JSON, the tasks can be written in the application layer and allow better cross-platform functionalities. Data are primarily stored in the Distributed File System, which is responsible for the distribution and replication of large datasets across the multiple servers (data nodes) (Ahmed Oussous et al., 2018). A unified access to the structured data is provided by the Distributed Database using the standard standard query language (SQL) interface.

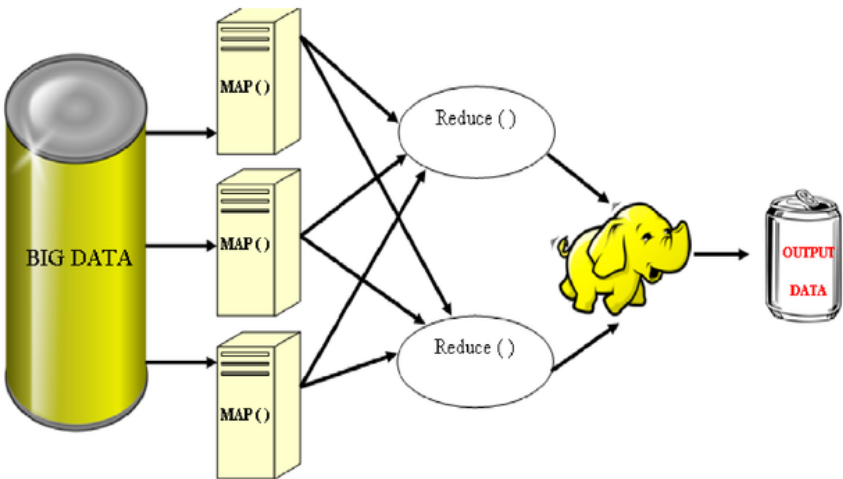
Hadoop is a distributed file system for storing and performing big data analytics, which provide reliability, scalability, and manageability. Hadoop consists of two main components:

the Hadoop distributed file system (HDFS) for the data storage and MapReduce (MR) for data analytics. MapReduce (MR) is an effective programming model used in cloud computing and large-scale data-parallel applications. The schematic of the MapReduce is shown in Figure 3. It is used for processing vast data sets with large number of parallel distributed algorithms on a cluster. Both the data nodes and the name node of HDFS provide a redundant and reliable distributed file system, which is optimized for large files, where a single file is split into blocks and distributed across cluster nodes. Tachyon is the distributed file system that extends HDFS and provides access to the distributed data at memory speed across the cluster. Some of the features where Tachyon has outsmarted HDFS are in-memory data caching for enhanced performance and backwards compatibility to work seamlessly with Spark as well as MR tasks without any code changes required to the programs. The other Big Data storage methods are object-based mechanism, that are able to scale to very high capacity and enormous files so that enterprises can take the advantage of Big Data as the DNS way of doing things over the Internet (Bhat, 2018). MapReduce has also been shown to have significant problems with more complex algorithms, like conjugate gradient, fast Fourier transform and block tri-diagonal linear system solver (Cohen, 2009). MapReduce used to predict whether, agricultural commodity price analysis and crop disease analysis (Hu et al., 2015; Doreswamy & Gad, 2016). Yet another resource negotiator (YARN) is introduced that has taken Hadoop to an actually computationally-agnostic Big Data platform. MR runs as a service over YARN, while YARN handles scheduling and resource management related functionalities. This separation has made Hadoop suitable for implementing innovative applications. Apache Spark and Microsoft's Dryad provide the big-data processing framework as Apache Hadoop (Hosseini et al., 2019).

Agricultural Big Data are known to be highly heterogeneous (Ishii, 2014, Li et al., 2014). The heterogeneity of data concerns for example the subject of the data collected (i.e., what is the data about) and the ways in which data are generated. Data collected from the field or the farm include information on planting, spraying, materials, yields, in-season imagery, soil types, weather, and other practices.

In North America, several initiatives are undertaken to open up data transfer between several platforms and devices. The ISOBlue project facilitates data acquisition through the development of open-source hardware platforms and software libraries to forward ISOBUS messages to the cloud and develop applications for Android smart phones (Layton et al., 2014). They emphasize the capability

Figure 3. Representation of MapReduce in big data



to share data records easily between several workers within the farmer stakeholders outside the farm and the guarantee of long-term ownership of farmer's data (Welte et al., 2013, Ault et al., 2013).

In Europe, much work to realize an open infrastructure for data exchange and collaboration was done within their Future Internet programme. The focus of this programme was to realize a set of Generic Enablers (GEs) for e.g. cloud hosting, data and context management services, IoT services, security and Big Data Analysis which are common to all Future Internet applications for all kind of different sectors, called FIWARE (Wolfert et al., 2014, Kaloxylou et al., 2012; Kaloxylou et al., 2014). The Future Internet space project implemented the interface architecture into a real platform for business collaboration (Barmounakis et al., 2015; Wolfert et al., 2014). Janssen et al. (2017) provide examples of data and information technology structures that illustrate how private and public data components could be developed for such use cases. Jones et al. (2017) argue that the most important current limitation is data, both for on-farm decision support and for research investment and policy decision making. One of the greatest data challenges is to obtain reliable data on farm management decision making both for current conditions and under scenarios of changing bio-physical and socio-economic conditions.

Data integration combines data residing at different sources and provides the user with a unified view of these data is a key method. An effective solution to address the data integration problem is to learn good data representations from each individual data source and then to integrate the learned features at different levels (Berman, 2018). Srivastava et al. (2012) developed a novel application of deep learning algorithms to learn a unified representation by integrating real-valued dense image data and text data. Another tedious task is associated with high variety i.e. the data are often high dimensional and nonlinear, such as global climate patterns, stellar spectra, and human gene distributions. To deal with high-dimensional data, dimensionality reduction is an effective solution through finding meaningful low-dimensional structures hidden in their high-dimensional observations.

For big data, speed or velocity really matters, which is another emerging challenge for learning. Exploiting a variety of learning methods to analyze big datasets, it is to extract valuable information from massive amounts of data in the form of deep insight or commercial benefits. Therefore, value is also characterized as a salient feature of big data. However, to derive significant value from high volumes of data with a low value density is not straightforward. To handle this challenge, knowledge discovery in databases (KDD) and data mining technologies (Guarascio et al., 2019; Perera et al., 2016) comes into play. These technologies provide possible solutions to find out the required information hidden in the massive data. Few authors reviewed studies on applying data mining and KDD technologies to the IoT. Particularly, utilizing clustering, classification, and frequent patterns technologies to mine value from massive data in IoT, from the perspective of infrastructures and from the perspective of services were discussed in detail. Wu et al. (2014) characterized the features of the big data revolution and proposed big data processing methods with machine learning and data mining algorithms. Another challenging problem associated with the value of big data is the diversity of data meaning, i.e., the economic value of different data varies significantly, even the same data have different value if considering from different perspectives or contexts.

The other system used to handle the big data is the cloud agile technology with the advantages of data centric, storage and data management confluence, and storage information infrastructure-oriented. Cloud-based storage is able to provide virtually unlimited storage and flexible access to the data so as to carry out various applications and services via the internet (Elshawi et al., 2019).

3.2. Data Visualization

The main focus of data visualization is to convey the information or the knowledge, which are dug in the complex and large-scale data sets in a visual form (Rice, 2018). Data presentation is crucial, since the end user struggles to infer the vital information of the generated data from various sources, because they are usually unstructured and heterogeneous. For Big Data applications, data visualization is extremely difficult due to large size, high dimension of data and its crucial V's. To visualize large-scale

data, feature extraction and geometric modeling can be implemented, to decrease the data size before actual processing. Visualization based tools should fulfill three requirements such as expressiveness, effectiveness and appropriateness. Through climate visualization tools, various climate parameters are visualized. Some of well-known techniques for data visualization are Tag cloud, Clustergram, Motion charts and Dashboard (Van Hertem et al., 2017; Grainger et al., 2016). The operation data visualization is a web user interface where operational managers can configure various real-time visualizations of operational data and monitor the deployed predictive functions. The visualized data can be enhanced with the new trend indicators using various trend analysis methods. The operational data visualization framework also provides API for the implementation of the new trend indicators.

4. OPPORTUNITIES AND ADVANCEMENTS AND UTILIZATION OF BIG DATA IN AGRICULTURAL INDUSTRY

As opportunities for Big Data have surfaced in the agribusiness sector, big agriculture companies such as Monsanto and John Deere have spent hundreds of millions of dollars on technologies that use detailed data on soil type, seed variety, and weather to help farmers cut costs and increase yields (Faulkner & Cebul, 2014). Other players include various accelerators, incubators, venture capital firms, and corporate venture funds (Monsanto, DuPont, Syngenta, Bayer, DOW etc.) (Lane, 2015).

As the agri tech develops, an increasing number of small tech startups are launching products giving their bigger counterparts a run for their money. In the USA, start-ups like FarmLogs (Guild, 2014), Farm Link (Hardy, 2014) and 640 Labs challenge agribusiness giants like Monsanto, Deere, DuPont Pioneer (Plume, 2014). One observes a swarm of data-service start-ups such as FarmBot (an integrated open source precision agriculture system) and Climate Corporation. Their products are powered by many of the same data sources, particularly those that are freely available such as from weather services and Google Maps. They can also access data gathered by farm machines and transferred wirelessly to the cloud. Traditional agri-IT firms such as NEC and Dacom are active with a precision farming trial in Romania using environmental sensors and Big Data analytics software to maximize yields (NEC, 2014).

In the state of the art of big data applications in agriculture, the data are captured using sensors, open data, using biometric sensing genotype information (Faulkner and Cebul, 2014; Cole et al., 2012, Van 't Spijker, 2014)). The issue is the availability is limited and the quality of the sensors is immature due to the lack of developments (Tien, 2013). In data storage, the Cloud-based platform, Hadoop distributed file system and hybrid system are used. The system needs to improve quick and safe access of data along with installation and maintenance cost. Wireless Cloud-based platforms are used for data transfer even though it has the issues on safety, agreements on responsibilities and liabilities. To make Big Data applications for Smart Farming work, an appropriate technological infrastructure is essential.

The development in big data technologies improves the comprehensive productivity of agriculture and the overall productivity of agriculture using modern information technology and information systems (El Bilali et al., 2018). Because of volume, diversity, and complexity of agriculture data sets, there are number of challenges for innovative architecture and frameworks, algorithms, and analytics to manage, to extract values and hidden knowledge from it. Big data may play a vital role for effective decision making for farmers, policy makers, companies involved in agri-business and agro-inputs, banks, insurance companies and service providers. The major opportunities and challenges lie with setting a benchmark in agriculture sector, because the factors influencing the agriculture will vary with climate, area, soil type, culture and tradition.

The considerable challenges of agriculture big data exist at different stages, such as data collection, storage and analysis since the agriculture data set contains various data like soil, climate, seed, cultivation practices, irrigation facilities, fertilizers, pesticides, weeds, harvesting, post harvesting techniques, etc. A huge amount of data is generated and maintained by governments, universities,

research organizations, agri-business and agro-input companies for agriculture production, insurance, marketing, supply chain, packaging, distribution, etc. (Shrivastava et al., 2019). Because of the multimodal nature of data, it has several challenges like the improving methods for data collection, effective and efficient statistical and data analytical techniques to understand and support the functions of various agriculture verticals. The agriculture big data is not so sensitive. So, there is less security or privacy issues in agriculture data. The agriculture practices guide agriculture data mining (Kamilaris et al., 2017).

The challenges are also in implementation costs side. The challenge is to automate data acquisition in such a way that there are virtually no costs (Sonka, 2015). Because on-farm data will generally remain in the hands of individual companies, investments are needed in a common pool infrastructure to transfer and integrate data and finally make applications out of it. Poppe et al. (2015) suggests an coordination between the Agricultural Business centers and Data Exchange Facilities of the area. The challenge in this area is if the coordination will be closed, proprietary systems or if these will be more open. Another business-related challenge of Big Data is how the potential of information across food systems can be utilized (Sonka, 2015).

The biggest challenges of Big Data governance are probably how to ensure privacy and security (Orts & Spigonardo, 2014; Sonka, 2014; Van 't Spijker, 2014). Currently, this is sometimes inhibiting developments when data are in silos, guarded by employees or companies because of this issue. They are afraid that data fall into the wrong hands (e.g. of competitors) (Gilpin, 2015b). Hence privileged access to Big Data and building trust with farmers should be a starting point in developing applications (Van 't Spijker, 2014). Therefore, new organizational linkages and modes of collaboration need to be formed in the agri-food chain (Sonka, 2014). The ability to quickly access the correct data sources to evaluate key performance/core processes and outcome indicators in building successful growth strategies. Anonymization of data, so that it cannot be traced back to individual companies can also be a problem sometimes (Orts & Spigonardo, 2014).

Assessing impacts of new technologies and climate adaptation are challenging, most agricultural technology impact assessment is carried out after technologies have been disseminated. However, there is a growing recognition of the need for forward-looking, technology impact assessment designed to anticipate both intended and unintended impacts (Antle, 2011; Antle et al., 2014). One of the most important growing applications of forward-looking impact assessment is for climate adaptation and climate smart agriculture (Lipper et al., 2014).

Capalbo et al. (2017) present an illustrative analysis of how climate change may impact dry-land wheat producing farmers in the U.S. Pacific Northwest. Projected changes in climate are translated into changes in key climate factors affecting the grower's yields and these yield changes are transformed into net returns.

As Antle et al. (2017) observe that the users do not want models, but they want the information they can produce. This means that models must be embedded in decision support tools that have value to farm managers. One improvement could be to automate data collection using sensors on machinery and other mobile devices, as well as from web-based sources such as weather, and economic data such as prices. Another area of improvement is inter-operability of tools with farm accounting and tax preparation software, so that information can be entered once and then utilized in an integrated way across multiple analytical tools. The available approach of manually carrying out this integration on a case-by-case basis makes this type of analysis costly even in a small geographic region, and often makes integration infeasible across larger regions.

Second, the fact that virtually all stakeholders want access to model outputs, rather than access to the models themselves, means that there is a demand for "knowledge products," i.e., tools that facilitate access to model outputs and provide analytical capability to interpret model outputs for decision making. As Janssen et al. (2017) observe, "apps" on mobile devices, as is now being done for some types of decision making such as pesticide spray decisions – or as larger tablet computer dashboards for data visualization and additional processing through meta models and other analytical

tools. The fact that these knowledge products have been slow to materialize suggests some form of “market failure”. As suggested by Antle et al. (2017), one solution to these challenges appears to be private-public partnerships among the various organizations that have a mutual interest in assuring that the data are obtained efficiently and used appropriately for both private and public purposes.

There have been major developments in the world related to food safety and traceability. Some of the initiatives come from governments to protect the health of the citizens, the others are private initiatives by growers and retailers in order to meet the expectations of their customers with respect to food safety and environmental sustainability. Precision agriculture technology helps the food safety by developing a better decision support system for agriculture industry presents some specific challenges in information industry. Using network, internet and sensors gives a complete picture for precision agriculture along with the disease management and cattle movement. Analysis tools based on network theory can be used in the control and monitoring systems of food business operators by analyzing their commercial relations with each other (Baranyi et al., 2013; Chmiel et al., 2007). The management challenges in closely co-operating enterprises, as well as the mutual dependence of all participants in the food chain, necessitate the application of network science in this area (Fritz et al., 2008). The cattle identification and traceability are becoming a necessary pre-condition for the international competitiveness of cattle and the cattle-product export market (Schroeder et al., 2012). The network analysis approach has applied to analyze and prevent foot and mouth disease (Dube et al., 2009). Martinez-López et al. (2009) have analyzed the trans-boundary flow of animals with the purpose of implementing disease prevention measures and Bajardi et al. (2011, 2012) mapped the Italian cattle trade network and made great progress in analyzing dynamic patterns, using network science tools to optimize cattle farm surveillance.

Improving the technologies in data transmission is important for the agriculture sector. The lack of data or possible delays in providing updated records may make their use difficult, especially for time-varying patterns. The available data transmission models, i.e. cellular, wireless, and satellite are lagging behind the needs of agriculture production. The recent report Co-Bank (Tayler et al., 2016) released a report which briefs that the farmers utilize 30 plus gigabytes of data per month during the harvest period. This is more than double what they were using three years ago and the need keeps growing. Cellular and hard-wired providers are not able to reach much on the rural areas. However, satellite-based internet connectivity could provide an effective solution over larger geographic areas without signal interruption. The amount of data generated from drones will depend on the type, frequency, and quality of images that are being taken. Bennett (2016) suggests that cellular connectivity is the only viable option for coverage map sharing.

Insurance is another potential application of big data in the agriculture. It estimates the crop yields or crop losses within several days, thus allowing governments or insurance companies to respond to catastrophes in a much timelier manner (Castillo et al., 2016). Improved contract design and creative usage of satellite allow governments and the private sector to extend insurance coverage (William et al., 2010).

Utilization of real time information sharing between farmers and researchers enables service providers to supply real time and personalized services based on a wide range of factors such as: location, crop, management practices, mechanization level, irrigation type, farm size and soil type. These inform to the farmers about the different choices of cultivation and take necessary actions whenever needed. Minal Sawant et al. (2016) used PRIDE and mKRISHI business models. The farmers of Dindori tehsil of the Nashik district of Maharashtra state, India are trained by researchers in order to get more production within short span of time. Due to the personalized crop protocol, agro-advisory and timely alerts, the average increase in productivity was found to be 64% in 2013–2014 and 112% in 2014–2015. It also contributed to around a 90% increase in farmer participation in the second year.

5. CONCLUSION

Current Big Data platforms in agriculture are supported by various processing, analytical tools as well as dynamic visualization. Such platforms enable to extract knowledge and value from complex dynamic environment. They also support decision making through recommendations and automatic detection of anomalies, abnormal behavior or new trends. In this paper, the basics of big data and its important phases like data acquisition and storage, data transportation and data analysis in the agro environmental domain are discussed. The challenges in the big data in different phases are discussed and proposed its possible remedies. Also, we discussed the advanced applications of big data in agricultural industry. This discussion should lead to the improvement of methodologies and techniques required to efficiently provide access to and semantically interlink sources from different disciplines for agricultural industry. The successful introduction of Big Data in non-agricultural sectors consistently stresses the business issues and opportunities determine the success of the domain. This phenomenon is appropriate for agriculture business as well. Going with the existing technologies and applying the big data in the agribusiness chain will make agriculture business a profitable. Some of these applications will utilize data that is not directly captured from agricultural operations to provide enhanced product and service offerings to the sector. Data generation and transmission across business entities seems to be a necessary component of these applications. Linking data from operations with information and knowledge from laboratories and agricultural sites will be required to effectively optimize sector performance. In order to do so, the best possible solution is open data, which is a powerful evidence-based tool for long term sustainable development by improving economic opportunities for farmers and health of consumers. Open access to research, meta-analysis, and open publication of data are vital resources for nutritional security. Providing a more specialized Big Data analysis to extend the information about the features, benefits and limitations of each framework and language will help the industry. Hence, technological issues in many Big Data areas can be further studied and constitute an important research topic.

REFERENCES

- Huang, Y., Chen, Z. X., Tao, Y. U., Huang, X. Z., & Gu, X. F. (2018). Agricultural remote sensing big data: Management and applications. *Journal of Integrative Agriculture*, 17(9), 1915–1931. doi:10.1016/S2095-3119(17)61859-8
- Antle, J. M. (2011). parsimonious multi-dimensional impact assessment. *American Journal of Agricultural Economics*, 93(5), 1292–1311. doi:10.1093/ajae/aar052
- Antle, J. M., Basso, B., Connant, R., Godfray, C., Jones, J., Herrero, M., & Wheeler, T. et al. (2017). Towards a new generation of agricultural system data, models, and knowledge products: Design and improvement. *Agricultural Systems*, 155, 255–268. doi:10.1016/j.agsy.2016.10.002 PMID:28701817
- Antle, J. M., Stoerovogel, J., & Valdivia, R. (2014). New parsimonious simulation methods and tools to assess future food and environmental security of farm populations. *Philos.Trans. R. Soc. B*, 369(1639), 20120280. doi:10.1098/rstb.2012.0280 PMID:24535388
- Ault, A. C., Krogmeier, J. V., & Buckmaster, D. (2013). Mobile, Cloud-based Farm Management: A Case Study With Trello on My Farm, 2013 Kansas City, Missouri, July 21–July 24, 2013. American Society of Agricultural and Biological Engineers.
- Bajardi, P., Barrat, A., Natale, F., Savini, L., & Colizza, V. (2011). Dynamical Patterns of Cattle Trade Movements. *PLoS One*, 6(5), e19869. doi:10.1371/journal.pone.0019869 PMID:21625633
- Bajardi, P., Barrat, A., Savini, L., & Colizza, V. (2012). Optimizing Surveillance for Livestock Disease Spreading Through Animal Movements. *Journal of the Royal Society, Interface*, 9(76), 2814–2825. doi:10.1098/rsif.2012.0289 PMID:22728387
- Baranyi, J., Józwiak, A., Varga, L., Mézes, M., Beczner, J., & Farkas, J. (2013). Application Potentials of Network Science, Bioinformatics and Systems Biology to Food Science. *Hungarian Science*, 174(9), 1094–1102.
- Barmponakis, S., Kaloxylas, A., Groumas, A., Katsikas, L., Sarris, V., Dimtsa, K., & Wolfert, S. et al. (2015). Management and control applications in Agriculture domain via a Future Internet Business-to-Business platform. *Information Processing in Agriculture*, 2(1), 51–63. doi:10.1016/j.inpa.2015.04.002
- Baruth, B., Royer, A., Klisch, A., & Genovese, G. (2008). The use of remote sensing within the MARS crop yield monitoring system of the European Commission. *Proceedings of the Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* (pp. 935–940). Academic Press.
- Baseca, C. C., Sendra, S., Lloret, J., & Tomas, J. (2019). A Smart Decision System for Digital Farming. *Agronomy*, 9(5), 216. doi:10.3390/agronomy9050216
- Bennett, J. D. (2016). Value Of Map Sharing Between Multiple Vehicles In The Same Field While Using Automated Section Control [MAB Thesis]. Kansas State University, Manhattan, KS.
- Bennett, J. M. (2015). Agricultural Big Data: Utilisation to discover the unknown and instigate practice change. *Farm Policy Journal*, 12, 43–50.
- Bezlepkin, I. V., Adenauer, M., Kuiper, M. H., Janssen, S. J. C., Knapen, M. J. R., Kanellopoulos, A., & Ittersum, M. K. V. et al. (2010). *Using the SEAMLESS Integrated Framework for Ex-ante Assessment of Trade Policies*. Wageningen, Netherlands: Wageningen Academic Publishers.
- Bhat, W.A. (2018). Bridging data-capacity gap in big data storage. *Future Generation Computer Systems*, 87(October), 538–548.
- Bochtis, D., Sørensen, C.G., & Kateris, D. (2019). Advances and Future Trends in Agricultural Machinery and Management.
- Boogaard, H., Wolf, J., Supit, I., Niemeyer, S., & van Ittersum, M. (2013). A regional implementation of WOFOST for calculating yield gaps of autumn-sown wheat across the European Union. *Field Crops Research*, 143, 130–142.

- Borne, K. (2014). Top 10 Big Data challenges—A serious look at 10 Big Data V's. MAPR. Retrieved from <https://www.mapr.com/blog/top-10-big-data-challenges-%E2%80%93-serious-look-10-big-data-v%E2%80%99s#VLk8Iy6mRYo>
- Capalbo, S. M. S. M., Seavert, C., Antle, J. M., Way, J., & Houston, L. (2017). Understanding trade offs in the context of farm-scale impacts: an application of decision-support tools for assessing climate smart agriculture. In L. Lipper, N. McCarthy, D. Zilberman, G. Branca, & B. Blankespoor (Eds.), *Climate Smart Agriculture: Building Resilience to Climate Change*. Rome: FAO.
- Castillo, M. J., Boucher, S., & Carter, M. (2016). Index insurance: Using public data to benefit small-scale agriculture. *The International Food and Agribusiness Management Review*, 19, 93–114.
- Christiaensen, L., Demery, L., & Kühl, J. (2010). The (Evolving) Role of Agriculture in Poverty Reduction. UNU-WIDER, Helsinki.
- Cohen, J. (2009). Graph twiddling in a MapReduce world. *Computing in Science & Engineering*, 11(4), 29–41. doi:10.1109/MCSE.2009.120
- Cole, J. B., Newman, S., Foertter, F., Aguilar, I., & Coffey, M. (2012). Breeding and Genetics Symposium: really Big Data: processing and analysis of very large data sets. *Journal of Animal Science*, 90(3), 723–733. doi:10.2527/jas.2011-4584 PMID:22100598
- Cooper, M., van Eeuwijk, F. A., Hammer, G. L., Podlich, D. W., & Messina, C. (2009). Modeling QTL for complex traits: Detection and context for plant breeding. *Current Opinion in Plant Biology*, 12(2), 231–240. doi:10.1016/j.pbi.2009.01.006 PMID:19282235
- da Silva, C. A., Baker, D., Shepherd, A. W., Jenane, C., & Miranda-da-Cruz, S. (Eds.). *Agro-Industries For Development*. Food and Agriculture Organization of the United Nations (FAO).
- De Rango, F., Potrino, G., Tropea, M., Santamaria, A. F., & Fazio, P. (2019). Scalable and light way bio-inspired coordination protocol for FANET in precision agriculture applications Computers & *Electrical Engineering*, 74(March), 305–318.
- Dubé, C., Ribble, C., Kelton, D., & McNab, B. (2009). A Review of Network Analysis Terminology and its Application to Foot-and-Mouth Disease Modelling and Policy Development. *Transboundary and Emerging Diseases*, 56(3), 73–85. doi:10.1111/j.1865-1682.2008.01064.x PMID:19267879
- El Bilali, H., & Allahyari, M. S. (2018, December). Transition towards sustainability in agriculture and food systems: Role of information and communication technologies. *Information Processing in Agriculture*, 5(4), 456–464. doi:10.1016/j.inpa.2018.06.006
- Elshaw, R., Sakr, S., Talia, D., & Trunfio, P. (2018). Big Data Systems Meet Machine Learning Challenges: Towards Big Data Science as a Service. *Big Data Research*, 14(December), 1–11. doi:10.1016/j.bdr.2018.04.004
- Far, S. T., & Rezaei-Moghaddam, K. (2018, March). Impacts of the precision agricultural technologies in Iran: An analysis experts' perception & their determinants. *Information Processing in Agriculture*, 5(1), 173–184. doi:10.1016/j.inpa.2017.09.001
- Faulkner, A., & Cebul, K. (2014). *Agriculture Gets Smart: The Rise of Data and Robotics, Clean tech Agriculture Report*. Cleantech Group.
- Food and Agriculture Organization of the United Nations. (2009). How to Feed the World in 2050.
- Fritz, M., & G. Schiefer. (2008). Food chain management for sustainable food system development: a European research agenda. *Agribusiness*, 24(4), 440–452.
- Fritz, S., See, L., Juan, C. L. B., Waldner, F., Jacques, D., Becker-Reshef, I., & McCallum, I. et al. (2019). A comparison of global agricultural monitoring systems and current gaps. *Agricultural Systems*, 168, 258–272. doi:10.1016/j.agry.2018.05.010
- Funk, C., & Verdin, J. P. (2010). Real-time decision support systems: the famine early warning system network. In M. Gebremichael & F. Hossain (Eds.), *Satellite Rainfall Applications for Surface Hydrology* (pp. 295–320). Springer. doi:10.1007/978-90-481-2915-7_17

Gad, I. (2016). Big Data Techniques HADOOP and MAP reduce for weather Forecasting. *International Journal of Latest Trends in Engineering and Technology*.

Gallego, F. J., Carfagna, E., & Baruth, B. (2010). Accuracy, objectivity and efficiency of remote sensing for agricultural statistics. In R. Benedetti (Ed.), *Agricultural Survey Methods* (pp. 193–211). Chichester, UK: Wiley. doi:10.1002/9780470665480.ch12

Ge, M., Bangui, H., & Buhnova, B. (2018). Big Data for Internet of Things: A Survey. *Future Generation Computer Systems*, 87(October), 601–614. doi:10.1016/j.future.2018.04.053

Gilpin, L. (2015). How Big Data Is Going to Help Feed Nine Billion People by 2050. Tech Republic. Retrieved from <http://www.techrepublic.com/article/how-big-data-is-going-to-helpfeed-9-billion-people-by-2050/>

Dick, W. J., & Wang, W. (2010). Government Interventions in Agricultural Insurance. *Agriculture and Agricultural Science Procedia*, 1, 4–12. doi:10.1016/j.aaspro.2010.09.002

Grainger, S., Mao, F., & Buytaert, W. (2016). Environmental data visualisation for non-scientific contexts: Literature review and design framework. *Environmental Modelling & Software*, 85(November), 299–318. doi:10.1016/j.envsoft.2016.09.004

Groenveld, T., Kohn, Y. Y., Gross, A., & Lazarovitch, N. (2019, March). Optimization of nitrogen use efficiency by means of fertigation management in an integrated aquaculture-agriculture system. *Journal of Cleaner Production*, 212(1), 401–408. doi:10.1016/j.jclepro.2018.12.031

Guarascio, M., Manco, G., & Ritacco, E. (2019). Knowledge Discovery in Databases. In *Encyclopedia of Bioinformatics and Computational Biology* (Vol. 1, pp. 336–341). Academic Press.

Guild, M. & Danaher, T. (2014). Big Data Comes to the Farm. Financial Sense. Retrieved from <http://www.financialsense.com/contributors/guild/big-data-farm>

Hardy, Q. (2014). A Low-cost Alternative to Pricy Big Data on the Farm. *The New York Times Blogs*.

Hochman, Z., van Rees, H., Carberry, P.S., Hunt, J.R., McCown, R.L., Gartmann, A., Holzworth, D., van Rees, S., Dalgliesh, N.P., Long, W., Peake, A.S., Poulton, P.L., McClelland, T. (2009). Re-inventing model-based decision support with Australian dryland farmers. Yield Prophet® helps farmers monitor and manage crops in a variable climate. *Crop Past. Sci.*, 60(11), 1057–1070.

Holmes, M. (2014). Different Industries Debate the Potential of UAVs and the Need for Satellite. *SatelliteToday.com*. Retrieved from <http://www.satellitetoday.com/technology/2014/10/24/different-industries-debate-the-potential-of-uavs-and-theneed-for-satellite/>

Hosseini, B., & Kiani, K. (2019). A big data driven distributed density based hesitant fuzzy clustering using Apache spark with application to gene expression microarray. *Engineering Applications of Artificial Intelligence*, 79(March), 100–113. doi:10.1016/j.engappai.2019.01.006

Hu, Y., Cai, X., & DuPont, B. (2015). Design of a web-based application of the coupled multi-agent system model and environmental model for watershed management analysis using Hadoop. *Environmental Modelling & Software*, 70(August), 149–162. doi:10.1016/j.envsoft.2015.04.011

Ishii, K. (2014). Big Data analysis in medicine, agriculture and environmental sciences. *Seibutsu Kagaku Kaishi*, 92, 92–93.

Aubert, B. A., Schroeder, A., & Grimaudo, J. (2012, December). IT as enabler of sustainable farming: An empirical analysis of farmers' adoption decision of precision agriculture technology. *Decision Support Systems*, 54(1), 510–520. doi:10.1016/j.dss.2012.07.002

Janssen, S., 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: Developments from information and communication technology. *Agricultural Systems*, 155, 200–212. doi:10.1016/j.agsy.2016.09.017 PMID:28701813

John, H. (2018). Tibbetts, The Frontiers of Artificial Intelligence. *Bioscience*, 68(1), 5–10. doi:10.1093/biosci/bix136

- Jones, J. W., Antle, J., Basso, B., Boote, K., Conant, R., Foster, I., & Wheeler, T. et al. (2017). Towards a new generation of agricultural system data, models and knowledge products: State of agricultural systems science. *Agricultural Systems*, 155, 269–288. doi:10.1016/j.agsy.2016.09.021 PMID:28701818
- Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I., ... & Keating, B. A. (2017). Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. *Agricultural systems*, 155, 269–288. PMID:28701818
- Kaisler, S., Armour, F., Money, W., & Alberto, E. J. (2015). Big Data Issues and Challenges. In *Encyclopedia of Information Science and Technology* (3rd ed., Vol. 1, pp. 363–370). Academic Press. doi:10.4018/978-1-4666-5888-2.ch035
- Kaloxyllos, A., Eigenmann, R., Teye, F., Politopoulou, Z., Wolfert, S., Schrank, C., & Kormentzas, G. et al. (2012). Farm management systems and the Future Internet era. *Computers and Electronics in Agriculture*, 89, 130–144. doi:10.1016/j.compag.2012.09.002
- Kaloxyllos, A., Groumas, A., Sarris, V., Katsikas, L., Magdalinou, P., Antoniou, E., & Maestre Terol, C. et al. (2014). A cloud-based Farm Management System: Architecture and implementation. *Computers and Electronics in Agriculture*, 100, 168–179. doi:10.1016/j.compag.2013.11.014
- Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017). A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture*, 143(December), 23–37. doi:10.1016/j.compag.2017.09.037
- Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Ali, M., Kamaleldin, W., & Gani, A. et al. (2014). Big data: Survey, technologies, opportunities, and challenges. *The Scientific World Journal*. doi:10.1155/2014/712826
- Kitzes, J., Wackernagel, M., Loh, J., Peller, A., Goldfinger, S., Cheng, D., & Tea, K. (2008). Shrink and share: Humanity's present and future ecological footprint. *Philos. Trans. Royal Soc. B: Biol. Sci.*, 363(1491), 467–475. doi:10.1098/rstb.2007.2164 PMID:17652075
- Kwong, K. H., Wu, T.-T., Goh, H. G., Sasloglou, K., Stephen, B., Glover, I., & Andonovic, I. et al. (2012). Practical considerations for wireless sensor networks in cattle monitoring applications. *Computers and Electronics in Agriculture*, 81, 33–44. doi:10.1016/j.compag.2011.10.013
- Lane, J. (2015). Digital Soil: The Four Secrets of the New Agriculture. *Biofuels Digest*. Retrieved from <http://www.biofuelsdigest.com/bdigest/2015/03/09/the-four-secrets-of-the-newagriculture/>
- Layton, A. W., Balmos, A. D., Sabpisa, S., Ault, A., Krogmeier, J. V., & Buckmaster, D. (2014). ISOBlue: An Open Source Project to Bring Agricultural Machinery Data into the Cloud. *Proceedings of the 2014*, Montreal, Canada, July 13–July 16, 2014. American Society of Agricultural and Biological Engineers.
- Li, D., Hendricks Franssen, H.-J., Han, X., Jiménez-Bello, M. A., Martínez Alzamora, F., & Vereecken, H. (2018, September). Evaluation of an operational real-time irrigation scheduling scheme for drip irrigated citrus fields in Picassent, Spain. *Agricultural Water Management*, 208(30), 465–477. doi:10.1016/j.agwat.2018.06.022
- Li, X., Chen, S., & Guo, L. (2014). Technological innovation of agricultural information service in the age of Big Data. *J. Agric. Sci. Technol*, 16, 10–15.
- Lipper, L., Thornton, P., Shula, R., Tibu, A., & Torquebiau, E. F. (2014). Climate smart agriculture for food security. *Nature Climate Change*, 4(12), 1068–1072. doi:10.1038/nclimate2437
- Lokers, R., Knapen, R., Janssen, S., van Randen, Y., & Jansen, J. (2016). Analysis of Big Data technologies for use in agro-environmental science. *Environmental Modelling & Software*, 84, 494–504. doi:10.1016/j.envsoft.2016.07.017
- López-Lozano, R., Duveiller, G., Seguini, L., Meroni, M., García-Condado, S., Hooker, J., & Baruth, B. et al. (2015). Towards regional grain yield forecasting with 1 km-resolution EO biophysical products: Strengths and limitations at pan-European level. *Agricultural and Forest Meteorology*, 206, 12–32. doi:10.1016/j.agrformet.2015.02.021
- Lvova, L., & Nadporozhskaya, M. (2017). Chemical sensors for soil analysis: principles and applications. In *New Pesticides and Soil Sensors* (pp. 637–678). Academic Press.
- Manyika, J. (2011). *Big Data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

Mark, T. B., Griffin, T. W., & Whitacre, B. E. (2016). The role of wireless broadband connectivity on 'Big Data' and the agricultural industry in the United States and Australia. *The International Food and Agribusiness Management Review*, 19, 43–56.

Martínez-López, B., Perez, A. M., & Sánchez-Vizcaíno, J. M. (2009). Social Network Analysis. Review of General Concepts and Use in Preventive Veterinary Medicine. *Transboundary and Emerging Diseases*, 56(4), 109–120. doi:10.1111/j.1865-1682.2009.01073.x PMID:19341388

Misra, S., & Singh, S. (2012). Localized policy-based target tracking using wireless sensor networks. *ACM Transactions on Sensor Networks*, 8(3), 27. doi:10.1145/2240092.2240101

Moorthy, J., Lahiri, R., Biswas, N., Sanyal, D., Ranjan, J., Nanath, K., & Ghosh, P. (2015). Big Data: Prospects and Challenges. *Vikalpa*, 40(1), 74–96. doi:10.1177/0256090915575450

NEC. (2014). NEC and Dacom collaborate on precision farming solution to maximize yields and reduce costs.

Chmiel, A. M., Sienkiewicz, J., Suchecki, K., & Holyst, J. A. (2007). Networks of companies and branches in Poland. *Physica A*, 383(1), 134–138. doi:10.1016/j.physa.2007.04.095

Soulis, K. X., & Elmaloglou, S. (2018). Optimum soil water content sensors placement for surface drip irrigation scheduling in layered soils. *Computers and Electronics in Agriculture*, 152(September), 1–8.

Orts, E., & Spigonardo, J. (2014). *Sustainability in the age of big data*. IGEL/Wharton. University of PA.

Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences*, 30(4), 431–448.

Perera, C., & Vasilakos, A. V. (2016, October). A knowledge-based resource discovery for Internet of Things. *Knowledge-Based Systems*, 109(1), 122–136. doi:10.1016/j.knosys.2016.06.030

Maes, W. H., & Steppe, K. (2019, February). Perspectives for Remote Sensing with Unmanned Aerial Vehicles in Precision Agriculture. *Trends in Plant Science*, 24(2), 152–164. doi:10.1016/j.tplants.2018.11.007 PMID:30558964

Pivoto, D., Waquil, P. D., Talamini, E., Caroline, P. S. F., Vitor, F. D. C., & Mores, G. V. (2018). Scientific development of smart farming technologies and their application in Brazil. *Information Processing in Agriculture*, 5(1), 21–32. doi:10.1016/j.inpa.2017.12.002

Plume, K. (2014). The Big Data Bounty: U.S. Startups Challenge Agribusiness Giants.

Poppe, K. J., Wolfert, J., Verdouw, C. N., & Renwick, A. (2015). A European perspective on the economics of Big Data. *Farm Policy Journal*, 12, 11–19.

Rice, I. (2018). Improved data visualisation through nonlinear dissimilarity modelling. *Pattern Recognition*, 73(January), 76–88. doi:10.1016/j.patcog.2017.07.016

Rojas, O. (2015). *Protocol for Country-Level ASIS: Calibration and National Adaptation Process*. Food and Agriculture Organization of the United Nations.

Rojas, O., Vrieling, A., & Rembold, F. (2011). Assessing drought probability for agricultural areas in Africa with coarse resolution remote sensing imagery. *Remote Sensing of Environment*, 115(2), 343–352. doi:10.1016/j.rse.2010.09.006

Rosell, J. R., & Sanz, R. (2012). A review of methods and applications of the geometric characterization of tree crops in agricultural activities. *Computers and Electronics in Agriculture*, 81(February), 124–141. doi:10.1016/j.compag.2011.09.007

Sawant, M., Urkude, R., & Jawale, S. (2016). Organized Data and Information for Efficacious Agriculture Using PRIDE™ Model. *The International Food and Agribusiness Management Review*, 19, 115–130.

Schroeder, T. C., & Tonsor, G. T. (2012). International Cattle ID and Traceability: Competitive Implications for the US. *Food Policy*, 31(1), 31–40. doi:10.1016/j.foodpol.2011.10.005

Scott, V. E., Kerriska, K. L., Thompson, P. C., Lyons, N. A., & Garcia, S. C. (2015, January). Voluntary Cow Traffic and Behaviour in the Premilking Yard of a Pasture-Based Automatic Milking System with a Feed Supplementation Regime. *Livestock Science*, 171, 52–63. doi:10.1016/j.livsci.2014.11.002

- Sharma, R., Kamble, S. S., & Gunasekaran, A. (2018). Big GIS analytics framework for agriculture supply chains: A literature review identifying the current trends and future perspectives. *Computers and Electronics in Agriculture*, 155(December), 103–120. doi:10.1016/j.compag.2018.10.001
- Shrivastava, S. & Marshall-Colon, A. (2019). Big Data in Agriculture and Their Analyses. In *Encyclopedia of Food Security and Sustainability* (Vol. 1, pp. 233–237). Academic Press.
- Sonka, S. (2014). Big Data and the Ag sector: more than lots of numbers. *The International Food and Agribusiness Management Review*, 17.
- Sonka, S. (2015). Big Data: From hype to agricultural tool. *Farm Policy Journal*, 12, 1–9.
- Srivastava, N., & Salakhutdinov, R. R. (2012). Multimodal learning with deep Boltzmann machines. *Proceedings of Neural Information Processing Systems Conference (NIPS)* (pp. 2222–2230). Academic Press.
- Sun, Z., Du, K., Zheng, F., & Yin, S. (2013). Perspectives of research and application of Big Data on smart agriculture. *J. Agric. Sci. Technol.*, 15, 63–71.
- Supit, I., Hooijer, A. A., & Van Diepen, C. A. (1994). System Description of the WOFOST 6.0 Crop Simulation Model Implemented in CGMS.
- Talia, D. (2013). Clouds for scalable big data analytics. *Computer*, 46(5), 98–101. doi:10.1109/MC.2013.162
- Godfray, H. C. J., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Nisbett, N., ... & Whiteley, R. (2010). The future of the global food system. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 365, 2769–2777. PMID:20713383
- Tien, J. M. (2013). Big Data: Unleashing information. *Journal of Systems Science and Systems Engineering*, 22(2), 127–151. doi:10.1007/s11518-013-5219-4
- Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011, December 13). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, 108(50), 20260–20264. doi:10.1073/pnas.1116437108 PMID:22106295
- Van Hertem, T., Rooijakkers, L., Berckmans, D., Peña Fernández, A., & Vranken, E. (2017, June). Appropriate data visualisation is key to Precision Livestock Farming acceptance. *Computers and Electronics in Agriculture*, 138(1), 1–10. doi:10.1016/j.compag.2017.04.003
- Van 't Spijker, A. (2014). *The New Oil - Using Innovative Business Models to Turn Data into Profit*. Basking Ridge: Technics Publications.
- Verhoosel, J., van Bekkum, M., Verwaart, T. (2016). HortiCube, a platform for transparent, trusted data sharing in the food supply chain. *Proceedings in Food System Dynamics* (pp. 384–388). Academic Press.
- Welte, J. T., Ault, A. C., Bowman, C., Ellis, S., Buckmaster, D. R., Ess, D., & Krogmeier, J. V. (2013). An Approach to Farm Management Information Systems Using Task-Specific, Collaborative Mobile Apps and Cloud Storage Services.
- Wolfert, J., Sørensen, C. G., & Goense, D. (2014). A Future Internet Collaboration Platform for Safe and Healthy Food from Farm to Fork. *Proceedings of the 2014 Annual SRII Global Conference (SRII)*, San Jose, CA (pp. 266–273). IEEE.
- Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M.-J. (2017). Big Data in Smart Farming – A review. *Agricultural Systems*, 153, 69–80. doi:10.1016/j.agry.2017.01.023
- Wu, X., Zhu, X., Wu, G., & Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, 26(1), 97–107. doi:10.1109/TKDE.2013.109
- Yuan, S., Peng, S., Wang, D., & Man, J. (2018, October). Evaluation of the energy budget and energy use efficiency in wheat production under various crop management practices in China. *Energy*, 160(1), 184–191. doi:10.1016/j.energy.2018.07.006
- Zilberman, D., Lipper, L., McCarthy, N., & Gordon, B. (2018). Innovation in Response to Climate Change. In *Climate Smart Agriculture* (pp. 49–70). Springer. doi:10.1007/978-3-319-61194-5_4

Maya Gopal P.S. is currently pursuing a Doctoral Degree in the School of Computing Science and Engineering, VIT University, Chennai Campus, India. She has more than 15 years of industry, academic, and research experience. She received her M.E from Anna University, Chennai. Her research interests include data mining, machine learning, and data science.