

Digitalization and Big Data in Smart Farming – Bibliometric and Systemic Analysis

Jaqueline IAKSCH¹, Ederson FERNANDES and Milton BORSATO
Universidade Tecnológica Federal do Paraná (UTFPR)

Abstract. Agriculture has always had a great significance in the civilization development. However, modern agriculture is facing increasing challenges due to population growth and environmental degradation. Commercially, farmers are looking for ways to improve profitability and agricultural efficiency to reduce costs. Smart Farming is enabling the use of detailed digital information to guide decisions along the agricultural value chain. Thus, better decisions and efficient management control are required through generated information and knowledge at any farm. New technologies and solutions have been applied to provide alternatives to assist in information gathering and processing, and thereby contribute to increased agricultural productivity. Therefore, this article aims to gain state-of-art insight and identify proposed solutions, trends and unfilled gaps regarding digitalization and Big Data applications in Smart Farming, through a literature review. The current study accomplished these goals through analyses based on ProKnow-C (Knowledge Development Process - Constructivist) methodology. A total of 2401 articles were found. Then, a quantitative analysis identified the most relevant ones among a total of 39 articles were included in a bibliometric and text mining analysis, which was performed to identify the most relevant journals and authors that stand out in the research area. A systemic analysis was also accomplished from these articles. Finally, research problems, solutions, opportunities, and new trends to be explored were identified.

Keywords. Agriculture, smart farming, intelligent systems, smart technologies

Introduction

The agricultural sector has been improved with the arrival of the fourth industrial revolution [1]. New technologies and solutions have been applied to provide alternatives to assist in the collection and processing of information, contributing to the increase in agricultural productivity [2].

In recent years, technological development has introduced radical changes in the agricultural working environment (i.e., the use of electronic systems and data transmission). Besides that, these changes required the provision of updated information on systems, equipment, markets, and agents involved in production for the strategic and managerial decision-making processes [3].

¹ Corresponding Author, Mail: jaquelineiaksch@alunos.utfpr.edu.br.

Agriculture 4.0, as well as Industry 4.0, represents the combination and interaction of internal and external agricultural operations, enabling the use of detailed digital information to guide decisions along the agricultural value chain [4]. Also, in this context, the concept of Farm 4.0, or smart farm, emerges in the search for ensuring that a property is endowed with equipment, processes and people targeted to digital agriculture [5].

In addition to the introduction of new practices and tools, Agriculture 4.0 has as one of its main objectives the adaptability of production systems and the improvement of efficiency [6]. The latter occurs through increased productivity, which is the ability to collect, use and exchange data remotely in real-time [7]. Thus, automation and intelligent decision making become very important for the fulfillment of these objectives [8].

The development of information and communication technologies, IoT(internet of things), cloud computing and Big Data, is leveraging the development of smart agriculture [9]. Through the combination of technologies that make use of cloud computing and IoT, digital agriculture is evolving, as it is possible to take advantage of the immense amount of agricultural data generated by the operations carried out on the farms [10].

Big data technologies are being used to provide predictive information in agricultural operations, allowing real-time operational decisions to be made and redesigning business processes. These technologies play an essential role in the development of a scenario in which humans are only involved in high-level intelligence analysis and action planning [9].

Given these technologies that encompass smart agriculture and its applications, this article aims to verify the knowledge and previous studies on the subject. Thus, a bibliometric and systemic analysis based on ProKnow-C (Knowledge Development Process-Constructivist) was conducted to find out about the work already done, methodologies, technologies and tools used, and to identify gaps in the literature on the subject in question.

Section II presents the methodological aspects of the research, followed by the presentation of the sequence of activities to carry out the selection of the bibliographic portfolio in section III, as well as the bibliometric analysis and the systemic analysis in sections IV and V, respectively. Section VI presents the final considerations.

1. Methodological Aspects

Proknow-C was adopted as a methodological procedure. This instrument defines a flowchart for bibliometric review and is designed to assist researchers in managing information and knowledge in relevant content on a specific scientific issue [11].

The structured literature review process is a fundamental tool, since currently there is a huge availability of information, and which is used to manage the diversity of knowledge in academic research. It is possible to map and evaluate existing knowledge and define a research question to further develop existing knowledge when carrying out a structured literature review [12].

ProKnow-C consists of the following steps: selection of the bibliographic portfolio, bibliometric analysis, systemic analysis of the portfolio, and definition of the research question and objective [13]. These steps are shown in Figure 1.

In selection of the bibliographic portfolio step, the research axes and keywords are defined, as well as the databases and, subsequently, the researches. The bibliometric analysis step defines which authors, articles, journals, and keywords are relevant to the

research question. Finally, the systemic analysis interprets the articles in the bibliographic portfolio.



Figure 1. ProKnow-C steps.

This article will address only the first three steps of ProKnow-C, as this study aims to present the state of the art and research opportunities based on the systemic analysis step. To achieve this result, the software *EndNote*, *Mendeley*, and *Microsoft Excel* were used. *EndNote* and *Mendeley* supported bibliography management while the data tabulation was performed in *Excel*.

2. Bibliographic portfolio selection

The bibliographic portfolio selection aimed to gather publications with relevant content and scientific recognition related to the research topic. Figure 2 represents the steps for this selection.

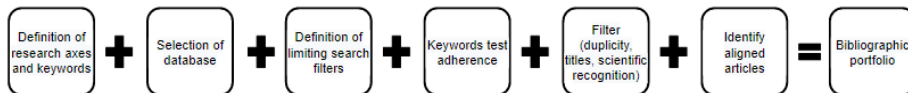


Figure 2. Steps for selection of bibliographic portfolio.

2.1. Research axes and keyword selection

The keywords used were divided into three categories (research axes): (i) Agriculture; (ii) Digitalization; (iii) Data Analysis.

Afterward, the keywords were combined into a search sentence with the booleans: ((“farming” OR “agriculture” OR “agrobusiness” OR “crop” OR “harvest”) AND (“digitalization” OR “connectivity” OR “CPS” OR “cyber-physical system” OR “iot” OR “internet of things” OR “data acquisition”) AND (“autonomous monitoring” OR “autonomous decision” OR “decision making” OR “knowledge-based” OR “data mining” OR “data analytics” OR “data modeling” OR “data processing” OR “data analysis”)).

2.2. Selection of database

In order to have access to a wide variety of academic and conference publications, research databases were chosen considering their connection with the engineering area, as well as their availability in the CAPES (Higher Education Personnel Improvement Coordination) journals portal. The bases chosen were: Engineering Village, Scopus, ProQuest, Springer, Emerald, Web of Science, Willey, Science Direct, EBSCO, IEEE, and World Scientific.

In the Emerald, Willey, Science Direct, IEEE and World Scientific databases no articles compatible with the topic were found, therefore, these databases were excluded from the process.

Table 1 Research Axes and their Keywords.

Research Axes	Keywords
Agriculture	<i>farming</i>
	<i>agriculture</i>
	<i>agrobusiness</i>
	<i>crop</i>
	<i>harvest</i>
Digitalization	<i>digitalization</i>
	<i>connectivity</i>
	<i>CPS</i>
	<i>cyber physical system</i>
	<i>iot</i>
	<i>internet of things</i>
Data analysis	<i>data acquisition</i>
	<i>autonomous monitoring</i>
	<i>autonomous decision</i>
	<i>decision making</i>
	<i>knowledge based</i>
	<i>data mining</i>
	<i>data analytics</i>
	<i>data modelling</i>
	<i>data processing</i>
	<i>data analysis</i>

2.3. Collection of articles

The initial search resulted in a total of 2401 publications. Only papers from congresses and journals published in the last five years (2015 to 2019) were considered. Also, adherence to the keywords was accomplished by reading the titles of the papers. Three articles were identified related to the research topic and it was found that the keywords were also related to those previously defined.

2.4. Filtering

In this step, all references were exported to *EndNote X7* and *Mendeley* software, as these tools assist to handle this big amount of references. To further refine the results, all duplicate references have been removed, resulting in 2260 papers.

Despite the pre-selection carried out directly in the databases, the presence of 13 references from previous years to those considered in this study, as well as 93 references from other types of references and documents from other areas that were not of interest (Medicine, Biology, etc.), leaving 2154 references.

Thereafter, an individual analysis of each paper was conducted. Each of the 2154 titles in the portfolio was read, and only 183 were related to the research. Then, the number of citations for each paper was verified to identify the scientific recognition of the papers. The criteria for selecting articles for the next stage of the analysis was that it

should be cited at least once. Papers published last year with related titles that were not cited were also considered. Thus, 153 references remained.

The last step started with reading and evaluating the abstracts, to select the papers that were in fact related with the research. Thus, a total of 39 papers were selected for the final review. These steps can be analyzed in Figure 3.

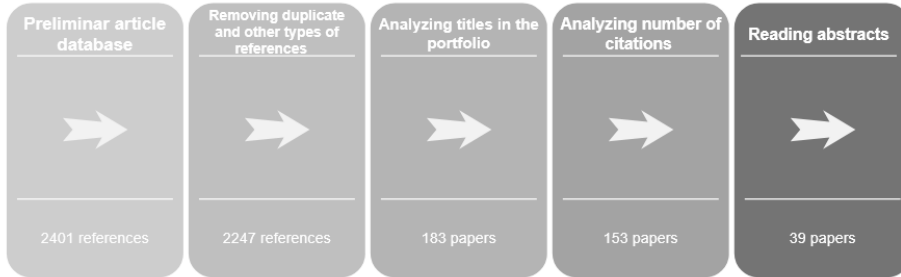


Figure 3. Steps followed to select bibliographic portfolio

3. Bibliometric analysis of bibliographic portfolio

Bibliometric Analysis allows to evaluate and interpret the bibliographic portfolio through statistical analysis, in order to generate knowledge of the bibliographic portfolio characteristics [13].

First, the results found for each database were analyzed (Table 2). In a total of 2401 papers, in Springer was found the large majority of articles, followed by Scopus, Web of Science, Engineering Village, ProQuest, and EBSCO. Databases as Wiley, Emerald, Science Direct, IEEE, and World Scientific didn't return any results.

Table 2. Returned articles per database researched.

Database	Number of articles returned
Springer	1370
Scopus	467
Web of Science	206
Engineering Village (COMPENDEX)	180
ProQuest	153
EBSCO	25
Wiley	0
Emerald	0
Science Direct	0
IEEE	0
World Scientific	0
Total	2401

It is possible to identify the year of papers' publication, the relevance of journals and congresses on the research topic, as well as the authors who stand out most in the research area of the portfolio.

The total number of articles in the portfolio published per year is shown in Figure 4. When analyzing the year of publication, it can be identified that the year 2018 has more publications, in addition to nine papers already published in the year 2019, so it is said that the number of articles published annually is growing, and this topic remains prominent.

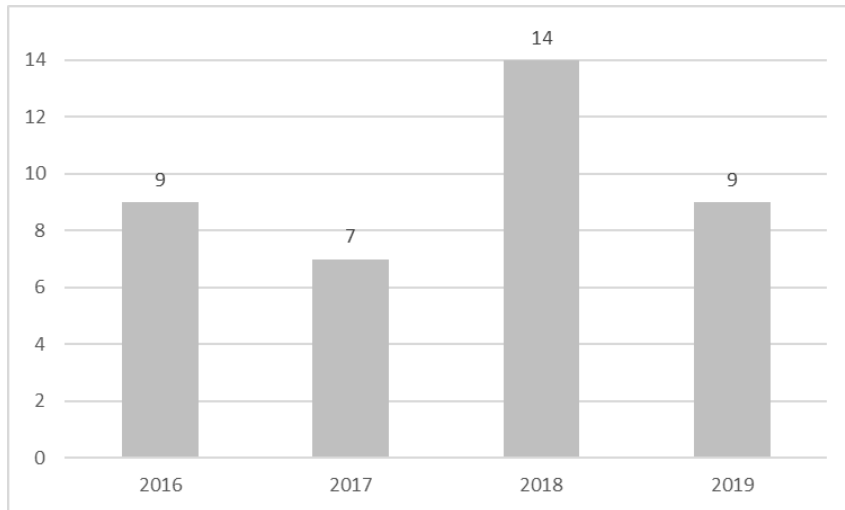


Figure 4. Number of articles published per year.

Continuing the analysis, with the exception of the authors Olakunle Elijah, Joseph Walsh, Sahitya Roy, Tharek Abdul Rahman, and Igbafe Orikumhi, who have two publications in the bibliographic portfolio, most of the publications were written by different authors.

Table 3 shows the ten most cited articles for this research. The most cited paper, [9], conducted a literature review on Big Data applications in Smart Farming and identified the related socio-economic challenges. The authors have also developed a conceptual framework for analyzing the topic that can be used for future studies.

The work of [14], [15] proposed intelligent monitoring systems to control environmental factors, such as humidity and temperature, and weather forecasting, using android, IoT and big data platforms. [16] in their article, proposed a semantic framework for smart agriculture applications based on IoT, which supports reasoning about many data flows from heterogeneous real-time sensors .

Table 3. Most cited articles¹.

Title	Citation	Year
Big Data in Smart Farming - A review	391	2017
Publicising Food: Big Data, Precision Agriculture, and Co-Experimental Techniques of Addition	65	2017
Intelligent Agriculture Greenhouse Environment Monitoring System Based on IOT Technology	65	2016
Internet of things: from internet scale sensing to smart services	57	2016
Big data in precision agriculture: Weather forecasting for future farming	54	2016
Agri-IoT: A Semantic Framework for Internet of Things-enabled Smart Farming Applications	52	2016
An Overview of Internet of Things (IoT) and Data Analytics in Agriculture: Benefits and Challenges	44	2018
Mobile Cloud Business Process Management System for the internet of things: A survey	40	2016
Scientific development of smart farming technologies and their application in Brazil	33	2018
How data analytics is transforming agriculture	29	2018

¹ October, 2019

After this step, systemic analysis stage started, which is presented in the next section.

4. Systemic Analysis and Review Discussion

The systemic analysis stage analyzed the content of the articles according to pre-established criteria. To complete this step, it was necessary to download the full texts, as only the citations were imported into *EndNote* and *Mendeley*.

To conduct the systemic analysis, each article was read in order to define the following aspects: (i) objective, (ii) methodology, (iii) main results, (iv) future recommendations, (v) research opportunities pointed out by the authors; and (vi) research opportunities found through critical analysis. This information was organized in an Excel spreadsheet and data were tabulated to facilitate global analysis.

When analyzing the articles, most of the tools and systems presented use IoT and big data to address issues involving aspects such as irrigation, pest control, crop growth, and weather forecasting. Some works like the [17] developed applications to support the interaction between IoT devices from different suppliers, seeking to recommend plants to increase agricultural productivity in a given location.

Several authors have tried to understand how the concepts of smart agriculture are being applied and what technologies are involved. Some works proposed solutions based on the development of IoT platforms, such as the study [18] that integrated IoT in agriculture. The authors considered a sugarcane plantation as an object of study and compared optimal parameters, regarding soil and environmental factors that should be considered to guarantee the maximum yield of the planting, with the captured data.

Other authors have created solutions based on the study of environmental factors to improve crop yields. [19], [20], and [21] implemented IoT technology to capture data for crop and soil monitoring, climate, and air monitoring, as well as machine work monitoring. According to [22], the majority of research involving IoT applications is focused on water management on farms, due to a lack of abundance.

[16] created a customizable online platform, based on IoT, that allowed large-scale data processing, analysis, and interpretation of data from different sources, such as sensory systems, surveillance cameras, climate and information, warnings, and alerts from government organizations. Although the platform seeks to integrate data from different sources, it is not flexible and adaptable to all agricultural scenarios and deals only with data from environmental factors.

The system proposed in [23] is based on IoT and is supported by a mobile interface, and seeks to create a connected agricultural network to share knowledge among farmers. The model proposed in [24] also uses the IoT approach for data acquisition via sensors, task control, and data management and analysis that are considered in the development of its system to help farmers manage aspects related to environmental factors and distribution information. [25] created a platform using CPS and IoT, capable of handling the needs of soilless culture in full recirculation greenhouses, using moderately saline water.

Although there are studies for the development of IoT platforms aimed at agriculture, they are specific to certain cultures and information. Thus, it is necessary to create generic platforms that can support any type of farm, regardless of their cultivation, being easily customizable and free from geographical restrictions [21]. According to the authors, it is necessary to develop IoT devices that take into account algorithms with advanced encryption, to increase the security of the captured data.

Big data technologies can be considered as the solution for various applications and can be used in decision making and to extract new ideas and knowledge for agriculture. The main challenges are to discover knowledge and correlations of historical records and

deal with unstructured data [15]. In their research, [15] aimed, through the use of different data sets, to improve the accuracy of rain forecasts, using different parameters.

Big data applications in agriculture are not strictly related to primary production but play an important role in improving the efficiency of the entire supply chain to minimize production costs [9]. [26] used predictive big data analysis techniques to analyze and predict harvest yield and decide the best harvest sequence based on yield information from previous crops on the same land, current information on soil nutrients, as well as predict the cost involved with fertilizers.

Farmers are interested in business models that support the generation of revenue from data captured using IoT technologies [27]. From the data collected and generated by the various communication devices and technologies, it is possible to build knowledge bases that store complex and unstructured information. However, having a model that has information in an organized and complete way, that provides relevant knowledge, that can be used universally and that actually implements smart agriculture, is still a challenge [28], [29].

Through reading the articles and the critical analysis carried out, some research opportunities were achieved. Five of them can be highlighted:

- Farm 4.0 characterization, with all technologies and information flows presented;
- Generic IoT platforms for agriculture creation;
- Development of IoT devices that take into account algorithms with advanced encryption;
- Development of a universal information model to implement smart agriculture;
- Event prediction considering all factors involved in the agricultural scenario.

5. Conclusion

The present work developed a structured review of the literature on smart farming, aiming to identify research trends and opportunities related to smart farming. The ProKnow-C procedure was applied with the objective of raising bibliographic and systemic analyzes of the proposed theme. From an initial survey of 2401 publications presented between 2015 and 2019, a portfolio of 39 articles was identified as representative for the given analysis.

Through bibliometric analysis, the portfolio was interpreted and evaluated. It was possible to identify a growing interest in the subject. In the systemic analysis, all papers were read in order to identify the main research problems addressed, the proposed objectives and resources, and future research opportunities.

This methodological procedure proved to be adequate for the literature review, however, the correct choice of keywords used in the searches is extremely important for the correct representation of the research topic. Besides, when using ProKnow-C it is necessary to record each step performed, which makes the process difficult. It is recommended in the future to carry out all the steps again, including the analysis of the references of the papers that make up the bibliographic portfolio and ensuring the analysis not only of the most recent articles.

Acknowledgement

The authors wish to thank CNPq - Conselho Nacional de Pesquisa e Desenvolvimento Tecnológico do Brasil (call 23/2018 – DAI) for supporting the present research (grant 157258/2019-0).

References

- [1] B. Ozdogan, A. Gacar, H. Aktas, Digital agriculture practices in the context of agriculture 4.0, *J. Econ. Financ. Account.*, 2017, vol. 4, pp. 184–191.
- [2] P. P. Ray, Internet of things for smart agriculture: Technologies, practices and future direction, *J. Ambient Intell. Smart Environ.*, 2017, Vol. 9, No 4, pp. 395–420.
- [3] D. Pivoto, P. D. Waquil, E. Talamini, C. P. S. Finocchio, V. F. Dalla Corte, e G. de Vargas Mores, Scientific development of smart farming technologies and their application in Brazil, *Inf. Process. Agric.*, 2018, Vol. 5, No 1, pp. 21–32.
- [4] I. Zambon, M. Cecchini, G. Egidi, M. G. Saporito, A. Colantoni, Revolution 4.0: Industry vs. agriculture in a future development for SMEs, *Processes*, 2019, Vol. 7, No. 1, p. 36.
- [5] M. J. O’Grady e G. M. P. O’Hare, Modelling the smart farm, *Inf. Process. Agric.*, 2017, Vol. 4, No 3, pp. 179–187.
- [6] D. Andritoiu, L. Bazavan, F. Besnea, H. Roibu, N. Bizdoaca, Agriculture autonomous monitoring and decisional mechatronic system, in *2018 19th International Carpathian Control Conference (ICCC)*, 2018, p. 241–246.
- [7] V. Bonneau, B. Copigneaux, L. Probst, B. Pedersen, *Industry 4.0 in Agriculture: Focus on IoT Aspects*, European Commission, 2017.
- [8] D. Yane, “Research and Analysis about System of Digital Agriculture Based on a Network Platform”, in *International Conference on Computer and Computing Technologies in Agriculture*, 2010, p. 274–282.
- [9] S. Wolfert, L. Ge, C. Verdouw, e M.-J. Bogaardt, “Big Data in Smart Farming – A review”, *Agric. Syst.*, vol. 153, p. 69–80, 2017.
- [10] S. Rotz et al., “The Politics of Digital Agricultural Technologies: A Preliminary Review”, *Sociol. Ruralis*, vol. 59, n° 2, p. 203–229, 2019.
- [11] J. E. Tasca, L. Ensslin, S. R. Ensslin, M.B.M. Alves, An approach for selecting a theoretical framework for the evaluation of training programs, *J. Eur. Ind. Train.*, 2010, Vol. 34, No. 7, 631–655.
- [12] D. Tranfield, D. Denyer, e P. Smart, Towards a methodology for developing evidence-informed management knowledge by means of systematic review, *Br. J. Manag.*, 2003, Vol. 14, No 3, pp. 207–222.
- [13] L. Ensslin, S. R. Ensslin, R. T. de O. Lacerda, J. E. Tasca, ProKnow-C, knowledge development process-constructivist, *Process. técnico com Pat. Regist. pendente junto ao INPI. Bras.*, 2010, vol. 10, n° 4, p. 2015.
- [14] L. Dan, W. Hongli, Z. Mengya, e X. Jianqiu, Intelligent Agriculture Greenhouse Environment Monitoring System Based on the Android Platform, in *2017 International Conference on Smart Grid and Electrical Automation (ICSGEA)*, 2017, p. 358–361.
- [15] M. R. Bendre, R. C. Thool, e V. R. Thool, Big data in precision agriculture: Weather forecasting for future farming, in *Proceedings on 2015 1st International Conference on Next Generation Computing Technologies, NGCT 2015*, 2016, p. 744–750.
- [16] A. Kamilaris, F. Gao, F. X. Prenafeta-Boldu, e M. I. Ali, Agri-IoT: A semantic framework for Internet of Things-enabled smart farming applications, in *2016 IEEE 3rd World Forum on Internet of Things, WF-IoT 2016*, 2017, p. 442–447.
- [17] D. Georgakopoulos e P. P. Jayaraman, Internet of things: from internet scale sensing to smart services, *Computing*, 2016, Vol. 98, No 10, pp. 1041–1058.
- [18] H. V. Abhijith, D. A. Jain, U. Adithya Athreya Rao, Intelligent agriculture mechanism using internet of things, in *2017 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2017*, 2017, vol. 2017-Janua, pp. 2185–2188.
- [19] C. Dewi R.-C. Chen, Decision making based on IoT data collection for precision agriculture, *Studies in Computational Intelligence*, 2020, Vol. 830. pp. 31–42.
- [20] Z. Chen et al., “Progress and perspectives on agricultural remote sensing research and applications in China”, *Yaogan Xuebao/Journal Remote Sens.*, 2016, Vol. 20, No 5, pp. 748–767.

- [21] O. Elijah, I. Orikumhi, T. A. Rahman, S. A. Babale, S. I. Orakwue, Enabling smart agriculture in Nigeria: Application of IoT and data analytics, in *2017 IEEE 3rd International Conference on Electro-Technology for National Development, NIGERCON 2017*, 2018, vol. 2018-Janua, pp. 762–766.
- [22] A.A.R. Madushanki, M.N. Halgamuge, W. A. H. S. Wirasagoda, A. Syed, Adoption of the Internet of Things (IoT) in agriculture and smart farming towards urban greening: A review, *Int. J. Adv. Comput. Sci. Appl.*, 2019, Vol. 10, No 4, pp. 11–28.
- [23] S. Roy et al., IoT, big data science & analytics, cloud computing and mobile app based hybrid system for smart agriculture, in *2017 8th Industrial Automation and Electromechanical Engineering Conference, IEMECON 2017*, 2017, pp. 303–304.
- [24] S. Suakanto, V. J. L. Engel, M. Hutagalung, D. Angela, Sensor networks data acquisition and task management for decision support of smart farming, in *2016 International Conference on Information Technology Systems and Innovation, ICITSI 2016 - Proceedings*, 2017.
- [25] M. A. Zamora-Izquierdo, J. Santa, J. A. Martinez, V. Martinez, A. F. Skarmeta, Smart farming IoT platform based on edge and cloud computing, *Biosyst. Eng.*, 2019, Vol. 177, pp. 4–17.
- [26] S. Rajeswari, K. Suthendran, e K. Rajakumar, A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics, in *Proceedings of 2017 International Conference on Intelligent Computing and Control, I2C2 2017*, 2018, vol. 2018-Janua, pp. 1–5.
- [27] O. Elijah, T. A. Rahman, I. Orikumhi, C. Y. Leow, M. N. Hindia, An Overview of Internet of Things (IoT) and Data Analytics in Agriculture: Benefits and Challenges, *IEEE Internet Things J.*, 2018, Vol. 5, No 5, pp. 3758–3773.
- [28] I. Mohanraj, K. Ashokumar, J. Naren, Field monitoring and automation using IOT in agriculture domain, *Procedia Comput. Sci.*, 2016, Vol. 93, pp. 931–939.
- [29] R. R. Shamshiri et al., Research and development in agricultural robotics: A perspective of digital farming, *Int. J. Agric. Biol. Eng.*, jul. 2018, Vol. 11, No 4, pp. 1–14.