EFFECTIVE USE OF BIG DATA FOR PRECISION AGRICULTURE-BASED SOIL ANALYSIS AND CROP FERTILITY PREDICTION

M.REVATHI

Research scholar, Department of Computer Science and Engineering, Dr. M.G.R. Educational and Research Institute, Chennai, India. Email id: revathimanickam@gmail.com

Dr. N. KANYA

Professor, Department of Information Technology, Dr. M.G.R. Educational and Research Institute, Chennai, India. Email id: kanya.v@drmgrdu.ac.in

Dr. VICTO SUDHA GEORGE

Professor, Department of Computer Science and Engineering, Dr. M.G.R. Educational and Research Institute, Chennai, India. Email id: victosudhageorge@drmgrdu.ac.in

Abstract

In today's era to handle the increasing challenges of agricultural production, we need to understand the complex agricultural ecosystems The usage of modern digital technologies can help in monitoring the physical environment, help in predicting large volumes of data in an unprecedented pace. By using (big) data we can extract the values from it and help the farmers to yield productivity .Big data has been used in various industries, but not yet used in agriculture sector. The main objective of this paper is to perform a review on recent studies and research works in agriculture which employ the current practice of big data analysis, in order to solve various relevant problems. Thirty-four different studies are presented, examining the problem they address, the proposed solution, tools, algorithms and data used, nature and dimensions of big data employed, scale of use as well as overall impact. Concluding, our review highlights the large opportunities of big data analysis in agriculture towards smarter farming, precision agriculture by showing that the availability of hardware and software, techniques and methods for big data analysis, as well as the increasing openness of big data sources, shall encourage more academic research, public sector initiatives and business ventures in the agriculture sector.

Introduction

Increase in population with the developing ecosystem has a great impact on food storage. In the past 40years the world's population has been increased to 3.5 billion to more than 6.2, which created an high demand in food [1] As estimated by (Food and Agriculture Organization of the UN in 2018,atates the increase in global population by 40% until 2045, which implies that the food production must reach to 70% . Due to various aspects like climate change, contamination of water, degradation of soil, sociocultural development (e.g. dietary preference of meat protein), market fluctuations and government policies has created massive uncertainties to food security [2] and to access the safe and nutritious food by all people on the planet. These vulnerabilities challenge farming to further develop efficiency, bringing down simultaneously its ecological impression, which as of now represents the 20% of the anthropogenic Nurseries Gas (GHG) discharges [3].

To fulfil these increasing demands, several studies and investigation has been introduced since the 1995s. Advancements in Global navigation satellite system we can monitor the crop yield

and productivity (e.g. GPS) [5] we can enable the precise localization of field, spatial variability maps, spatial images can be created which has helped in developing a concept known as "precision agriculture" [6]

These days, agriculture practice are being upheld by biotechnology [7] and arising advanced innovations, for example, remote detecting [8], distributed computing , cloud computing , IoT prompting the thought of "smart cultivating" . The development of n (ICT) for precision agriculture help in expanding the production in precision agriculture by enhancing the recent trends and techniques imposed in decision support system.

To tackle the important challenges faced in agriculture sector we have I posed Precision Agriculture to improve the sustainability, food security ,productivity and environmental impact .To enhance the quality and resource of environment we can directly link sustainable agriculture with smart farming

Related study

To employ big data in agriculture we need large investments in infrastructures for data processing[9] which helps in managing the real-time applications (e.g. In weather forecasting, monitoring for crops' pests and animals' diseases). Thus, "big data analysis" is used to portray a new way of agriculture practice [10] which has designed so that farmers can extract economic values from these large volume of agricultural datasets which inherits data from spatial satellite.
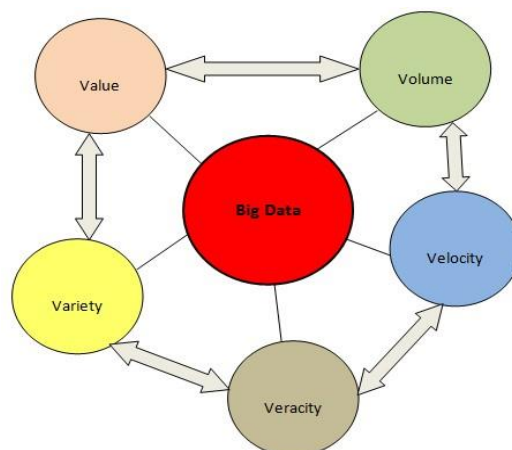
Big data analysis is used in various sectors like banking and finance, in understanding customer behaviour in online shopping for understanding customer need and personalise their needs based on environmental studies. As [11] show, legislative associations utilize Big Data ito upgrade their capacity to serve their residents addressing public difficulties connected with economy, medical care, work creation, cataclysmic events and psychological oppression.

Although big data seems to be very successful and popular in many domains, it has been applied to agriculture in recent days [12] when stakeholders started to see expected success. According to the survey conducted by various agricultural corporations they sated that employing Big Data in agriculture we can Increase the anual global profit of crops up to $30 million dollar.To overcome the challenges of smart farming and sustainable agriculture, as [13, 14] point out, the complex data and to analyse the multivariate and unpredictable data and to provide better understanding of the ecosystem. The development of digital technologies helps in understanding the various aspects of the physical environment, and help in managing large quantities of data in an unprecedented pace. This suggests, as note, the requirement for enormous scope assortment, capacity, pre-handling, demonstrating and investigation of immense measures of information coming from different heterogeneous sources.

The inspiration for setting up this review comes from the way that big data in agribusiness is cutting edge with developing prominence, with recent advancement and application of big data in different areas show its huge potential Current important overviews[15] cover for the most hypothetical parts of this procedure (for example conventional framework, economics, business processes, partners' organization) or spotlight on specific sub-spaces like remote detecting and geospatial examination .Accordingly, the fundamental commitment of this study is that it presents a more engaged outline of the specific issues experienced in horticulture, contrasted with existing reviews, where information investigation is a vital perspective and arrangements are found inside the enormous information domain. Our overview features big data utilized, the strategies and procedures utilized, giving explicit bits of knowledge according to a specialized viewpoint on the potential and chances of enormous information examination, open issues, obstructions and ways of beating them.

Big Data

Big Data" was first presented by Roger Magoulas in 2005 to characterize a large volume of data that are used in conventional application can't deal with and process because of its intricacy and size. It is the assortment of different organized, semi-organized and unstructured informational indexes with enormous volume in Terabytes and complex in nature. A new review says that 80% of the informational indexes are unstructured [16]. The biggest challenge of big data is the   space need to store and manage   unstructured information and to organize them before examination. The big data 5Vs are diagrammatically displayed in the accompanying figure1,

Figure 1: Characteristics of Big data

The Figure 1 shows the Volume, Velocity, Variety, Veracity and Value of the big data. It collect, store, manage and process huge amount of data with speed and time to get proper information to make right decisions.

Volume: It characterizes the data catching that are being created from various information sources in enormous volumes and are in the exceptionally colossal structure generally estimated in Exabyte and Zettabytes. The enormous information begins with least information size as Terabytes [5][6]. This is the amount of information gathered that encased the significant data for its own motivation of assortment to process and examine with their own models.

Velocity: The immense volume of various kinds of information produced from different sources must be process fast and speed. The velocity manages the speed that information gathered from various sources to be assembled and handled rapidly which is absurd with existing conventional applications [17].

Variety: It refers to the different kinds of information captured. This information might be organized as unstructured or semi structured data. Big data can handle unstructured data for example, sensor output, basic text, video, sound, pictures, comments, feedback and documents in different formats [18]. The organized information can be handled fast and effectively when contrasted with different configurations which make more perplexing and hard to process and examine.

Veracity: The data collected from various sources are not perfect and accurate and the data present in large dataset found to be in poor quality and cannot be handled precise, the big data provides ways to work with these types of data [6]. Because of these, establishing the trust is a very big challenge in big data when the number of data points and types increases.

Value: It helps in analysing the data and utilizing the data in the right way [6]. Data can be collected completely based on the duration and availability. The data volume and variety are interconnected with the value of the collected data.

Table 1-Software used in Big Data for Agriculture

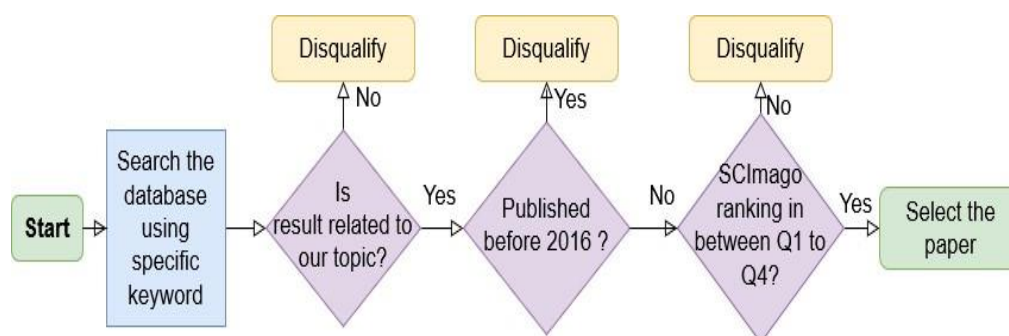| CATEGORY	TOOLS
Image Processing & Statistical Tools	Weka, k, IM& VTK Toolkit
ML TOOLS	R, Weka, Mlpy, Mlpack, Oryn,Mllib
Computation and Analysis using Cloud Platform	MapReduce, Cloudera, Apache Pig, Picotal, Greenplum, Astu SQL Analysis
GIS System	MapInfo,Miramon
Big Database	Give, HadoopDB, Google Bigtable
Modeling, Stimulation and Time series analysis	State, BFAST, Matlab,

Big Data Technologies

The Big data technology is used to store, process, manage the large dataset and to extract data to meet the complexity [7], we cannot process these data with the traditional software, so Big data gather and process huge amount of real time data from various sources with its analysing tools to show the possible predictions in perfect time to reduce the risks and fraudulent acts quick and fast [7]. The Big data technologies are widely classified into two types as per their properties are,

- Operational
- Analytical

The emergence of diverse technologies is shifting traditional farming practices. 5]. was presented after the fast development of innovation over the most recent couple of many years. [19] characterized accuracy horticulture as an aspect of site-explicit yield the executives, which addresses another sort of cultivating the board idea instead of the customary ideas that incorporate perception, checking, and exact estimation of variables in food creation. Huge Information helps the two ranchers, policymakers, and individuals associated with the food creation and store network to oversee direction, organization arranging, and better handling [11]. Many related endeavors incorporate big data stages and applications zeroing in on different issues in the agrarian area, for example, supply chain management, distributed channel.

Methodology

The bibliographic examination of this study incorporates three stages i.e.; (1) gathering important examinations, (2) sifting the works that are straightforwardly connected with the subject of the review, (3) point by point examination of related articles. Figure 2 outlines the course of the writing screening process. In the initial step, a quest for significant works were performed through eminent logical information bases, predominantly IEEE Investigate, Science Direct, and Google Researcher - a web logical ordering administration. The hunt inquiry utilized explicit catchphrases of; "Huge Information" AND ["Precision Agribusiness" OR "Savvy Cultivating" OR "Agriculture"]. To restrict the distribution range, just distributions from 2016 to 2020 were considered for the survey. For better-quality data, the creators sifted the gathered papers as per SC Imago Diary Positioning (SJR pointer). The distributions from diaries bearing a position scope of Q1 to Q4 were considered as great wellsprings of data. The distributions with practically no diary positioning were not thought of.

Figure 2: Literature screening process

After the filtering process, the authors considered only 25 papers. The chosen studies met the authors' requirements and showed the actual relation of Big Data in the field of precision

agriculture. Finally, in the analysis phase, the authors analyzed the chosen papers considering their purposes, methods, platforms, tools used, their fields of application, and impacts achieved to determine the research questions mentioned before.

Table 2-Tools used in Big Data and its Application

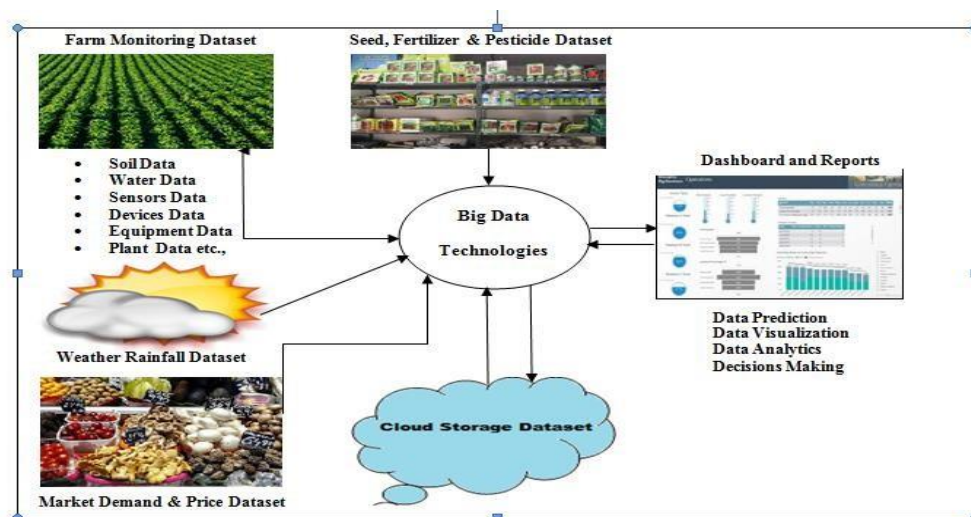| Sl.no	Solution	Area Of Implementation	Metrics	Tools
1.	ML models [27, 28, 30, and 36].	To improve the system storage	Prediction and clustering algorithm, Bayesian Algorithm, support vector machine, k-Nearest algorithm.	• Feature selection • Optimal Classifier
		Wheat, winter wheat, LAI	Artificial Neural Network	X-band Spectrometer
		Spinach, Biomasses, plant height ,soil moisture content	CERES(Crop Environmental Resource Synthesis)	UAV
		Pineapple yield Mapping	Logistic Regression, Sparse	Aerial Photograph
		Nitrogen stress, weed	Artificial Neural Network , support vector machine	Airborne[Hyper spectral]
		Water content	Artificial Neural Network	L- Radiometer
		Sugar vet	Support vector Machine, MapReduce algorithm	Field Spectrometer
		Leafgreenness, Chlorophyll	Support vector Machine	Aerial
2	Sensors[31,34,35]	Thermal images of agriculture land based on location and changing parameter	Predict height and density of soil, characteristic of soil, topological condition	GIS, GPS, Spatial variability of Hydraulic conductivity
		Predicting weather in agriculture field	For predicting the input in the farm to transit analysis	Unnamed Aerial surveillance
		Combing Big data input from various parameters like weather prediction Report various input are gathered from different sensors used in ecosystem	Predict the temperature level, soil moisture	Supervised and non-supervised learning model
3	Weed Detection[19,21]	Image Processing	Detecting the condition of weed	Logistic Regression
4	Weather Prediction using Big Data[15,16]	Accurate prediction of weather required for farming	Farming decision support to increase the efficiency and utilize the resource management	MapReduce, Linear Regression

Role of Big Data in Agriculture

These days' different big data innovations help to upgrade and get the farming system which safeguards plant development and liberated from plant killing sicknesses. The Enormous

information upholds the plant producers and ranchers from limited scope to huge scope cultivating by giving proper forecasts to accept choices according to the constant information analysis[26]. After the web-based entertainment blasting on the planet, information has filled in an unpredicted way practically in all industry. Consistently the information is being made with various sources and gathered in servers, cloud and put away for some time later. The Large information advances having the office to gather gigantic volume of datasets, stockpiling, access, break down with various calculations and handle terabytes of information effectively to address the improvement of horticultural cycle and business. It is an exceptionally basic issue to deal with the horticultural activity with ranchers to make it fruitful in their yield and creation to fulfill the market need

It helps to improve the crop yields, soil information, fertilizers, pesticides, rain water storage and production forecasting to optimize and plan for their labourers, agri equipment's with respect to the size of the farming area. The agricultural operation data, agri business data, social media, weather data, sensors, smart devices, machineries and tractors data are huge volume while collecting and very difficult to manage. Big data can collect, store all the data from different sources and visualize the forecasting to make appropriate decision making [28]. The soil, climatic conditions, seeds, plants, soil supplements, water irrigation and the yield are the major things which influence the profitability of the farming business. These technologies facilitate the farmers in a safe way to analyse a wide range of data sources for better decisions to avoid huge loss. The figure 3 shows the role and the other influencing criteria of big data in agriculture,

Figure 3: Bigdata in Agriculture

In the above figure, the market demand and the price data set helps the farmers to know which to cultivate during the appropriate season. The availability of seeds, fertilizers and the pesticides are also needed for planning of the procurement to protect crop health [29]. The dashboards, reports and the other analytical information are the outcome that farmer can use, make decisions and utilize for farm operation [30]. Big data technologies provides interactive

graphs, dashboards and charts in easy understandable manner to make proper insights of their farming requirements.

Table 3-Source of BigData Techniques used in Agriculture

Agricultural Area	Big Data sources	Technique	Reference
Climate Condition[21]	Weather stations, Remote sensing, static historical data, climate & environmental data, spatial data	ML, MapReduce, Support vector Machine, GIS, Geospatial data	Fuchus and wolf [2016].
Bio diversity[26]	SER Dataset for wildlife species, historical information about soil, animals, microorganisms and ground sensors	Bayesian Believe Network, ML(decision Tree),k-nearest	Sakemato [2011]
Remote Sensing[18]	Historical data which contain data from satellite, drone, web based data GIS geospatial data.	Cloud, GIS geospatial data, image processing, IOT(Internet of things), decision support system	Karmas[2014]
Crop [24]	Ground sensors to predict electricity, moisture level , digital image using drone and satellite	Mo, Fourier transform, wavelet filtering	Meyer [2009]
Weeds[20]	Ground sensors, surveys and database of statistical data	Mo, Logistic, regression, neural network	Gutierrez [2008]
Security & food availability[25]	Earth land images (datasets), web-based data, WMO datasets	Iot, modelling stimulation, image processing, farthest first clustering	Urbibiera[2016]
Soil moisture level[23]	Remote sensing, static historical information, GIS, web-based data	MapReduce, mobile applicants, AI, decision support system	Beeha [2018[

Conclusion

Therefore, there is a biggest need for future research to meet the food demand and maintaining a good healthy agricultural environment around the globe. The farmers, agriculture equipment manufacturers, government, research universities, weather stations, technological companies, research scholars, young innovators and other related bodies can be joined as team to solve the critical issues in agriculture by using Big Data technologies. In this survey, the big data technologies, roles and challenges of big data in agriculture are discussed briefly and the future need of research are also suggested as per the growing demands.

References

1. The Food and Agriculture Organization of the United Nations Homepage, Available online: http://www.fao.org/plant-health-2020 (accessed on May 18, 2020).

2. Daniel Frona, Janos Szenderak and Monika HarangiRakos, University of Debrecen, 4032 Debrecen, Hungary. "The Challenge of Feeding the World", Published in MDPI Journal , 2019, 11, 5816; doi:10.3390/su11205816

3. Sjaak Wolfert, Lan Ge, Cor Verdouw and Marc - Jeroen Bogaardt, "Big Data in Smart Farming - A review", published in ScienceDirect Agricultural Systems, Vol 153, May 2017, Pages: 69-80. https://doi.org/10.1016/j.agsy.2017.01.023

4. Miguel A Zamora-lzquierdo, Jose Santa, Juan A Martinez, Vicente Martinez and Antonio F.Skarmeta, " Smart farming IoT platform based on edge and cloud computing", published in Elsevier, BioSystem Engineering, Vol:177, Jan 2019, Pages:4-17. https://doi.org/10.1016/j.biosystemseng.2018.10.014

5. Saeid Sadeghi, Darvazeh, Iman Raeesi Vanani and Farzaneh Mansouri Musolu, "Big Data Analytics and Its Applications in Supply Chain Management", published in "New Trends in the Use of Artificial Intelligence for the Industry 4.0", Mar 2020, DOI:10.5772/intechopen.89426

6. Reihaneh H. Hariri, Erik M. Fredericks and Kate M. Bowers," Uncertainity in big data analytics: survey, opportunities and challenges", published in Springer Open, Journal of Big data, Vol 6, Article no:44(2019).

7. Ahmed Oussous, Fatima -Zahra Benjelloun, Ayoub Ait Lahcen and Samir Belfkih, "Big Data Technologies - A Survey", published in Journal of King Saud University - Computer and Information Sciences. Vol: 30, Issue:4, October 2018, Pages: 431-448. https://doi.org/10.1016/j.jksuci.2017.06.001

8. Orange-Rogla, Sergio, Chalmeta and Ricardo, "Framework for implementing a big data ecosystem in organizations" published in Communications of the ACM Journal, 2019, 62(1).

9. Nawsher Khan, Ibrar Yaqoob, Ibrahim Abaker Targio Hashem, Zakira Inayat, Waleed Kamaleldin Mahmoud Ali, Muhammad Alam, Muhammad Shiraz and Abdullah Gani," Big Data: Survey, Technologies, Opportunities and Challenges", published in Hindawi The Scientific world journal, Vol:2014, Article Id:712826, https://doi.org/10.1155/2014/712826.

10. Anurag Agrahari, Prof D.T.V. Dharmaji Rao, "A Review paper on Big data: Technologies, tools and Trends",published in International Research Journal of Engineering and Technology Vol: 04 Issue: 10, 2017.

11. Ifeyinwa Angela Ajah and Henry Friday Nweke, " A Review on Big Data and Business Analytics: Trends, Platforms, Success Factors and Applications", published in MDPI big data and cognitive computing journal 2019, Vol:3,32; doi:10.3390/bdcc3020032

12. PAT Research, Dark data analytics Inc, Canada company website, www.predictiveanalyticstoday.com/teradata rainstor-analytics-with-archived-data (May 21, 2020).

13. Toby Wolpe, Big data Analytics, Article published on 2013, Available online: https://www.zdnet.com/article/splunks-big-data-hunkgives-hadoop-muscle-to-non-techies(May 29, 2020)

14. Neelam Tyagi, Big Data, Article published on Mar 2020 as Top 10 Big data Technologies in 2020, Available online: https://www.analyticssteps.com/blogs/top-10-bigdata-technologies-2020 (accessed May 29,2020)

15. Varsha C. Pande and Dr. Abha S.Khandelwal, "Clustering And Classification Evaluation Using Rapid Miner", International Journal of Emerging Technologies and Innovative Research,Vol.5, Issue 9, page no.936-939, Sept 2018, http://www.jetir.org/papers/JETIR1809788.pdf

16. The Presto Homepage "Distributed Query Engine for Big data", Available online: https://prestodb.io (accessed on Jun 18, 2020)

17. Karmas, Han Wu, Shang Zhihao and KatinkaWolter,"Performance Prediction for the Apache Kafka Messaging System", Published in IEEE 5th International Conference on Data Science and Systems on Aug 2019, DOI: 10.1109/HPCC/SmartCity/DSS.2019.00036

18. The KNIME Homepage, https://www.knime.com (accessed on Jun 18, 2020)

19. Gutierrez. Madhavi Vaidya, "Big Data Storage Mechanisms and Survey of MapReduce Paradigms", Article published on Apr 2020, Available online:https://analyticsindiamag.com /big-data-storage-mechanisms-and-survey-of-mapreduceparadigms/

20. Vlod Krotov, Fuchus and wolf ,"A Quick Introduction to R and R Studio", technical report published in Research Gate on Nov 2017, DOI: 10.13140/RG.2.2.10401.92009

21. The Apache Website," An Advanced unified Programming Model", https://beam.apache.org (accessed on Jun 22 ,2020) The Airflow Apache Website, https://airflow.apache.org/docs/stable (accessed on Jul 01, 2020)

22. Beena, Evert FK, Fountas S, Jakovetic D, Crnojevic V, Travlos I and Kempenaar C. (2017). "Big Data for weed control and crop protection" published in Weed Research, https://doi.org/10.1111/wre.12255

23. Meyer, E. and Rutschmann, E., 2018. Big data analytics and demand forecasting in supply chains: a conceptual analysis... The International Journal of Logistics Management, Volume 29, pp. 739-766.

24. Desamparados Blazquez and Josep Domenech, "Big Data sources and methods for social and economic analyses", published in Elsevier Journal Technical forecasting and social change, 2018, Vol: 130, Page: 99-113. https://doi.org/10.1016/j.techfore.2017.07.027

25. Urbibieria, J.U. Kumar, B.K. Kumar, C. SekharBig data analytics on indian crop planning to increase agricultural production Adv Sci Technol Lett, Vol:147 (2018), pp. 211-216

26. The Talend Website, Topic: Big data and Agriculture https://www.talend.com/resources/big-data-agriculture

27. Andres Villa, Henriksen, Gareth T.C. Edwards Lissa A. Pesonen, Ole Green and Claus Aage Gron Sorensena, "Internet of Things in arable farming: Implementation, applications, challenges and potential", published in Biosystem Engineering on Mar 2020, Vol: 191, P: 60-84.

28. Veronica Saiz-Rubio and Francisco Rovira-Mas, "Review on From Smart Farming towards Agriculture 5.0: A Review on Crop Data Management", published on MDPI Agronomy journal, Feb 2020, Vol: 10, 207. https://doi.org/10.3390/agronomy10020207

29. Mekruksavanich and T. Cheosuwan, "Visual Big Data Analytics for Sustainable Agricultural Development," in 2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing, iSAI-NLP 2018 - Proceedings, 2018, pp. 1–5, doi: 10.1109/iSAINLP.2018.8692910.

30. J. Zhao and J. Guo, "Big data analysis technology application in agricultural intelligence decision system," in 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), 2018, pp. 209–212, doi: 10.1109/ICCCBDA.2018.8386513.

31. J. Majumdar, S. Naraseeyappa, and S. Ankalaki, "Analysis of agriculture data using data mining techniques: application of big data," J. Big Data, vol. 4, no. 1, p. 20, 2017, doi: 10.1186/s40537-017-0077-4.

32. S. Rajeswari, K. Suthendran, and K. Rajakumar, "A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics," in 2017 International Conference on Intelligent Computing and Control (I2C2), 2017, pp. 1–5, doi: 10.1109/I2C2.2017.8321902.

33. T. Řezník et al., "Disaster risk reduction in agriculture through geospatial (Big) data processing," ISPRS Int. J. Geo-Information, vol. 6, no. 8, p. 238, 2017, doi: 10.3390/ijgi6080238.

34. K. Charvat et al., "Advanced Visualisation of Big Data for Agriculture as Part of Databio Development," in IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, 2018, pp. 415–418, doi: 10.1109/IGARSS.2018.8517556.

35. L. J. Young, M. Hyman, and B. R. Rater, "Exploring a Big Data Approach to Building a List Frame for Urban Agriculture: A Pilot Study in the City of Baltimore," J. Off. Stat., vol. 34, no. 2, pp. 323–340, 2018, doi: 10.2478/jos-2018-0015.

36. S. Parvin et al., "Smart Food Security System Using IoT and Big Data Analytics," in 16th International Conference on Information Technology-New Generations (ITNG 2019), 2019, pp. 253–258, doi: 10.1007/978-3-030-14070-0_35.