Computers and Electronics in Agriculture 195 (2022) 106813

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Review

Data analytics platforms for agricultural systems: A systematic literature review

Ngakan Nyoman Kutha Krisnawijaya [a], Bedir Tekinerdogan [a,*], Cagatay Catal [b], Rik van der Tol [c]

[a] Information Technology Group, Wageningen University & Research, the Netherlands
[b] Department of Computer Science and Engineering, Qatar University, Qatar
[c] Farm Technology Group, Wageningen University & Research, the Netherlands

ARTICLE INFO

Keywords:
Data analytics platforms
Agriculture
Systematic literature review
Big Data

ABSTRACT

With the rapid developments in ICT, the current agriculture businesses have become increasingly data-driven and are supported by advanced data analytics techniques. In this context, several studies have investigated the adopted data analytics platforms in the agricultural sector. However, the main characteristics and overall findings on these platforms are scattered over the various studies, and to the best of our knowledge, there has been no attempt yet to systematically synthesize the features and obstacles of the adopted data analytics platforms. This article presents the results of an in-depth systematic literature review (SLR) that has explicitly focused on the domains of the platforms, the stakeholders, the objectives, the adopted technologies, the data properties and the obstacles. According to the year-wise analysis, it is found that no relevant primary study between 2010 and 2013 was found. This implies that the research of data analytics in agricultural sectors is a popular topic from recent years, so the results from before 2010 are likely less relevant. In total, 535 papers published from 2010 to 2020 were retrieved using both automatic and manual search strategies, among which 45 journal articles were selected for further analysis. From these primary studies, 33 features and 34 different obstacles were identified. The identified features and obstacles help characterize the different data analytics platforms and pave the way for further research.

1. Introduction

As the growth of the world population is increasing significantly, food security is at the core of the United Nations (UN) development agenda of the Sustainable Development Goals (SDGs) (United Nations). A profound change in the agriculture system is needed to assist the farmer in improving their yields, reducing waste, and making better management decisions to ensure sustainable agriculture (United Nations). The smart farming concept is a broadly adopted concept in agriculture that emphasizes the use of new agricultural technologies and the application of information and communication technology in the farm management domain (Bacco et al., 2019). As smart machines and various sensors are more commonly installed on the farms, farming processes will become increasingly data-driven and data-enabled (Wolfert et al., 2017). With this, data analytics is becoming an important practice to extract valuable information from big data to support the farm management decision-making process (Perakis et al., 2020).

To understand data analytics platforms, it is essential to know which features were supported, what technologies were adopted, which architecture patterns were applied, and what obstacles were faced. Developers of data analytics platforms encounter many challenges in providing the expected data analytics features. Besides, the end-users of data analytics platforms face many problems while using such a system. For example, data analytics result does not provide proper information to the end-users (Perakis et al., 2020). The various level of user's knowledge to understand the summary of data analytics results is also an obstacle faced by the end user (Perakis et al., 2020; Laurent et al., 2019; Baseca et al., 2019). Therefore, this paper provides the systematic literature review (SLR) related to the obstacles encountered by the data analytics platforms' developers and those challenged by the end-users.

As big data analytics gain prominence, data analytics platforms have been investigated in several studies with their applications in several agricultural domains. However, the main characteristics and overall findings on these platforms are scattered over the various studies. To the

* Corresponding author.
E-mail addresses: ngakan.krisnawijaya@wur.nl (N. Nyoman Kutha Krisnawijaya), bedir.tekinerdogan@wur.nl (B. Tekinerdogan), ccatal@qu.edu.qa (C. Catal), rik.vandertol@wur.nl (R. Tol).

https://doi.org/10.1016/j.compag.2022.106813
Received 9 March 2021; Received in revised form 18 January 2022; Accepted 21 February 2022
Available online 9 March 2022

best of our knowledge, there has been no attempt yet to systematically synthesize the features, technical aspects, and obstacles of the adopted data analytics platforms. To address this issue, this paper applies a systematic literature review (SLR) to identify the adopted data analytics platforms for agricultural systems described in the scientific literatures. This study contributes with an SLR to the limited view on good practices and known pitfalls described in selected studies related to the application of big data analytics in agricultural domains. Furthermore, identifying and describing the technical aspects, features, and obstacles are important for practitioners who aim to build a data analytics platform for agricultural domains and researchers who study this topic.

The paper is further organized as follows: Section 2 provides the research methodology, the SLR. Section 3 describes the results of the SLR. Section 4 presents the discussion. Section 5 presents the related work. Finally, Section 6 discusses the conclusions and future work.

2. Research methodology

For the SLR we adopt the guidelines as proposed by Kitchenham et al. (Kitchenham et al., 2009). Fig. 1 shows the steps of the adopted SLR process. In this study, before performing a systematic review, we defined a review protocol based on Tummers, *et al.*, Gurbuz, *et al.*, as well as Köksal and Tekinerdogan (Tummers et al., 2019; Gurbuz and Tekinerdogan, 2018; Köksal and Tekinerdogan, 2017), which also followed Kitchenham *et al.* (Kitchenham et al., 2009). The first step of our protocol was defining the research questions, and then we defined our search protocol (constructing and performing search strings to find the potentials studies), including the database selection. Once the relevant studies were ready, they were selected and assessed using a set of selection criteria and the quality assessment technique. In the fifth step, we utilized a data extraction form to extract the selected papers in order to answer our defined research questions. The third, fourth and fifth steps were iterative process. In the last step, we performed data synthesis and presented the results of the extracted data.

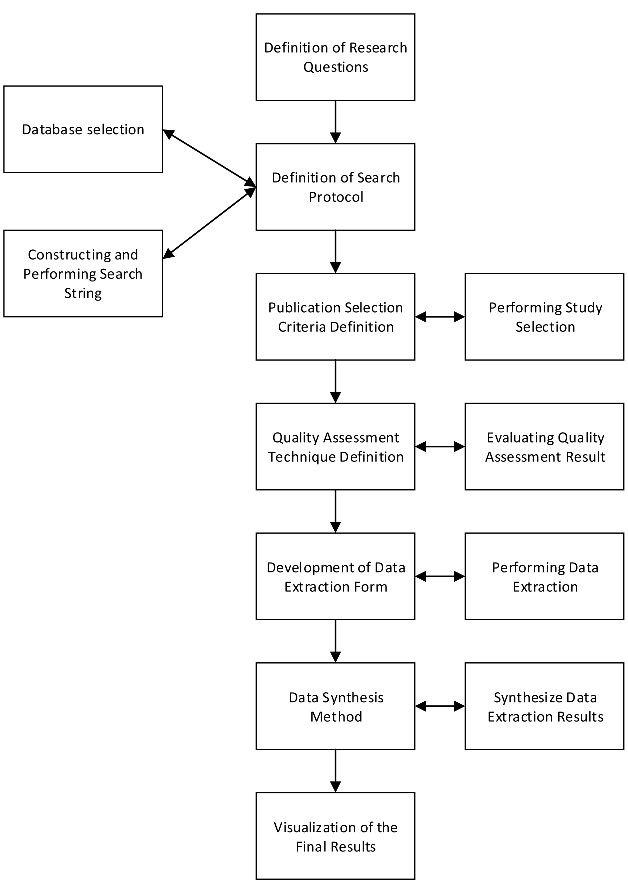

Fig. 1. SLR process used in this study.

2.1. Research questions

This SLR study aims to review the studies published in the domain of data analytics in the agricultural sectors. The definition of research questions was based on our study's purposes, and these questions were answered using the selected studies. For this SLR study, these questions can be seen in Table 1.

The first research question is addressed by identifying which agricultural domains the data analytics platforms were applied (e.g., crop, orchard, greenhouse). The second research question is answered by identifying the person or groups of persons mentioned in the selected studies who can impact or be impacted by the project. As regards the third question, we identified the purposes of performing data analytics in the primary studies.

To understand data analytics platforms, it is essential to know what technologies were adopted, so in question four, we identified the deployed technologies for data analytics as described in the primary articles, which can be classified into four categories, such as the used platforms, programming languages, applied databases, and used software. In addition, this question is followed by four sub-questions. The first sub-question is addressed by observing the features of data analytic platforms. For the second sub-question, we identified the libraries implemented for data analytics. The third sub-question is answered by observing the data analytics task mentioned in the selected studies. For the last sub-question, we identified the data analytics algorithm implemented in the primary studies. To enrich our result, we also provided the architecture pattern used in the selected studies.

Since data analytics is intended to solve data-related agricultural problems, it is essential to know the properties of the data used in the selected studies. In the fifth research question, we checked how the data is obtained, whether the data is public data, proprietary data, or upon-requested data. We also observed the data format mentioned in the selected studies (e.g., JPG, XML, PNG). Furthermore, as one of this study's purposes is to capture the obstacles of data analytics, the following research questions are addressed by identifying the obstacles mentioned in the selected studies and the proposed solutions to tackle those obstacles.

2.2. Search protocol

A systematic search method was performed to find potential articles that would be used to answer our research questions. Both automated and manual searches were utilized to broaden our searched results. In an automated search, we selected digital databases that published high-quality articles: ACM Digital Library, IEEE Xplore, Science Direct, Scopus, Springer, and Wiley Online Library. The range of publication years used in this study was eleven years, considering that the research of data analytics in agricultural sectors is a popular topic from the recent years

Table 1
List of research questions.

No	Questions
RQ1	In which agricultural domain have data analytics platforms been applied?
RQ2	Who are the identified stakeholders?
RQ3	What are the objectives/goals of these data analytics platforms?
RQ4	What are the adopted technologies for data analytics platforms?
RQ4.1	What are the adopted features of these data analytics platforms?
RQ4.2	What are the implementation libraries for data analytics?
RQ4.3	What are the adopted data analytics tasks?
RQ4.4	What are the adopted data analytics algorithms?
RQ5	What are the adopted data properties?
RQ6	What are the obstacles and possible solutions?

so that the results from before 2010 are likely less relevant. This selection was based on previous SLRs, such as (Tummers et al., 2019; Gurbuz and Tekinerdogan, 2018; Köksal and Tekinerdogan, 2017). The automated search performed using search strings in the databases as mentioned earlier. Each database had a different syntax to perform. The general syntax of the search string can be seen in the following query:.

("data analytics" OR "big data" OR "machine learning" OR "deep learning") AND ("agriculture" OR "farm") AND ("Platform" OR "Infrastructure" OR "Software Architecture").

The second column of Table 2 presents the search string results, and the number of studies achieved by performing the search query is 480 studies. Most studies were found in Springer link with 142 papers, and the database with the smallest number of studies was ACM digital library with eight articles. After performing an automated search, this study also applied a manual search to find as many potential studies as possible. The search was performed by manually searching the potential studies in the reference list of studies found by automated search (a.k.a., backward/forward snowballing process). We found 55 articles after applying the snowballing process. In total, 535 papers were gathered, which are shown in Table 2.

2.3. Selection criteria

The search string provides a large number of potential research papers, including journal articles, conference papers, and book chapters. From the searched results, we performed the selection criteria presented in Table 3 to identify the most relevant studies. Firstly, the selection criteria were applied by choosing only journal articles and reading the studies' title and abstract. After that, the next stage was reading the studies thoroughly and performing the exclusion criteria. After applying these criteria, 45 of the 535 papers were selected. The selection process is presented in Table 2.

2.4. Quality assessment

We also consider assessing the quality of 45 studies before doing the data synthesis process. The quality assessment was a part of data extraction in this research. As in the process of reading the articles, we applied quality criteria presented in Table 4. These criteria based on quality instruments were adopted from Kitchenham *et al.* (Kitchenham et al., 2009) and other SLRs research. The quality assessments process aimed to rank the studies based on the quality of their final product, and we used a three-point scale (yes, partial, no), for each criterion found during the process. In this stage, we also consider excluding the papers with a score below four points of eight. We can see in Fig. 2 that all of our primary studies had a higher score than the minimum mark after assessing the entire article. Thus, there was no exclusion decision regarding all of those studies, and all 45 will be used in the data extraction process.

Table 2
Overview of search query result and selection process.

Sources	After Automated and manual search	After applying selection criteria
ACM Digital Library	8	0
IEEE Xplore	54	0
Science Direct	77	7
Scopus	83	4
Springer	142	3
Wiley Online Library	116	1
Manual search	55	30
Total	535	45

Table 3
List of selection criteria.

No	Criteria
EC 1	Papers that do not have full text available
EC 2	Papers which not written in English
EC 3	The duplicate publication that found in multiple sources
EC 4	Papers do not discuss the agricultural domain
EC 5	Papers do not discuss the architecture
EC 6	Papers do not relate to the data analytics platform
EC 7	Experience and survey papers
EC 8	Papers do not validate the proposed study

Table 4
Quality assessment criteria.

No	Question	Yes (1)	Partial (0.5)	No (0)
Q1	Aims clearly stated			
Q2	Scope and Context clearly defined			
Q3	Variable valid and reliable			
Q4	The research process documented adequately			
Q5	All study questions answered			
Q6	Negative findings presented			
Q7	The main findings clearly stated			
Q8	Conclusions relate to the aim of the purpose of the study			

2.5. Data extraction

The research questions should be responded to by reading carefully and entirely the 45 primary studies. Therefore, a data extraction form was developed to collect and retrieve all needed information from those studies. The development of the data extraction form was an iterative process. Firstly, we select several articles randomly and read them rigorously to create the data extraction form. After that, the initial data extraction form was used on other selected articles to extract the data, and if found that the form did not cover all of the information yet, then we updated the form and used the revised form on the following articles. It was a repeated process. Eventually, our final data extraction form includes general information such as the title of the study, the authors, year of publication, publication venue, and publication type. It also contains the specific information which directly answers the research questions, such as targeted domain, stakeholders, data analytics objective, data analytics task, adopted technologies, data properties, features, architectural patterns, obstacles, and proposed solution. The details of the data extraction form were provided in Appendix.

2.6. Data synthesis

In Tummers *et al.* (Tummers et al., 2019), the authors used data synthesis: umbrella concepts, to group all variations of the resulting data from the data extraction stage. For instance, it is common to find multiple names for the same underlying features with the same function. This information needs to be synthesized to be able to describe the correct trend in the articles. We adopted and performed that synthesis technique in our specific data such as targeted domain, features, obstacles, proposed solutions, adopted technologies, and stakeholders. The tables and charts were utilized to visualize and present the resulting data to make data synthesis results easy to understand and communicate.

3. Results

In this section, we will discuss the results of the data extraction and data synthesis stage. The first sub-section will discuss the general statistics of the primary studies, and the next sub-sections will present the results corresponding to the research questions.

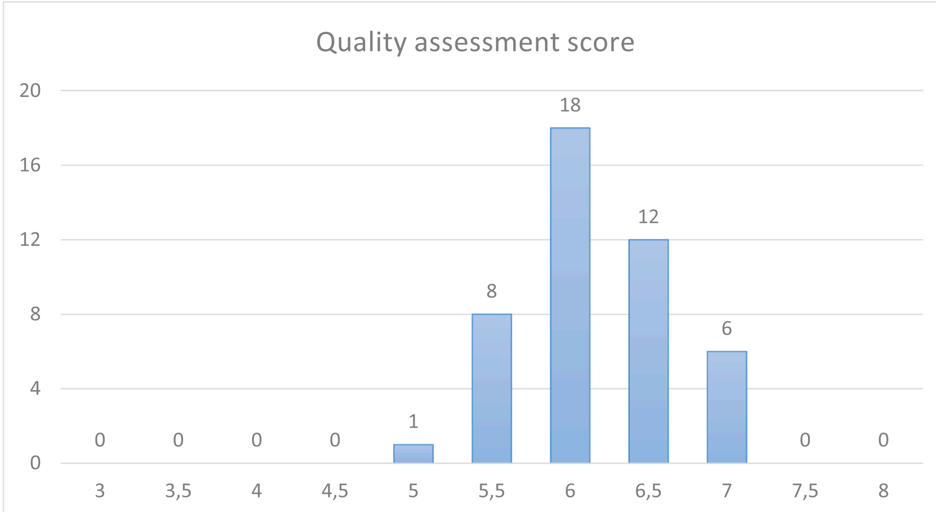

Fig. 2. Quality score distribution of the selected papers.

3.1. General statistics

Fig. 3 shows the year-wise distribution of the selected studies ranging from 2010 to 2020, with the most articles published in 2019 and 2020. No relevant primary study between 2010 and 2013 was found. The resulting data strengthens our assumption that big data and data analytics in agricultural sectors shows an increasing trend in the last five years and has become more prominent over the last two years. The 45 journal articles that were used in this research are listed in Table 5.

The pie chart (Fig. 4) shows that Spain and India are the major contributors to data analytics for agricultural system development literature since those countries produced the most papers, with 12 and 9 documents, respectively. They are followed by China, USA, and Greece, with 7, 6, and 5 documents. Saudi Arabia, France, UK, and Brazil are in sixth place, with three documents.

After conducting data synthesis, it was found that there are 100 institutions that contributed to the literature of data analytics in agricultural domains. However, only five of them has published more than one paper, as shown in Table 6.

Furthermore, based on Fig. 5, it is clearly seen that Computers and Electronics in Agriculture and Sensors journals are the most preferred journals among researchers, with 8 and 4 papers respectively. They are followed by Agronomy, Remote Sensing of Environment, and Remote Sensing, with two articles.

According to Fig. 6, the most cited papers with more than 200 citations belong to Dong et al. (2016) and Zamora-Izquierdo et al. (2019), with 397 and 202 citations, respectively. They are followed by papers of Ferrández-Pastor et al. (2016), Popović et al. (2017), Kaloxylos et al. (2014), Ampatzidis and Partel (2019), and Keswani et al. (2019) The number of citations per paper are represented in Fig. 6.

3.2. RQ-1: In which agricultural domain have data analytics platforms been applied?

According to Röling *et al.* (Röling et al., 2014), the domain is a potential 'system of interest' among stakeholders in a concrete situation, which is attempted to be changed (unfreeze, bypass, or develop). We have identified five agricultural domains as cases in our primary studies. In Fig. 7, it can be seen that the domain that was mentioned the most is Crop with 21 studies. Following this was Orchard with five studies, and it was followed by Greenhouse, Livestock, and Multidomain, each with four studies. In this study, the multidomain was the study that stated or

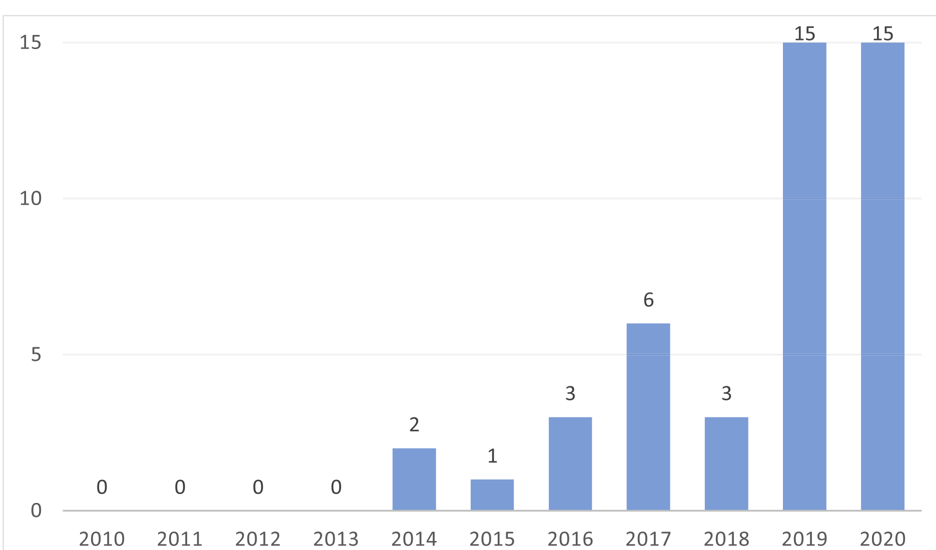

Fig. 3. Year of publication of the primary studies.

Table 5
The 45 selected journal articles.

Study	Year	Study	Year	Study	Year
Perakis et al.	2020	Zamora-Izquierdo et al.	2019	Ampatzidis and Partel	2019
Laurent et al.	2019	Saranya and Nagarajan	2020	Fawcett et al.	2019
Swain et al.	2020	Dong et al.	2016	Singh et al.	2020
Baseca et al.	2019	Kaloxylos et al.	2014	Chen et al.	2019
Ampatzidis et al.	2020	López-Riquelme et al.	2017	Tsipis et al.	2020
Kumar and Sharma	2020	Popović et al.	2017	Subahi and Bouazza	2020
Alonso et al.	2020	Souza et al.	2020	Muñoz et al.	2020
Kamilaris et al.	2018	Liu	2016	Lee and Wang	2020
Yang et al.	2018	Keswani et al.	2019	Meena and Sujatha	2019
Jeppesen et al.	2018	Ferrández-Pastor et al.	2016	Taneja et al.	2020
Pavòn-Pulido et al.	2017	Silva et al.	2014	Cipolla et al.	2019
Li	2019	Cañadas et al.	2017	Vincent et al.	2019
Triantafyllou et al.	2019	Chen et al.	2015	Campos et al.	2019
McCarty et al.	2017	Sawant et al.	2017	Laurent et al.	2020
Bendre and Manthalkar	2019	Salamí et al.	2019	Bahri et al.	2020

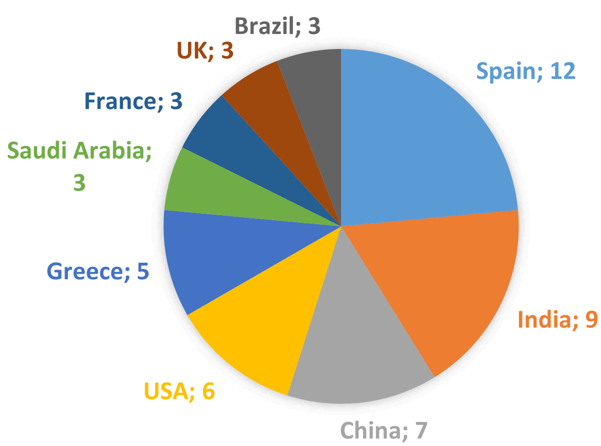

Fig. 4. Number of studies per country.

Table 6
Number of studies per institution.

Institutions	Number of Documents
Iowa State University	2
Iowa Soybean Association	2
University of Florida	2
University of Huelva	2
Technical University of Cartagena	2
University of Almeria	2

mentioned more than one domain as their case studies. For instance, some papers report on multiple domains, such as Perakis *et al.*[P1] mentioned several domains in their study like crop, arable, livestock, and aquaculture. Furthermore, Laurent *et al.*[P2] focused on soya and corn sectors, while Popović et al.[P21] worked in crop and aquaculture domains. The last study that had more than one domain was by Lee and Wang [P38] that mentioned aquaculture and hydroponics as their domains. Finally, we have classified the studies which did not mention any specific domain as the general category, and seven studies were found in this category.

3.3. RQ-2: Who are the identified stakeholders?

A stakeholder is defined as an individual or group who have an interest or ownership in the project, which can contribute in the form of knowledge or support and impact the project or be impacted by the project (Bourne and Walker, 2008). The information about stakeholders involved in the selected studies' research is presented in Fig. 8 and Table 7, where each stakeholder's concerns are described in Table 8. About 15 different stakeholders have been identified. Farmers were the stakeholders that were mentioned the most in the primary studies. In the graph, the researcher and agronomist became the second most mentioned stakeholder in the primary studies. It was also found that one study could have one or more stakeholders involved in the research.

Table 9 presents the stakeholders who engaged in a certain domain. Farmers, of course, is engaged in each domain since mainly the research was conducted to help farmers manage their farms. The livestock domain has the highest number of stakeholders involved; ten out of fifteen stakeholders found in this domain. From this table, we can observe which stakeholders are involved in which domains. For instance, the environmental experts are found only on two domains: crop and livestock.

3.4. RQ-3: What are the objectives of these data analytics platforms?

The authors of selected studies had some purposes in performing data analytics in their research. After conducting data synthesis, in total, nine data-analytics objectives have been identified. Increasing production was the most mentioned objective in the studies. Other frequently mentioned objectives include controlling environmental field, reducing resources, inducing cost-effectiveness, maintaining the quality of the product, and detecting the diseases. The complete overview of the identified objectives of data analytics is shown in Fig. 9 and Table 10.

Fig. 10 shows the correlation analysis between the identified domains and the reported objectives of the primary studies. We can observe that seven out of nine objectives were found in the multi-domain category. Another interesting point from this picture is that the crop domain has the highest rate of increasing production objectives.

Fig. 11 presents the correlation analysis between years and objectives. The graph shows various research purposes in 2020, 2019, and 2017. Increasing production is the most popular objective in 2020 and 2017, while controlling the environment is the crucial issue in 2019. Only three kinds of objectives were found in 2018 and 2016. Two out of eight purposes were found in 2015, and in 2014 only one of them was found. Increasing production and controlling the environment seem to be the most popular objective among researchers since these two objectives almost appear every year.

3.5. RQ-4: What are the adopted technologies for data analytics platforms?

The results related to the data analytics method are described in this section. Firstly, the adopted technologies for data analytics as described in the primary articles were classified into four categories: the used platforms, programming languages, applied databases, and used software. We also provided information about the adopted features, the implemented libraries, and the data analytics task in the following sub-questions. Finally, the architecture patterns of the data analytics

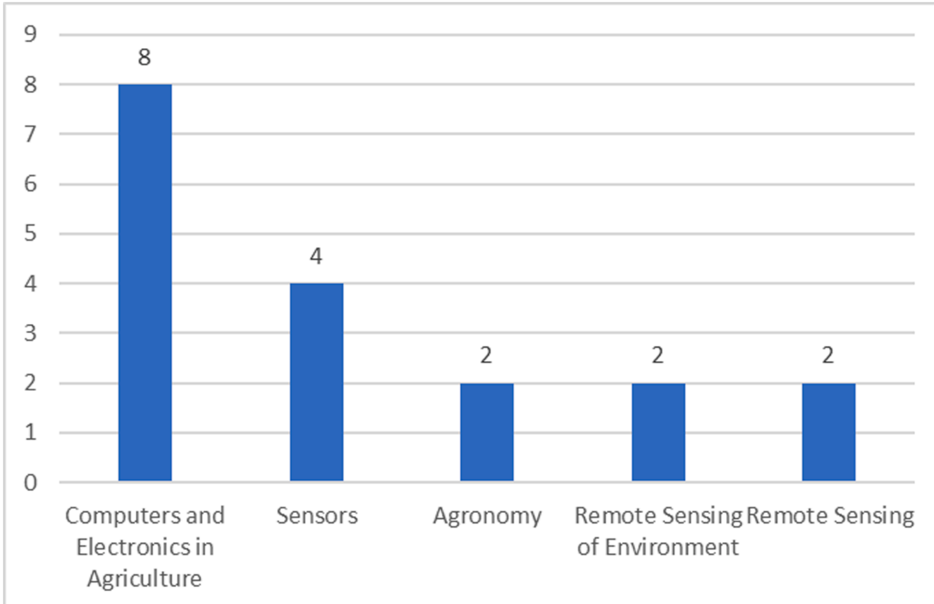

Fig. 5. Top five publication venues.

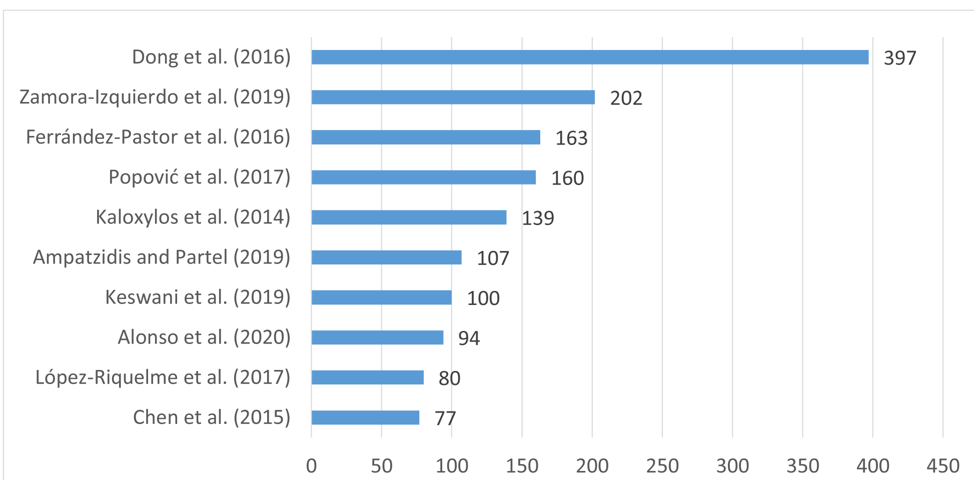

Fig. 6. Top ten cited papers.

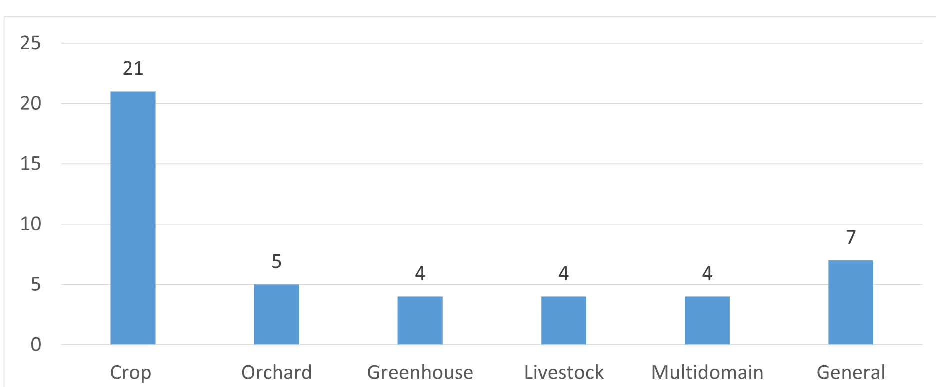

Fig. 7. Number of primary studies based on their specific domain.

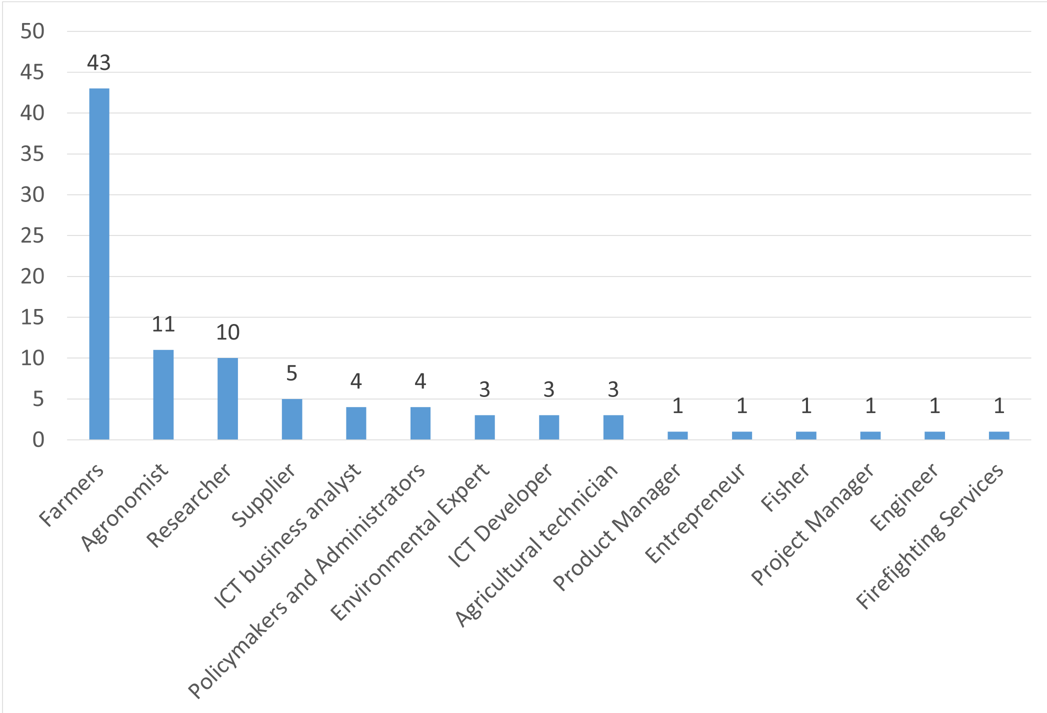

Fig. 8. The types of stakeholders that is identified in the primary studies.

platforms are discussed.

In this study, the data analytics platforms were divided into two subtypes, which are data storage and management platform and data analytics platforms. Data storage and management platform is the integrated software and hardware to capture, prioritize, and manage the data, while data analytics platform is the platform used to perform analysis on the data to retrieve or extract the valuable information.

In Table 11, we can see that Hadoop Distributed File System (HDFS) was the most mentioned data storage platform in the primary studies, followed by Amazon and Web Services, and Google cloud. With regards to the data analytics platform, there are 12 primary studies that mentioned some sort of home-designed platforms as their platform for data analytics. Such a platform was built by the researchers of the particular study. The details of the platforms are mentioned in the primary studies can be seen in Table 11.

Table 12 shows the trend of data storage platforms used for the last six years. As shown in this table, the first three platforms, HDFS, AWS, and Google Cloud, are being used for three years in projects, and they are still used to manage the large-scale data. These three systems are also the most important data storage platforms in software industry. Another interesting point is that IBM Cloudant, ThingSpeak platform, and Neo4j graph-based database appear to be the new platforms in 2020.

Table 13 shows that researchers or practitioners used many open-source tools from the Apache project within the last three years, from 2018 to 2020. Apache Hadoop is the most popular data analytics platform, and Apache Spark and Mahout follow it. Regarding Google products, Google Earth Engine and Google App Engine were also used in 2016 and 2017, respectively.

Table 14 shows the platforms employed for a certain purpose. Generally, Apache projects are almost engaged in every identified goal. This is also strengthened by the result from Table 13, which shows that Apache projects are the most used platforms for analytical purposes. Another observation is that IoT platforms, such as Ubidots, ThingSpeak, and IBM Watson, can be used to achieve several purposes, such as inducing cost-effectiveness, controlling the environmental field, reducing resources consumption, and detecting diseases.

After identifying the applied platforms, this study also investigated the programming languages applied in studies. Fig. 12 illustrates the three programming languages mentioned the most in the primary studies, such as JavaScript and PHP, with 16% and 14% in turn. It was followed by CSS and HTML with a similar percentage at 11%. The details of the programming languages that are mentioned in the primary studies can be seen in Table 15.

Furthermore, in total, 13 databases were described in the selected studies. Furthermore, five of them are mentioned in two or more studies, as depicted in Fig. 13 and Table 16. From the pie chart and the table, it can be seen that the most popular database was MySQL, which occurs in 10 studies. It was followed by Hadoop Distributed File System (HDFS) and Google datastore, with 4 and 3 studies, respectively.

This study also aims to capture the applied software in the paper to perform the data analytics. We could identify 55 software systems that are mentioned in the primary studies, which can be seen in Table 17. The most common software platform is WEKA, which is followed by ArcGIS, Orion Context Broker (OCB), MATLAB, and Pix4Dmapper.

3.5.1. *RQ-4.1: What are the provided features of these data analytics platforms?*

Thirty-three features could be identified from the primary studies. As can be seen in Table 18, the most occurring feature was data visualization, with 28 studies mentioning this feature. Some features were mentioned more than five times in the primary studies, such as data analysis, acquisition, storage, pre-processing, and processing.

Fig. 14 shows the correlation analysis between platforms and features. We found 20 platforms and 33 features. Hence, only top five mentioned features, such as data visualization, data analysis, data acquisition, data storage, and data pre-processing are shown in this figure. Generally, most papers stated that they developed their own platform to apply data analytics. Therefore, the owned platform was the most commonly used for all the features. Apache Hadoop is the second most used one for data analysis, and Apache Mahout and Spark follows.

3.5.2. *RQ-4.2: What are the implementation libraries for data analytics?*

A library is important to programmers to support reuse and capture the specific knowledge (Gregor et al., 2005). The information in

Table 7
The stakeholders mentioned in the primary studies.

Study	Stakeholder Categories														
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
(Perakis et al., 2020)	√	√									√	√			
(Laurent et al., 2019)	√	√	√												
Swain et al. (2020)	√														
(Baseca et al., 2019)	√		√		√				√						
Ampatzidis et al. (2020)				√											
Kumar and Sharma (2020)	√					√									
Alonso et al. (2020)	√	√		√	√			√		√				√	
Kamilaris et al. (2018)	√	√				√			√						
Yang et al. (2018)															
Jeppesen et al. (2018)	√	√	√	√											
Pavòn-Pulido et al. (2017)	√		√	√											
Li (2019)	√														
Triantafyllou et al. (2019)	√														
McCarty et al. (2017)	√				√										
Bendre and Manthalkar (2019)	√														
Zamora-Izquierdo et al. (2019)	√														
Saranya and Nagarajan (2020)	√														
Dong et al. (2016)	√														
Kaloxylos et al. (2014)	√		√		√	√		√							
López-Riquelme et al. (2017)	√		√												
Popović et al. (2017)	√	√						√					√		
Souza et al. (2020)	√														
Liu (2016)	√														
Keswani et al. (2019)	√														
Ferrández-Pastor et al. (2016)	√		√												
Silva et al. (2014)	√						√		√						
Cañadas et al. (2017)	√														
Chen et al. (2015)	√														
Sawant et al. (2017)	√	√													
Salamí et al. (2019)	√														
Ampatzidis and Partel (2019)	√														
Fawcett et al. (2019)	√														
Singh et al. (2020)	√		√												
Chen et al. (2019)	√														
Tsipis et al. (2020)	√	√					√								√
Subahi and Bouazza (2020)	√														
Muñoz et al. (2020)	√			√											
(Lee and Wang, 2020)	√														
Meena and Sujatha (2019)	√														
Taneja et al. (2020)	√														
Cipolla et al. (2019)	√														
Vincent et al. (2019)	√														
Campos et al. (2019)	√	√	√												
Laurent et al. (2020)	√	√	√												
Bahri et al. (2020)	√		√			√	√								

S1: Farmer	S4: Supplier	S7: Environmental Experts	S10: Product Manager	S13: Project Manager
S2: Researcher	S5: ICT Business Analyst	S8: ICT Developer	S11: Entrepreneur	S14: Engineer
S3: Agronomist	S6: Policy Makers and Administrators	S9: Agricultural Technician	S12: Fisher	S15: Firefighting Service

Table 19 shows that there are 25 libraries mentioned in the primary studies. The most common library stated by the primary studies was MCMCglmm, and Shiny (R package), with two studies, mentioned it. The details and description of the libraries can be seen in the table below.

3.5.3. *RQ-4.3: What are the adopted data analytics tasks?*

Data analytics task is the problem-solving being made based on the obtained data and the problems being asked (Provost and Fawcett, 2013). For instance, the classification task predicts whether something will happen, and the regression task predicts how much something will happen.

Furthermore, the data analytics task results described in primary studies are presented in Fig. 15. The graph shows that multiclass classification is the most stated data analytics task in our primary studies, with 21 times mentioned. Other than that, recommendation, anomaly detection, clustering, forecasting, and regression also occur frequently in more than five studies.

Table 20 presents that multiclass classification, recommendation, and anomaly detection are the most popular data analytics tasks and widely used over the years, which have been applied in 21 and 11 primary studies, respectively. Meanwhile, the primary studies' least-mentioned data analytics task is profiling, with only one study mentioning it in 2019.

We also focused on four types of data analytics commonly encountered in the data science system, such as descriptive, diagnostic, predictive, and prescriptive, which will be explained based on Husamaldin and Saeed (Husamaldin and Saeed, 2020) as follows:.

- Descriptive analytics

Descriptive analytics is commonly applied to the historical data to answer the question "what has happened". For instance, to find the average product sales annually.

- Diagnostic analytics

Diagnostic analytics is usually performed when analysing the

Table 8
The identified stakeholders and their concerns regarding data analytics in agriculture.

No	Stakeholder	Concern
S1	Farmers	Have a responsibility to take care of the farm. This stakeholder is the main stakeholder of the most identified studies and is the data analytics platform's end-user.
S2	Researcher	Have a job to discover, investigate and solve agricultural challenges systematically. Their research result can be used as a new insight for further research regarding data analytics platforms.
S3	Agronomist	They use knowledge extracted by the data analytics system to determine and maximize the field's conditions, which will lead to the improvement of crop production.
S4	Supplier	a person or organization that provides agricultural products needed by the customers. The data analytics results in decision or strategic support to improve their revenues.
S5	ICT business analyst	a person or organization who has the knowledge to identify system requirements, make a plan of the system and its documentation, design the system to meet the user's business needs, and then evaluate the existing system to improve its performance if needed.
S6	Policymakers and administrators	a person who decides new policies or rules for government or organizations. Data analytics' results will help them deepen their understanding of certain policy issues, leading to better policymaking.
S7	Environmental expert	Have a responsibility to monitor and identify environmental issues and then recommend the solutions. Data analytics system will help the environmental experts identify environmental issues to give a suitable recommendation to solve the encountered problems.
S8	ICT developer	They have abilities to develop the data analytics system, write or create computer software or applications.
S9	Agricultural technician	They use their agricultural knowledge to improve and find effective ways for farmers to run their business by using a data analytics system to observe, find, and evaluate the valuable information regarding the agricultural practices.
S10	Product manager	Their primary duty is to control crop production, and then they also have some duties such as evaluating and managing agricultural factors like weather, soil condition, diseases, market conditions, and other field conditions. They use the data analytics system results to enhance their knowledge related to maximizing crop production.
S11	Entrepreneur	a person or organization concerned with agricultural products and has the innovation to increase the agricultural products' value by making new or innovative products. Data analytics gather valuable information that can be used to make decisions related to business strategy.
S12	Fisher	They have a job to catch fish, especially for a living. They are the end-user of the data analytics system.
S13	Project manager	Their responsibility is to define the scope of the project, the schedule, planning, procurement, and project execution based on the information they gained from the data analytics system.
S14	Engineer	a person who knows how to design, construct, and use agricultural engines.
S15	Firefighting Services	Have a responsibility to prevent or fight fires, which can be supported by predictive and prescriptive data analytics.

historical data and the data pattern to answer the question "Why something happened ". For instance, an e-commerce manager wants to review their web click pattern and then get valuable information regarding their customer activities.

- Predictive analytics

Using predictive analytics to transform raw data into valuable information in order to make predictions about the future or build information about unknown events and answer the questions "What will happen". An example is building a system to help farmers estimate their production the following day.

- Prescriptive analytics

Prescriptive analytics is applied when developing a system to provide the end-users with the predictions and then suggest advice options to take advantage of them. This analytics type helps to answer the question "What should I do?". In this case, the best example is the decision support system to help the managers determine their strategy to maximise revenue.

The data analytics types and their corresponding number of studies are shown in Fig. 16. The graph shows that descriptive analytics occurred the most in the primary studies. It was followed by diagnostic analytics and predictive analytics, mentioned in 15 and 8 studies in turn. The least frequent data analytics type in primary studies was prescriptive analytics.

Fig. 17 presents the correlation analysis between publication year and data analytics type. According to the graph, it is evident that descriptive and diagnostic analytics are the most common data analytics types mentioned in the primary studies, which have been extensively applied over the six years. Predictive analytics started to be recognized in the past three years, while only a few research papers mentioned prescriptive analytics.

Furthermore, we also make the correlation analysis between data analytics (DA) types and objectives. From previous results, the most used DA type is descriptive analytics, which appears in almost all objectives except in analyzing public opinion. The largest number of research which used descriptive analytics was found regarding the increasing production objective. Diagnostic and predictive analytics were commonly used in the controlling environmental field objective. Furthermore, prescriptive analytics was used for reducing resources consumption. The analysis results are shown in Fig. 18.

3.5.4. RQ-4.4: What are the adopted data analytics algorithms?

There were 60 algorithms found in the primary studies, but only 12 algorithms were mentioned in two or more studies, as illustrated by the graph in Fig. 19. K-Nearest Neighbors (kNN) and linear regression analysis algorithms were mentioned in five studies. The details of identified algorithms in primary studies, as well as their categories, are described in Table 21.

Table 22 shows the data analytic tasks and the employed algorithm to do the task. For instance, to do multiclass classification, we can use several well-known machine learning algorithms such as SVM, Random Forests, kNN, Naïve Bayes, etc., and also some deep learning techniques such as CNN, U-Net, or R-CNN. This table presents the correlation between analytic tasks and algorithms.

We also make a correlation analysis between DA types and algorithms to get more information regarding algorithms used in the literature. Machine learning (e.g., SVM, kNN) and deep learning techniques (e.g., CNN, U-Net, R-CNN) were dominant in every DA type. It seems that these techniques are still the most used ones in data analytics research. Other techniques like statistical analysis, linear regression and Monte Carlo method were found to be used only for descriptive analytics. The analysis results are shown in Table 23.

Table 9
The correlation analysis between identified domain and stakeholders in primary studies.

Domain	Stakeholder Categories														
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
Crop	√	√	√	√	√	√	√		√						
Orchard	√		√	√											
Greenhouse	√														
Livestock	√	√		√	√	√	√	√	√	√				√	
Multidomain	√	√	√					√			√	√	√		
General	√	√	√		√	√	√	√							√

S1: Farmer; S2: Researcher; S3: Agronomist; S4: Supplier; S5: ICT Business Analyst; S6: Policy Makers and Administrators; S7: Environmental Experts; S8: ICT Developer; S9: Agricultural Technician; S10: Product Manager; S11: Entrepreneur; S12: Fisher; S13: Project Manager; S14: Engineer; S15: Firefighting Service

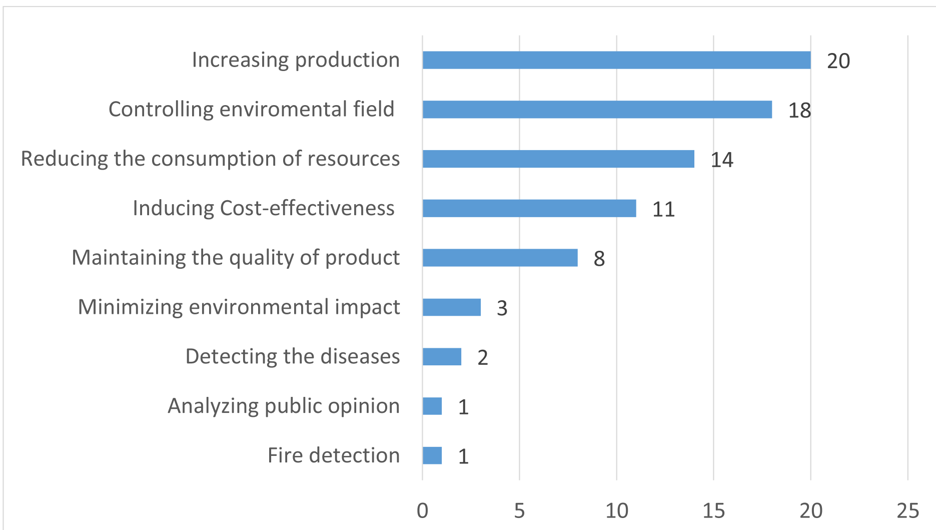

Fig. 9. The types of data-analytics objectives reported in the primary studies.

3.5.5. *Architecture pattern*

Buschmann *et al.* define the architecture pattern as follows: "Architectural patterns express fundamental structural organization schemas for a software system. They provide a set of predefined subsystems, specify their responsibilities, and include rules and guidelines for organizing the relationships between them." (Buschmann et al., 1996). In this research, we consider looking at the system architecture of our primary studies. Fig. 20 and Table 24 shows that five different architecture patterns have been used in the primary studies. The authors' most common architecture pattern was layered pattern, and it was followed by a blackboard pattern with more than five times mentioned in these studies. Some papers also used multiple patterns, such as layers and broker patterns found in Triantafyllou *et al.* [P13], Zamora-Izquierdo *et al.* [P16], López-Riquelme *et al.*[P20], Ferrández-Pastor. [P25], and Muñoz *et al.*[P37], as well as blackboard and layers patterns in Perakis *et al.[*P1]. Furthermore, it was found that around 20 % of the selected studies did not describe their proposed system's architecture pattern.

Based on Buschmann *et al.*, each architecture pattern are explained as follows (Buschmann et al., 1996):.

- The Layered pattern helps structure and split up the application into a group of layers in which each layer has a particular task.
- The Blackboard pattern is commonly used for large data-intensive systems in which data access is performed by multiple actors, and a controller moderates the access to the blacboard.
- The Pipes and Filters pattern is an architectural design pattern for the system that allows for stream processing. This pattern consists of two main components, namely filters and pipes. "Pipes" is the connector between two filters and filters encapsulate the given data.
- The Model-View-Controller (MVC) pattern divides an interactive application into three components. The model has a responsibility to manage data. The views provide and display the information to the end-user. The controllers handle the user inputs and send them to the other components.
- The Broker pattern is commonly used to design a distributed software system, and this pattern uses a broker component as a central unit to coordinate and communicate between components.

In addition, we also linked the publication year with architecture patterns. Based on Table 25, it can be concluded that the layered pattern is the most mentioned architecture pattern in primary studies that have been used since 2014 and is still the most popular until 2020. Meanwhile, blackboard and broker patterns are started to be used in the last two years.

3.6. *RQ-5: What are the adopted data properties?*

This SLR research is interesting when we also analyze the data used in each of the selected studies. We observed how the input data were accessed and obtained and their formats as the input data. Therefore, three data access types were defined: public data, proprietary data, and upon-request data. Public data are the data that can be accessed or provided publicly. The proprietary data are collected and owned privately by an individual, an organization, and/or a group. The upon-request data can be available once the authors make a request or

Table 10
The data analytics objectives mentioned in the primary studies.

Study	Data analytics Objectives								
	O1	O2	O3	O4	O5	O6	O7	O8	O9
(Perakis et al., 2020)	√	√		√	√	√			
(Laurent et al., 2019)	√			√					
Swain et al. (2020)	√								
(Baseca et al., 2019)		√							
Ampatzidis et al. (2020)			√	√					
Kumar and Sharma (2020)								√	
Alonso et al. (2020)	√		√		√				
Kamilaris et al. (2018)	√					√			
Yang et al. (2018)	√	√							
Jeppesen et al. (2018)	√								
Pavòn-Pulido et al. (2017)	√		√						
Li (2019)	√								
Triantafyllou et al. (2019)	√		√		√				
McCarty et al. (2017)	√								
Bendre and Manthalkar (2019)		√							
Zamora-Izquierdo et al. (2019)	√	√			√				
Saranya and Nagarajan (2020)	√								
Dong et al. (2016)		√							
Kaloxylos et al. (2014)		√							
López-Riquelme et al. (2017)	√		√						
Popović et al. (2017)	√				√	√	√		
Souza et al. (2020)		√		√					
Liu (2016)		√	√						
Keswani et al. (2019)		√	√						
Ferrández-Pastor et al. (2016)				√					
Silva et al. (2014)		√							
Cañadas et al. (2017)	√	√	√						
Chen et al. (2015)	√			√					
Sawant et al. (2017)		√							
Salamí et al. (2019)		√		√					
Ampatzidis and Partel (2019)				√					
Fawcett et al. (2019)				√					
Singh et al. (2020)		√							
Chen et al. (2019)			√						
Tsipis et al. (2020)				√					√
Subahi and Bouazza (2020)	√	√	√						
Muñoz et al. (2020)			√	√					
(Lee and Wang, 2020)		√	√						
Meena and Sujatha (2019)					√				
Taneja et al. (2020)							√		
Cipolla et al. (2019)			√		√				
Vincent et al. (2019)					√				
Campos et al. (2019)		√	√						
Laurent et al. (2020)	√								
Bahri et al. (2020)	√								

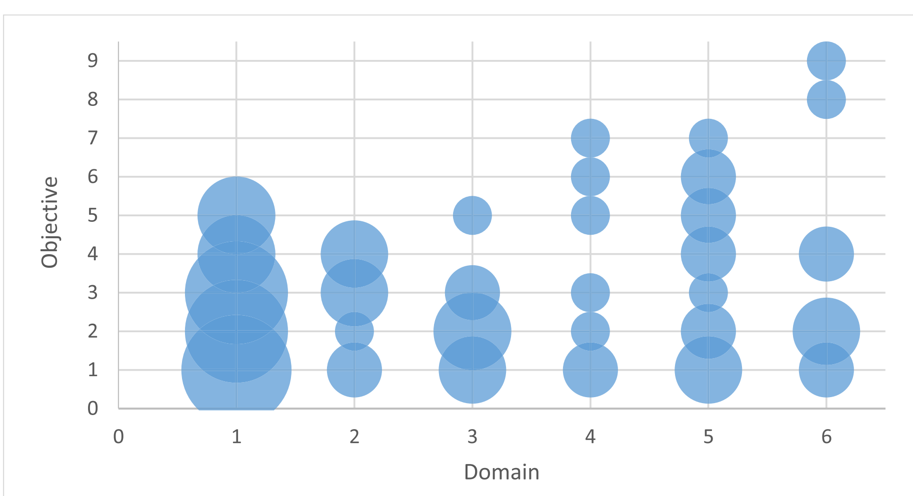

Fig. 10. The correlation between identified domains and objectives in primary studies.

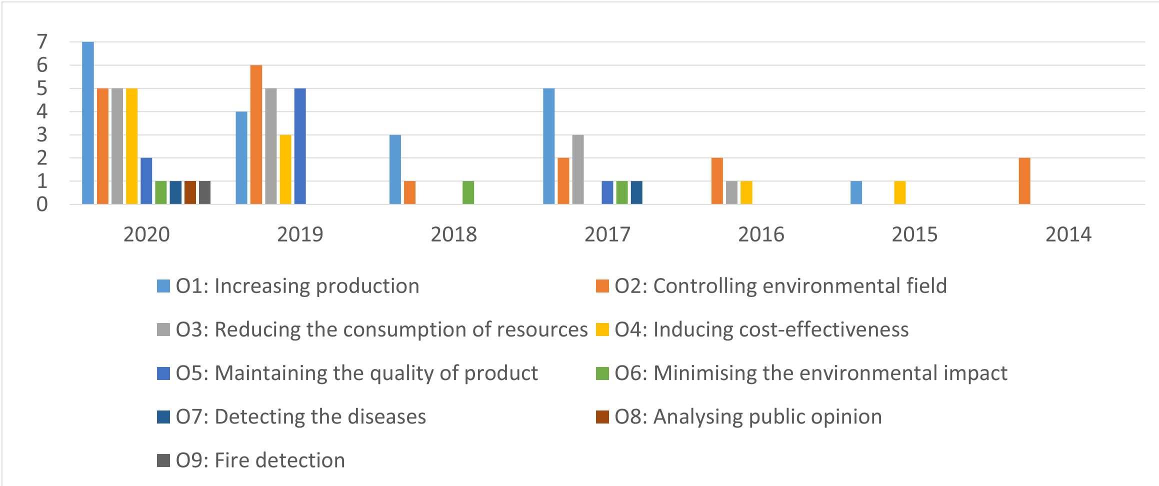

Fig. 11. The correlation analysis between publication year and objectives.

permission to access them.

Fig. 21 and Table 26 present the types of data access and the number of studies in which these were mentioned. In this figure, proprietary data was the most common way to obtain raw data. Following this was public data, and the least common way was upon request data. It was also found that the studies can have one or more data access types to gather their input data. The sources of public data were also presented.

Regarding the input data format, Fig. 22 and Table 27 show that JSON format was the most widely used data format in our primary studies. The XML and CSV formats were the second and third most commonly used as the input data, in turn. This study also identified well-known image-data formats such as RGB, JPG, JPEG, and PNG. The unique image-data formats were also found such as Jpeg2000, GeoTIFF, DNG, and RNB, where they usually were utilized in geographical data analytics. The other formats were found as follows: HTML, sound, TSV, and excel.

3.7. RQ-6: What are the obstacles and possible solutions?

Table 28 gives information about the 34 different obstacles mentioned in the primary studies with the number of corresponding studies. Looking at the details, the vast amounts of data volume occurred the most in the primary studies. Furthermore, obstacles related to the great variety of data, data processing complexity, insufficient user knowledge, and quality of trustworthiness of the data were also mentioned frequently in more than five studies.

According to Table 29, vast amount of data volume and the complexity of data processing and analysis are the most prominent issues that the authors of primary studies have faced for several years. Furthermore, in the past two years, there are several new challenges that are emerged and need to be tackled, such as insufficient of user knowledge, the trustworthiness of the data, the complexity of the implemented application, high variety of data, poor data analytics result, and low system performance.

Table 30 presents the proposed solution to handle the commonly encountered challenges in data analytics approaches with the studies. In the data synthesis, the obstacles that have possible solutions were categorized into two main categories: big data and communication. In Big data issue, two challenges and their proposed solutions have been identified: the problem of handling vast amounts of data volume and the

Table 11
Platform mentioned in the primary studies.

PLATFORMS					
Data storage and management platforms	**Frequency of occurrence**	**Percentage**	**Data analytics platforms**	**Frequency of occurrence**	**Percentage**
Hadoop Distributed File System (HDFS)	4	19%	Owned Platform	12	32%
Amazon Web Services (AWS)	2	10%	Apache Hadoop	5	13%
Google Cloud	2	10%	Apache Mahout	2	5%
Apache Hive	1	5%	Apache Spark	2	5%
GeoNode	1	5%	Amazon Web Services (AWS)	2	5%
Advanced Data Analytics Platform (ADAPT)	1	5%	Google App Engine Backend	1	3%
Cosmos GE	1	5%	Apache Storm	1	3%
ownCloud	1	5%	Apache Kafka	1	3%
Baidu Cloud	1	5%	Google Cloud	1	3%
Ubidots IoT platform	1	5%	QGIS	1	3%
PostGIS	1	5%	Advanced Data Analytics Platform (ADAPT)	1	3%
Neo4j graph-based database system	1	5%	Cosmos GE	1	3%
ThingSpeak platform	1	5%	Baidu Cloud	1	3%
IBM Cloudant	1	5%	Ubidots IoT platform	1	3%
			G2	1	3%
			FIWARE platform	1	3%
			ThingSpeak platform	1	3%
			IBM Watson IoT Platform	1	3%
			Google Earth Engine	1	3%

Table 12
The correlation analysis between data storage platform and year.

Number	Technology	Year					
	Data storage and management platforms	2020	2019	2018	2017	2016	2015
1	Hadoop Distributed File System (HDFS)	√	√√	√			
2	Amazon Web Services (AWS)	√	√				
3	Google Cloud	√			√		
4	Apache Hive			√			
5	GeoNode			√			
6	Advanced Data Analytics Platform (ADAPT)				√		
7	Cosmos GE				√		
8	ownCloud				√		
9	Baidu Cloud					√	
10	Ubidots IoT platform					√	
11	PostGIS						√
12	Neo4j graph-based database system	√					
13	ThingSpeak platform	√					
14	IBM Cloudant	√					

Table 13
The correlation analysis between data analytics platform and year.

Number	Technology	Year					
	Data analytics platforms	2020	2019	2018	2017	2016	2015
1	Apache Hadoop	√ √	√ √	√			
2	Apache Mahout	√		√			
3	Apache Spark	√	√				
4	Amazon Web Services (AWS)	√	√				
5	Google App Engine Backend				√		
6	Apache Storm	√					
7	Apache Kafka	√					
8	Google Cloud	√					
9	QGIS			√			
10	Advanced Data Analytics Platform (ADAPT)				√		
11	Cosmos GE				√		
12	Baidu Cloud					√	
13	Ubidots IoT platform					√	
14	G2				√		
15	FIWARE platform	√					
16	ThingSpeak platform	√					
17	IBM Watson IoT Platform	√					
18	Google Earth Engine					√	

Table 14
The correlation analysis between objective and data analytic platform.

No	Data analytics platform	Objective						
		O1	O2	O3	O4	O5	O6	O7
1	Apache Hadoop	√	√		√	√	√	
2	Apache Spark	√	√	√	√	√	√	
3	Apache Storm	√	√		√	√	√	
4	Apache Mahout	√	√		√	√	√	
5	Apache Kafka	√	√		√	√	√	
6	AWS			√	√	√		
7	Google Cloud	√		√		√		
8	QGIS	√						
9	Google App Engine Backend	√		√				
10	ADAPT	√						
11	Google Earth Engine		√					
12	Cosmos GE	√		√				
13	Baidu Cloud		√	√				
14	Ubidots IoT platform				√			
15	G2	√	√	√				
16	FIWARE platform			√	√			
17	ThingSpeak platform		√	√				
18	IBM Watson IoT Platform							√

O1: Increasing production
O2: Controlling environmental field
O3: Reducing the consumption of resources
O4: Inducing cost-effectiveness
O5: Maintaining the quality of product
O6: Minimising the environmental impact
O7: Detecting the diseases

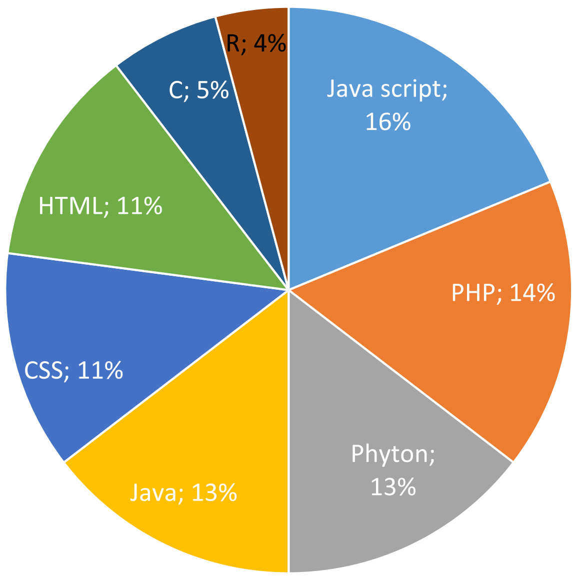

Fig. 12. Programming languages that appear two times or more in the primary studies.

Table 15
Programming languages mentioned in the primary studies.

Programming Language	Frequency of the Occurrence	Percentage
JavaScript	9	16%
PHP	8	14%
Phyton	7	13%
Java	7	13%
CSS	6	11%
HTML	6	11%
C	3	5%
R	2	4%
C++	1	2%
Ruby	1	2%
Perl	1	2%
Ames Stereo Pipeline (ASP)	1	2%
Bootstrap 4	1	2%
Vue.js	1	2%
Visual basic	1	2%
Neo4j Cypher	1	2%

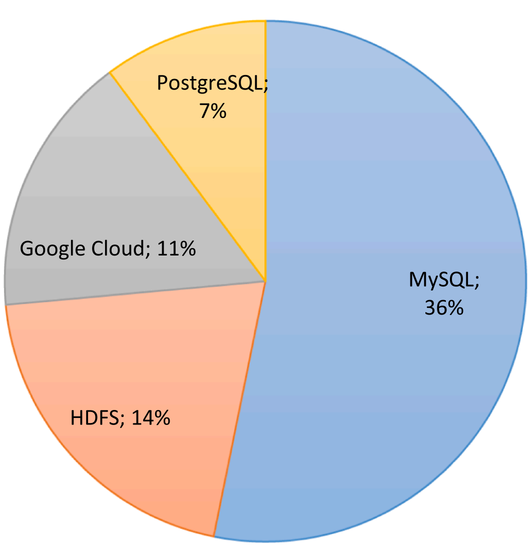

Fig. 13. The databases that occur two times or more in the primary studies.

Table 16
Databases mentioned in the primary studies.

Database	Frequency of the Occurrence	Percentage
MySQL	10	36%
Hadoop Distributed File System (HDFS)	4	14%
Google Cloud	3	11%
PostgreSQL	2	7%
Apache Hive	1	4%
SQL Lite	1	4%
Spatial DBMS	1	4%
Microsoft Access	1	4%
IBM Cloudant	1	4%
Neo4j graph-based database system	1	4%
Microsoft SQL	1	4%
Geodatabase ArcGIS	1	4%
MongoDB	1	4%

Table 17
Software mentioned in the primary studies.

Software			
WEKA	3	The Application Mashup or Wirecloud GE	1
ArcGIS	2	WIDHOC Field Status	1
Orion Context Broker (OCB)	2	WIDHOC Multisensor Monitor	1
MATLAB	2	WIDHOC Multisensor Indicator	1
Pix4Dmapper	2	Laravel framework	1
Docker	1	NOOBS Linux	1
SLURM	1	REST API Ubidots libraries	1
Torque	1	ArcGIS Macros	1
Kubernetes	1	SCADA system	1
MESOS	1	Tianditu	1
YARN	1	Eclipse Mosquitto	1
IReS	1	ArduPilot Mission Planner software	1
Apache Airflow	1	Agisoft Photoscan Professional V1.4.2	1
Tensorflow	1	CloudCompare, V2.9.1	1
MIME	1	CloudSIM	1
Bonita Software	1	IrrSch software	1
Tomcat web server	1	IDAS	1
Open Drone Map (ODM), version 0.3.1	1	Frameworks Django	1
AgriCatVIZ	1	Firebase platform	1
OpenStreetMap	1	React Native framework	1
GeoServer	1	React Native Firebase	1
PostGIS	1	Mosquito	1
OpenSensorHub	1	Microsoft CRM (Customer Relationship Management)	1
Cygnus	1	ESRI ArcView Desktop GIS	1
Apache Karaf	1	Oxdata H2O	1
Jersey	1	Comet	1
Mule ESB	1	R	1
Apache Shrro	1		

data velocity issue. Two challenges were also found regarding communication: latency and network failure, and the complexity of node networks.

4. Discussion

4.1. General discussion

Data analytics research in the agricultural sector progressed significantly in the last two years, between 2019 and 2020. It can be implied that there is a correspondence between the deployment of new technologies for field-level crop management and data analytics to form new insights for the primary user in that period. Several emerging digital technologies, such as the Internet of Things (IoT), remote sensing, cloud computing, image processing, are being applied to support agricultural activities (Kamilaris et al., 2017) and generate vast field data. In order to acquire insights from these data and read their pattern, data analytics

Table 18
Identified features that reported in the primary studies.

Data visualization	28	Data aggregation	3	Data query	1	Data handling	1
Data analysis	25	Data streaming	3	Data loading	1	Data annotation	1
Data acquisition	25	User management	2	Data transmission	1	Data communication	1
Data storage	14	Data service	1	Data preparation	1	Data synchronization	1
Data pre-processing	9	Data check-in	1	Data comparison	1	Data monitoring	1
Data processing	8	Repository	1	Interface	1	Reporting	1
Knowledge base	5	Big Data analytics	1	Security data	1		1
Data management	5	Feature selection	1	Data transformation	1		
Data fusion	3	Knowledge extraction	1	Extract, Transform, and Load (ETL)	1		

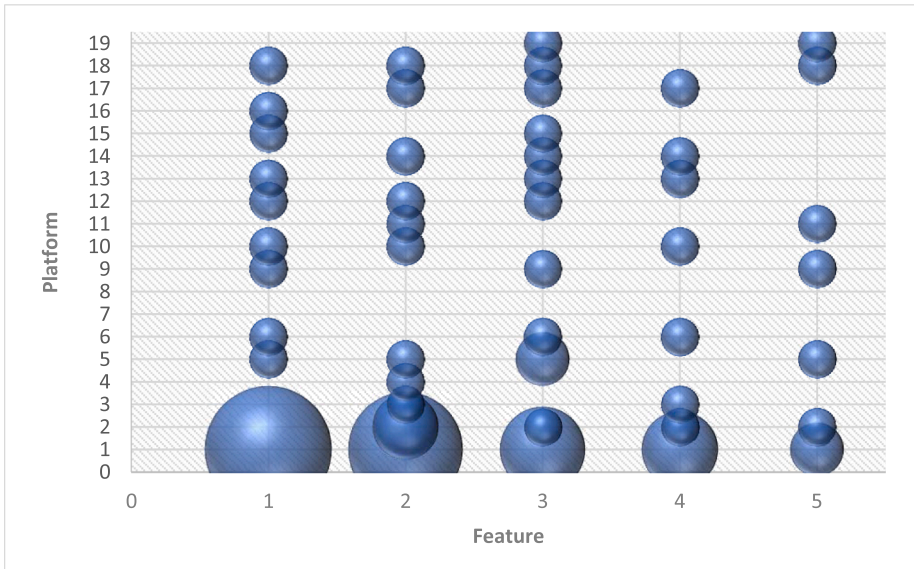

Fig. 14. The correlation analysis between platform (RQ 4) and identified feature (RQ 4.1).

Table 19
Types of libraries mentioned in primary studies.

No	Library	Frequency of Occurrence	Description	Study
1	MCMCglmm	2	R library to implement Markov chain Monte Carlo Sampler	(Laurent et al., 2019), Laurent et al (2020)
2	Shiny (R package)	2	R package that makes it easy to build interactive web apps	(Laurent et al., 2019), Laurent et al (2020)
3	Geographic Data Abstraction Library (GDAL)	2	A library for reading and writing raster and vector geospatial data formats	Jeppesen et al. (2018), McCarty et al. (2017)
4	Tensorflow	1	A library to implement machine learning	Swain et al. (2020)
5	TensorFlow Lite	1	A library to run Tensorflow model on mobile, embedded system, and IoT devices	Alonso et al. (2020)
6	MLBase	1	A library to implement machine learning	(Perakis et al., 2020)
7	Eclipse DeepLearning4j	1	A JAVA's library to implement deep learning	(Perakis et al., 2020)
8	GraphX	1	Apache Spark's Library for graphs and graph-parallel computation	(Perakis et al., 2020)
9	RestNet101 network	1	MATLAB's library to run Deep learning	Ampatzidis et al. (2020)
10	Darknet19 network	1	MATLAB's library to run Deep learning	Ampatzidis et al. (2020)
11	OpenLayers	1	JavaScript library for displaying map data	Jeppesen et al. (2018)
12	C++ Curl library	1	A library for transferring the data using a variety of protocols	Pavòn-Pulido et al. (2017)
13	Octave	1	A data analytics tools similar as SciPy	Popović et al. (2017)
14	SciPy	1	Python's library for analysing scientific and technical computing	Popović et al. (2017)
15	PyWPS	1	Python's library for implementing Geospatial Web Processing Service	Sawant et al. (2017)
16	R lidR package	1	R package for LiDAR data manipulation and visualization	Fawcett et al. (2019)
17	YALMIP toolbox	1	MATLAB's library to model optimization problem	Muñoz et al. (2020)
18	PouchDB	1	JavaScript database library for storing data locally while in the offline status	Taneja et al. (2020)
19	PyETo	1	Python package to calculate crop evapotranspiration	Campos et al. (2019)
20	PyAstronomy	1	Python package for astronomy-algorithms implementation	Campos et al. (2019)
21	XGBoost	1	Machine learning library to process Gradient Boosting Regression Tree	Campos et al. (2019)
22	Leaflet	1	R packages to produce interactive maps	Laurent et al. (2020)
23	Plotly	1	R packages to produce interactive figures	Laurent et al. (2020)
24	R Markdown	1	R packages to produce documents in a variety of formats including HTML, MS Word, PDF, and Beamer.	Laurent et al. (2020)
25	ggplot2	1	R packages to produce interactive figures	Laurent et al. (2020)

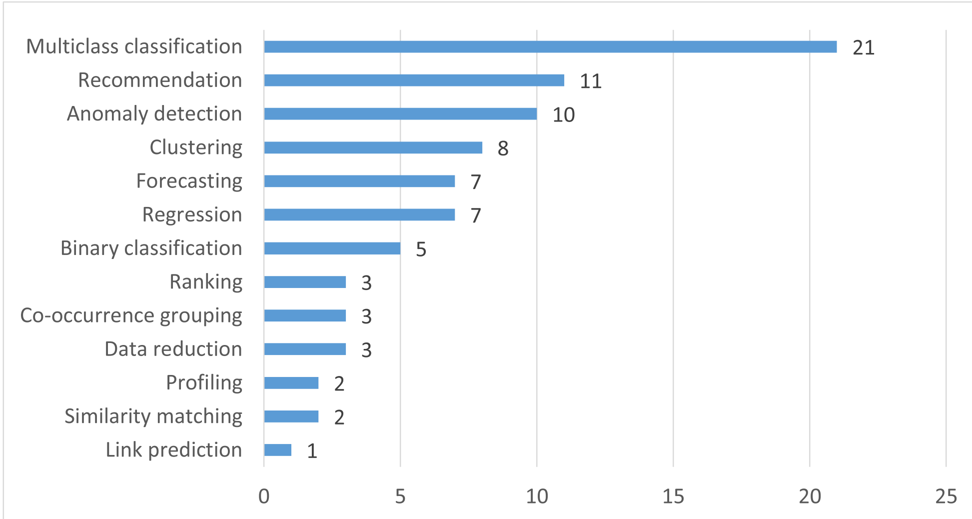

Fig. 15. The types of data analytics task that is identified in the primary studies.

Table 20
The correlation analysis between publication year and data analytics task.

Data analytics task	Year 2020	2019	2018	2017	2016	2015	2014	Total of occurrence in primary studies
Multiclass classification	8	9		1		1	2	21
Recommendation	2	6		1			2	11
Anomaly detection	4	3		2		1		10
Data reduction			1	1	1			3
Regression	5	2						7
Forecasting	2	4		1				7
Binary classification	3		1		1			5
Clustering	6	1		1				8
Ranking	2	1						3
Similarity matching					2			2
Link prediction			1					1
Profiling		1		1				2
Co-occurrence grouping			1	2				3

are deployed. Therefore, many researchers pay attention to data analytics as a tool to handle these generated data.

To the best of our knowledge, this is the first SLR study on (big) data analytics in the domain of agricultural systems. In this respect, more than five hundred papers were identified, from which only 45 high-quality journal articles are selected. The observations were made in the agricultural domain, whereby the crop domain was represented the most in the selected studies. This phenomenon might be caused by the fact that the emerging technologies applied to manage crop practices are relatively simple compared to livestock or aquaculture. There are also several papers that report on multiple domains, where they mentioned more than one domain as their case studies. For instance, Lee and Wang (Lee and Wang, 2020) were focused on aquaponics, which means the combination of hydroponics and aquaculture. Although the main focus is aquaculture, it was categorized as multidomain since the proposed system also gave the advantages for hydroponics side. Besides, Popović *et al.* (T. Popović , N. Latinović , A. Pešić , ărko Zečević , B. Krstajić , and S. Djukanović , , 2017) clearly stated that the proposed system was built for aquaculture and cropped domains.

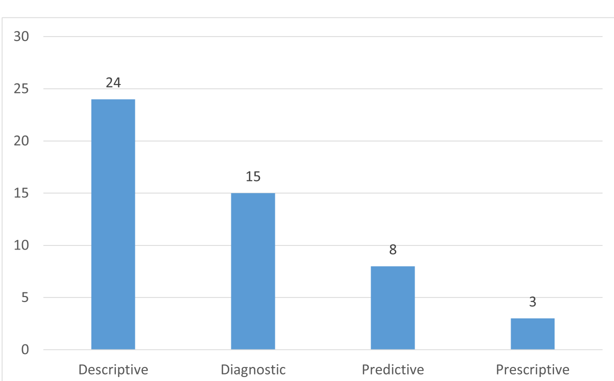

Fig. 16. Number of primary studies based on data analytics type.

Furthermore, the massive interest in conducting data analytics is based on several purposes and motivations, such as increasing production as the main goal in data analytics, while there are many other objectives that could serve the primary goal, such as reducing resources, inducing cost-effectiveness, maintaining the quality of the product, as well as detecting the diseases. Based on these facts, data analytics can be used as a strategy to help the user improve their skills and practices in the field. From data analytics, the user could learn and receive informative results and then make valuable decisions.

On top of that, we also found impressive results regarding the adopted technologies behind the data analytics platforms. This study observed several aspects of the data analytics platform, such as the used platform, the applied programming languages, the databases, the implemented libraries, and the adopted data analytics algorithms. To make our observation more comprehensive, we also reviewed the adopted features of the data analytics platform, the data analytics tasks,

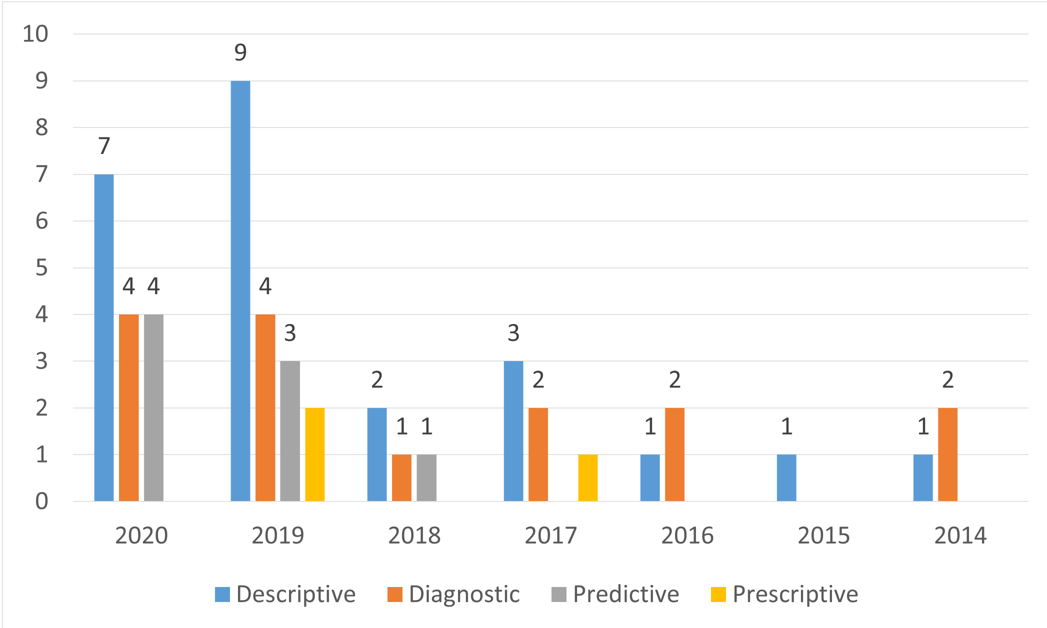

Fig. 17. The correlation analysis between publication year and data analytics type.

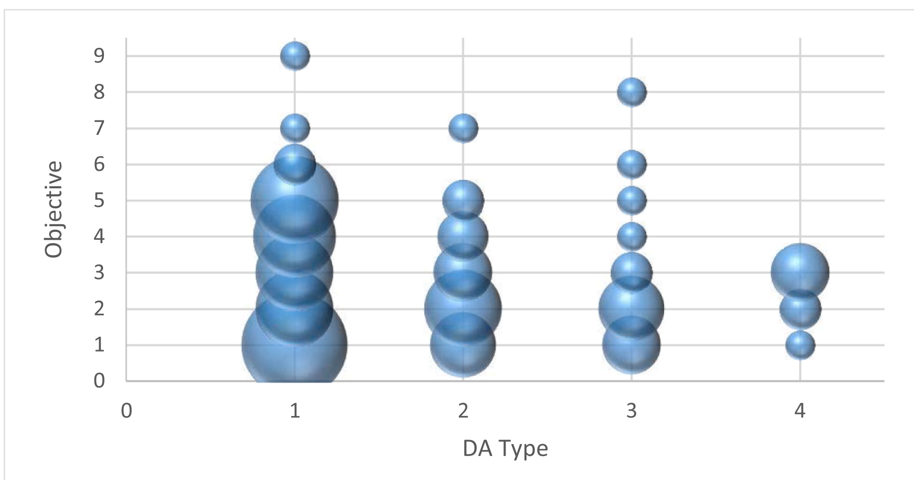

Fig. 18. The correlation analysis between data analytics (DA) types and objectives.

and types, and the applied architectural design pattern. In addition, the platform results (see Table 11) show that we have separated two types of platforms: data storage/management platform and data analytics platform, to have a clear understanding of the implemented platforms in the studies. In the former result, HDFS which part of the Apache Hadoop platform stood in the first rank to manage the datasets. We believe that the use of HDFS is crucial in handling vast amounts of data that have many formats, which were encountered by most studies. In the present, to make a robust data analytics platform, we should be able to manage all of the data types: structured, semi-structured, or unstructured data. HDFS is the NoSQL database in the distributed model, and it is powerful to handle unstructured data like sensor data, social media data, and etc.

Moreover, we have identified other platforms which are AWS, Google Cloud, and IBM. These platforms are well-known platforms to provide many services to handle both data storage and data analytics purposes of their customers. Looking at the data analytics platform, it was found that about twelve of the studies stated that they developed their own platform to perform several modules and software. We have identified that ISOFAST, is the platform to help the users analyze and retrieve useful information based on the datasets (Laurent et al., 2019; Laurent et al., 2021), which appear two times in the primary studies. The other data-analytics platforms were the product of Apache such as Apache Hadoop, Apache Mahout, Apache Spark, Apache Storm, and Apache Kafka, most of them are used to handle big data analytics, and they performed machine learning or deep learning approaches.

Regarding the most mentioned programming languages, we found that JavaScript, HTML, PHP, and CSS were common languages used for deploying a visualization system. The visualization system is crucial in the data analytics platform since it delivered and presented the analytics results in a user-friendly interface. Thus, it can be a communication tool between the end-user and the system. By visualizing the result in a precise, coherent, and user-friendly way, will help the end-user from various levels of knowledge or hi-tech expertise to understand the result of the data analytics platform. In order to develop a data analytics

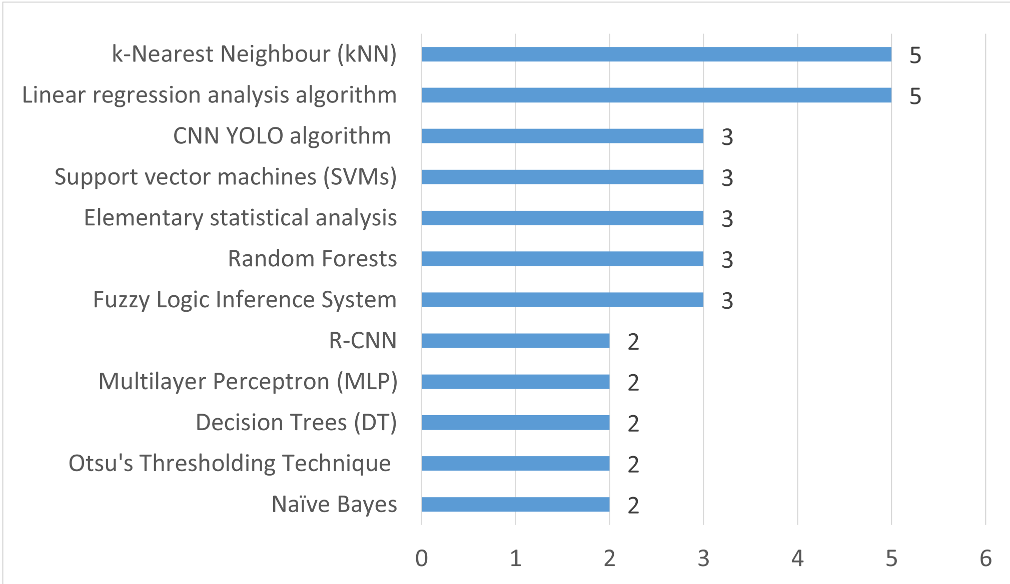

Fig. 19. The types of algorithms that mentioned two times or more in the primary studies.

platform, it was found that several well-known programming languages, such as Python, C, and Java, were utilized. The development will be helped by implementing some standard analytics libraries, like MCMCglmm and Shiny (R package). Furthermore, MySQL was the most common applied database in primary studies.

In the data analytics platform, several features were identified. Data visualization was the most prominent feature in the identified data analytics platforms, and this finding could be the reason why HTML, PHP, and CSS became the most mentioned programming languages. Furthermore, looking at the results (see Table 18), most of the identified features (e.g., data visualization, data analysis, data acquisition, data storage, data processing, data management, data service, and data transformation) were related to managing, handling, processing, storing, and visualizing the data. From the observations, it is clear that proprietary data was the most used type of data. It is evident for data acquisition when implementing IoT, sensor devices, Unmanned Aerial vehicles (UAV), and video cameras. The three formats of the input data most seen in the data analytics platforms were image data (RGB, JPG, JPEG, and PNG), JSON, and XML format. Finally, feature diagrams are made to give a general picture of data analytics platforms' features based on our observations. Feature models are widely used to identify and capture the features of the software system and are commonly utilized to contribute to the architecture design of the system (De Vylder, 2011; Van Geest et al., 2021). The feature diagrams are used to represent the feature models. Fig. 23 shows the feature diagrams of the data analytics platforms based on the results of our SLR study. As regards the diagram, there are several important features in the data analytics platforms, such as data acquisition, data pre-processing, data processing, data analysis, data visualization, data management, and data security. These features are the basic features that should be provided when developing a data analytics platform. Furthermore, regarding the data visualization feature, it has a mandatory feature like a user-friendly

Table 21
Types of algorithms mentioned two times or more in primary studies.

No	Algorithm	Frequency	Studies	Category
1	k-Nearest Neighbour (kNN)	5	Kumar and Sharma (2020), Bendre and Manthalkar (2019), Souza et al. (2020), Fawcett et al. (2019), Taneja et al. (2020)	Machine Learning Approach
2	Linear regression analysis algorithm	5	Alonso et al. (2020), Saranya and Nagarajan (2020), Fawcett et al. (2019), Campos et al. (2019), (Lee and Wang, 2020)	Regression Analysis
3	CNN YOLO algorithm	3	(Perakis et al., 2020), Ampatzidis et al. (2020), Ampatzidis and Partel (2019)	Deep Learning
4	Support vector machines (SVMs)	3	Swain et al. (2020), Kumar and Sharma (2020), Yang et al. (2018)	Machine Learning Approach
5	Statistical analysis	3	Pavòn-Pulido et al. (2017), Liu (2016), Ferrández-Pastor et al. (2016)	Statistical models
6	Random Forests	3	Yang et al. (2018), Triantafyllou et al. (2019), Campos et al. (2019)	Machine Learning Approach
7	Fuzzy Logic Inference System	3	Keswani et al. (2019), Cañadas et al. (2017), Singh et al. (2020)	Decision Support System
8	R-CNN	2	(Perakis et al., 2020), Ampatzidis et al. (2020)	Deep Learning
9	Multilayer Perceptron (MLP)	2	Kumar and Sharma (2020), Vincent et al. (2019)	Machine Learning Approach
10	Decision Trees (DT)	2	Kumar and Sharma (2020), Yang et al. (2018)	Machine Learning Approach
11	Otsu's Thresholding Technique	2	McCarty et al. (2017), Salamí et al. (2019)	Image processing technique
12	Naïve Bayes	2	Kumar and Sharma (2020), Kaloxylos et al. (2014)	Machine Learning Approach

Table 22
The correlation analysis between data analytics tasks and algorithms.

Data analytics task	Algorithm(s)	Data analytics task	Algorithm(s)
Multiclass classification	Support vector machines (SVMs), CNN YOLO algorithm, U-Net, Rete pattern-matching algorithm, R-CNN, Random Forests, Autoregressive Integrated Moving Average (ARIMA), k-Nearest Neighbour (kNN), Naïve Bayes, Multi-criteria Decision Aid (MCDA) method using Outranking approach, Monte Carlo (MC) method, the Proportional-Integral-Derivative (PID) control, Association rule mining, Multilayer Perceptron (MLP)	Forecasting	ANN with Population-based incremental learning, Fuzzy Logic Inference System, Variable learning rate gradient descent (VLRGD), Soil Water Stress Factor (SWFAC) Algorithm, Chandler burning index (CBI), Association rule mining, Decision Stump, RepTree
Recommendation	Collaborative Filtering, Nearest-neighbour with Fuzzy Logic, Multi-criteria Decision Aid (MCDA) method using Outranking approach, Fuzzy Logic Inference System, Soil Water Stress Factor (SWFAC) Algorithm, Model Predictive Control (MPC), Multilayer Perceptron (MLP)	Binary classification	k-Nearest Neighbour (kNN), Support vector machines (SVMs), Multilayer Perceptron (MLP), Naïve Bayes, Decision Trees (DT), Adaboost algorithm, Random Forests
Anomaly detection	Angle-Based Outlier Detection, Local Outlier Factor, Cluster Based Local Outlier Factor, Isolation Forest, Fuzzy Logic Inference System, Inverse distance weighting (IDW) interpolation, Kriging, Spline interpolation, Reference evapotranspiration estimation, Soil Water Stress Factor (SWFAC) Algorithm, Chandler burning index (CBI), Linear regression analysis algorithm, Generalized ESD (Extreme Studentized Deviate) algorithm, Chauvenet, Z-Score, Modified Z-Score, RepTree	Clustering	CNN YOLO algorithm, U-Net, R-CNN, ANN with Population-based incremental learning, k-Nearest Neighbour (kNN), Chandler burning index (CBI), Association rule mining
Data reduction	Statistical analysis	Ranking	Collaborative Filtering, Histogram-Based Outlier Score
Regression	Bayesian model, Support vector machines (SVMs), Linear regression analysis algorithm, Gradient Boosting Regression Tree (GBRT)	Similarity matching	Statistical analysis

Table 23
The correlation analysis between data analytics (DA) type and algorithm.

DA type	Algorithm(s)	DA type	Algorithm(s)
Descriptive	Bayesian model, Rete pattern-matching algorithm, R-CNN, CNN YOLO algorithm, Linear regression analysis algorithm, Statistical analysis, Random Forests, Isolation Forest, k-Nearest Neighbour (kNN), Inverse distance weighting (IDW) interpolation, Kriging, Spline interpolation, Speeded up robust features (SURF), Stereo vision-based segmentation, Monte Carlo (MC) method, Chandler burning index (CBI), Multilayer Perceptron (MLP)	Predictive	CNN YOLO algorithm, U-Net, Hidden Markov Modelling, Support vector machines (SVMs), Multilayer Perceptron (MLP), Decision Trees (DT), Naïve Bayes, Fuzzy Logic Inference System, k-Nearest Neighbour (kNN), ANN with Population-based incremental learning, Gradient Boosting Regression Tree (GBRT), Decision Stump, RepTree, Random Forests
Diagnostic	Bayesian model, Support vector machines (SVMs), Decision Trees (DT), Adaboost algorithm, Collaborative Filtering, Random Forests, Phenology and pixel-based paddy rice (PPPM), Naïve Bayes, Nearest-neighbour with Fuzzy Logic, Statistical analysis, Multi-criteria Decision Aid (MCDA) method using Outranking approach, Fuzzy Logic Inference System, Model Predictive Control (MPC), Linear regression analysis algorithm, k-Nearest Neighbour (kNN)	Prescriptive	Variable learning rate gradient descent (VLRGD) and Gradient descent, Fuzzy Logic Inference System, Soil Water Stress Factor (SWFAC) Algorithm

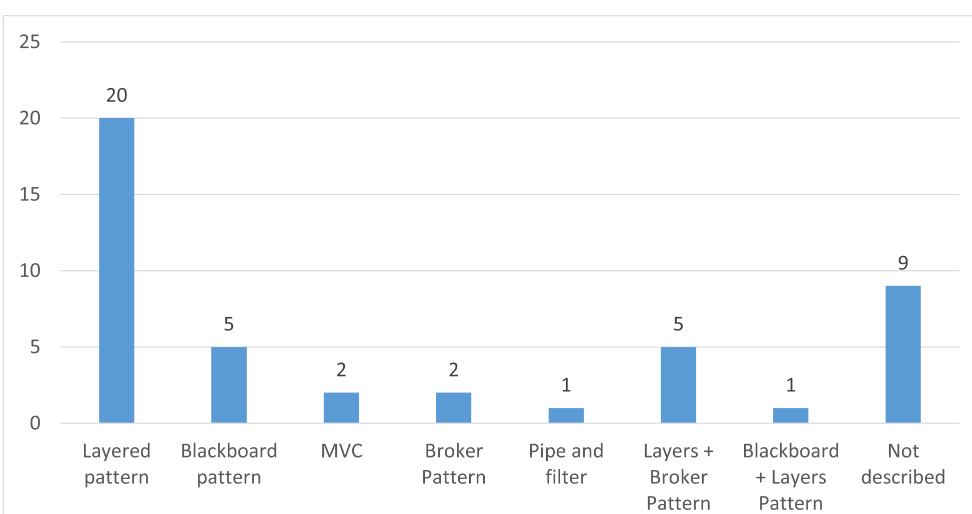

Fig. 20. Number of studies per architecture pattern.

Table 24
Architecture patterns mentioned in primary studies with the corresponding studies.

Architecture	Number of appearances	Studies
Layered pattern	20	(Baseca et al., 2019), Ampatzidis et al. (2020), Alonso et al. (2020), Kamilaris et al. (2018), Yang et al. (2018), Jeppesen et al. (2018), Pavòn-Pulido et al. (2017), McCarty et al. (2017), Bendre and Manthalkar (2019), Kaloxylos et al. (2014), Liu (2016), Cañadas et al. (2017), Chen et al. (2015), Sawant et al. (2017), Tsipis et al (2020), Subahi and Bouazza (2020), Taneja et al. (2020), Cipolla et al. (2019), Campos et al. (2019), Bahri et al. (2020)
Blackboard pattern	5	(Laurent et al., 2019), Kumar and Sharma (2020), Souza et al. (2020), (Lee and Wang, 2020), Vincent et al. (2019)
MVC	2	Popović et al. (2017), Silva et al. (2014)
Broker Pattern	2	Salamí et al. (2019), Singh et al. (2020)
Pipe and filter	1	Fawcett et al. (2019)
Layers + Broker Pattern	5	Triantafyllou et al. (2019), Zamora-Izquierdo et al. (2019), López-Riquelme et al. (2017), Ferrández-Pastor et al. (2016), Muñoz et al. (2020)
Blackboard + Layers Pattern	1	(Perakis et al., 2020)
Not described	9	Swain et al. (2020), Li (2019), Saranya and Nagarajan (2020), Dong et al. (2016), Keswani et al. (2019), Ampatzidis and Partel (2019), Chen et al. (2019), Meena and Sujatha (2019), Laurent et al. (2020)

Table 25
The linked between publication year and architecture pattern.

Architecture pattern	Year							Total of occurrence in primary studies
	2020	2019	2018	2017	2016	2015	2014	
Layered pattern	6	4	3	4	1	1	1	20
Blackboard pattern	3	2						5
MVC				1			1	2
Broker Pattern	1	1						2
Pipe and filter		1						1
Layers + Broker Pattern	1	2		1	1			5
Blackboard + Layers Pattern	1							1

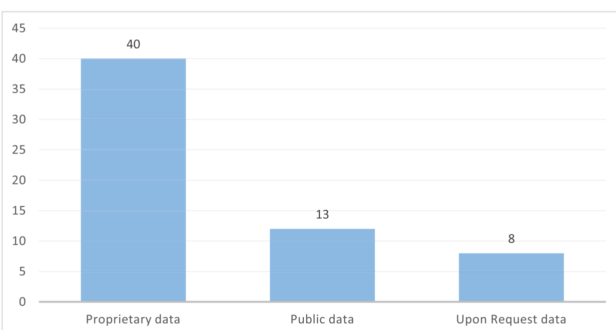

Fig. 21. The number of studies based on their types of data access.

interface with sub-features like reporting and data monitoring.

In another observation, it was found that data analytics have some obvious tasks, and it was affected by the problems to answer and the given datasets. In this SLR, we also considered what kind of tasks are required in the identified data analytics platforms and the algorithms applied to solve these tasks. Multiclass classification, recommendation, anomaly detection, clustering, as well as forecasting and regression were the top five identified tasks. Regarding data analytics types, descriptive analytics stands in the top rank, followed by diagnostic analytics.

Furthermore, the algorithms applied to solve the tasks were k-Nearest Neighbors (kNN) in the top rank, followed by linear regression analysis, CNN YOLO algorithm, Support vector machines (SVMs), Random Forests, Fuzzy Logic Inference System, and elementary statistical analysis. Statistical analysis is employed in descriptive analytics. This result proves that apart from the machine learning approach, simple algorithm such as statistical analysis was still used in the data analytics platforms. Most of the identified data analytics platforms present the design of their system. The Layered pattern was the most popular pattern used to deliver the system's design, and the Blackboard pattern stood in second place and then followed by the MVC pattern, Broker pattern, and pipes and filters pattern. In addition, multiple patterns were also found such as layers and broker patterns, as well as layers and blackboard patterns. Related to the studies that used layers and broker patterns, mostly developed a data analytics platform using IoT technology as their data acquisition. The broker system acted as the central system to gather, control, and manage the input data from various sensors, and this system is generally placed between physical and data layers. However, there were some papers do not explicitly describe their system architecture.

In this review, we identified 34 different obstacles and some of them were problems concerning big data issues, such as vast amounts of data volume, a wide variety of data, the trustworthiness of the data, high speed of data velocity, and latency. We also discovered some possible solutions, like system scalability both vertical or horizontal techniques or big data platform and parallel computing implementation, to tackle those identified obstacles.

Finally, several correlation analyses had been made to enhance the SLR results. In this study, the correlation analysis between publication years and research questions (RQs) was conducted to derive knowledge regarding the trend in specific RQ. For instance, in Table 9, we analyzed the link between years and technologies (RQ 4) to obtain information regarding recent or outdating technologies used in data analytics in agriculture. Another analysis is presented in Fig. 11. This figure shows the trend of data analytics objectives in recent years. Increasing production and controlling environmental field objectives almost appear every year. Furthermore, to enrich our results, we also analyzed among RQs, such as analyzing between domain (RQ1) and stakeholder (RQ2) to get insight into which stakeholders most engaged in which domain. To determine which objective frequently occurs in which domain, we analyzed between domain (RQ1) and objective (RQ3). We also linked the objective and other RQs, such as data analytics platform (RQ4) and data analytics type (RQ4.3), to know which technologies are primarily

Table 26
Types of access for data sources in the primary studies.

No	Study	Public data	Proprietary data	Upon request data	Public sources
1	(Perakis et al., 2020)		√		
2	(Laurent et al., 2019)		√		
3	Swain et al. (2020)		√		
4	(Baseca et al., 2019)		√		
5	Ampatzidis et al. (2020)		√		
6	Kumar and Sharma (2020)	√			from twitter.com
7	Alonso et al. (2020)		√		
8	Kamilaris et al. (2018)		√		
9	Yang et al. (2018)		√		
10	Jeppesen et al. (2018)	√	√	√	from governmental offices
11	Pavòn-Pulido et al. (2017)		√		
12	Li (2019)	√			https://www.bookcrossing.com
13	Triantafyllou et al. (2019)		√		
14	McCarty et al. (2017)	√	√	√	https://espa.cr.usgs.gov/
15	Bendre and Manthalkar (2019)	√			https://www.indiawaterportal.org
16	Zamora-Izquierdo et al. (2019)		√		https://redhook.gsfc.nasa.gov/
17	Saranya and Nagarajan (2020)	√			https://www.eomf.ou.edu/photos/
18	Dong et al. (2016)	√			
19	Kaloxylos et al. (2014)	√	√		https://www.myweather2.com
20	López-Riquelme et al. (2017)		√	√	
21	Popović et al. (2017)	√	√	√	Not mentioned
22	Souza et al. (2020)		√		
23	Liu (2016)		√		
24	Keswani et al. (2019)		√		
25	Ferrández-Pastor et al. (2016)		√		
26	Silva et al. (2014)		√		
27	Cañadas et al. (2017)		√		
28	Chen et al. (2015)		√		
29	Sawant et al. (2017)		√		
30	Salamí et al. (2019)	√	√		https://centrodedescargas.cnig.es
31	Ampatzidis and Partel (2019)		√		
32	Fawcett et al. (2019)		√	√	
33	Singh et al. (2020)		√		
34	Chen et al. (2019)	√	√	√	https://api.openweathermap.org
35	Tsipis et al. (2020)		√		
36	Subahi and Bouazza (2020)		√		
37	Muñoz et al. (2020)		√		
38	(Lee and Wang, 2020)	√	√		https://water.usgs.gov/software/DOTABLES
39	Meena and Sujatha (2019)		√		
40	Taneja et al. (2020)		√		
41	Cipolla et al. (2019)		√		
42	Vincent et al. (2019)		√		
43	Campos et al. (2019)		√		Brazilian national weather station
44	Laurent et al. (2020)	√	√	√	https://mesonet.agron.iastate.edu/
45	Bahri et al. (2020)		√		
Total		13	40	7	

applied to achieve the objectives. In addition, we presented the correlation between data analytics properties (e.g., task and type) and algorithms that are mainly mentioned in primary studies.

4.2. Potential threats to validity

Threats of validity are one of the essential mechanisms to ensure the rigorous finding of an SLR. The possible threats to validity can be the construct, internal, external, and conclusion (Zhou et al., 2016). We believe that the construct and internal validity are addressed for this SLR study since we had a discussion among authors in formulating all research questions in this study, as well as defining the relationship between research questions and research objectives. Thus, we hope that the possibility of the narrow relationship between our findings is already minimized. In addition, the focus of the literature search was on journal articles since this study intended to use high-quality research papers only. Initially, we started checking conference papers as well, but we found that most of them are short and very brief; hence they cannot provide sufficient information needed to address the research questions defined in this research. In addition, the review procedure of the many conferences is not similar to the journal review procedures that take several months. Journal articles pass a thorough and meticulous review from the field experts prior to getting published in the highly reputed journals. Most of the authors also extend their conference papers to be published in highly reputed journals, or the conference organizers encourage and provides the authors an opportunity to extend their papers and publish them in high-quality journals. Furthermore, recent SLR papers also used only journal articles as the primary studies, and they have been published in a highly reputed journals (Catal et al., 2022)–(Pathak et al., 2019). Due to these reasons, we decided to include only journal articles as primary studies. We believe that 45 articles selected in this research are able to present the-state-of-the-art in this field.

Furthermore, we systematically integrate both automated and manual search strategies to find the existing literature. We also thoroughly evaluate search performance among authors since we realize that each targeted database has a different query format and performs different logical operators, which means that the query design needs to be adjusted across databases. The search results were evaluated by reading the abstract of the returned papers. The query was modified again if the papers were not relevant. After modifying the query, the search was performed again. These are the iterative process until the search query captures the relevant papers. To capture as many relevant

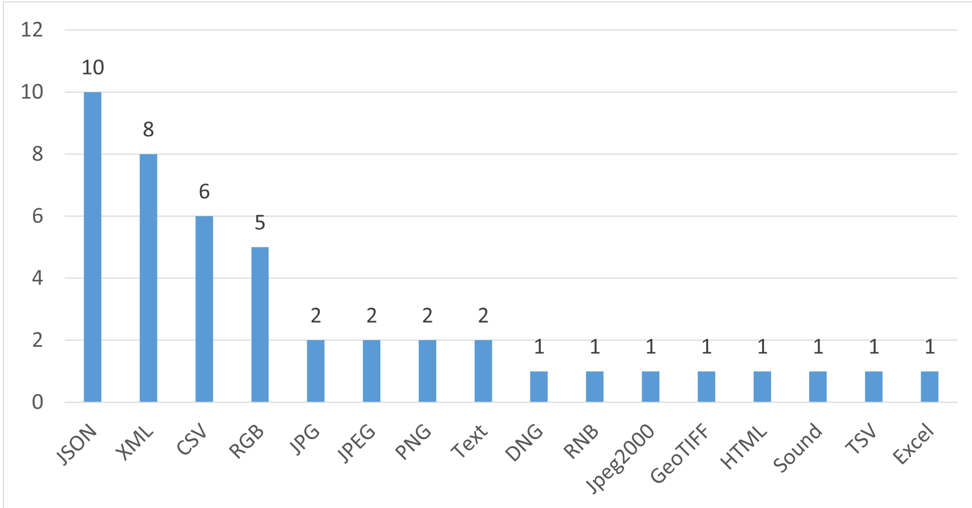

Fig. 22. The types of data format mentioned in the primary studies.

Table 27
Data format mentioned in primary studies.

Data Format	Studies	Frequency of the Occurrence	Percentage
JSON	(Baseca et al., 2019), Yang et al. (2018), Jeppesen et al. (2018), Pavòn-Pulido et al. (2017), Zamora-Izquierdo et al. (2019), López-Riquelme et al. (2017), Popović et al. (2017), Sawant et al. (2017), Muñoz et al. (2020), Taneja et al. (2020)	10	22%
XML	(Baseca et al., 2019), Jeppesen et al. (2018), Triantafyllou et al. (2019), Kaloxylos et al. (2014), López-Riquelme et al. (2017), Chen et al. (2015), Sawant et al. (2017), Subahi and Bouazza (2020)	8	18%
CSV	Kamilaris et al. (2018), Popović et al. (2017), Souza et al. (2020), Ferrández-Pastor et al. (2016), Sawant et al. (2017), Subahi and Bouazza (2020)	6	13%
RGB	Ampatzidis et al. (2020), Triantafyllou et al. (2019), Liu (2016), Ampatzidis and Partel (2019), Fawcett et al. (2019)	5	11%
JPG	McCarty et al. (2017), Chen et al. (2015)	2	4%
JPEG	Ampatzidis et al. (2020), Salamí et al. (2019)	2	4%
PNG	Jeppesen et al. (2018), Chen et al. (2015)	2	4%
Text	Kumar and Sharma (2020), Meena and Sujatha (2019)	2	4%
DNG	Salamí et al. (2019)	1	2%
RNB	Ampatzidis and Partel (2019)	1	2%
Jpeg2000	Jeppesen et al. (2018)	1	2%
GeoTIFF	McCarty et al. (2017)	1	2%
HTML	(Perakis et al., 2020)	1	2%
Sound	(Perakis et al., 2020)	1	2%
TSV	Sawant et al. (2017)	1	2%
Excel	Laurent et al (2020)	1	2%

papers as possible, we also performed snowballing as a complementary method to avoid the possibility of missing any relevant paper.

Another repeating process is when defining the data extraction form. We also consider that the predefined data extraction form is not cover all of the useful data. Therefore, the form is updated during the extraction process to ensure that the form already covered the selected papers' invaluable data. After data extraction, data synthesis process was conducted. Due to this process, the name of some items, such as stakeholders, objectives, architecture patterns, features, obstacles, and proposed solutions are not necessarily similar to those mentioned in selected studies.

For conclusion validity, most of our time conducting this SLR study was spent discussing among the authors. The discussion aims to minimize bias interpretation among the researchers and keep following the research protocols from designing research questions to making all conclusions derived from the collected data. After the measurements described, we believe that all the potential threats of this study have been tackled.

5. Related work

In this section, we present previous review papers related to big data analytics in agriculture. To the best of our knowledge, there is a lack of literature review papers on this topic. This study contributes with an SLR to the limited view on known pitfalls and good practices described in primary studies regarding the application of big data analytics in agricultural domains. Specifically, we found six closely related articles pertinent to Big Data application in agriculture, as we present below.

Tantalaki *et al.* presented the existing challenges, opportunities, and promising areas of Big Data application (Tantalaki et al., 2019). They extracted information from 121 scientific papers, which were published between 2015 and 2018. This study identified nine challenges related to Big Data adoption in agriculture, such as data quality issues, data heterogeneity, data availability, data security holes, and privacy concern, spatiotemporal autocorrelation of data, the high dimensionality of data, non-stationarity of data and velocity, voluminous datasets, and data interpretation. Furthermore, Kamilaris *et al.* reviewed 34 different Big Data analysis papers in agriculture to identify the agricultural domain, the specific problem tackled, and the solution to overcome the problems

Table 28
Identified obstacles reported in the primary studies.

Obstacle	Number of papers	Studies
Vast amounts of data volume	10	(Perakis et al., 2020; Baseca et al., 2019), Yang et al. (2017), Pavòn-Pulido et al. (2017), McCarty et al. (2017), Bendre and Manthalkar (2019), Keswani et al. (2019), Tsipis et al (2020), Subahi and Bouazza (2020), Meena and Sujatha (2019)
The complexity of data processing and analysis	7	Ampatzidis et al. (2020), Kumar and Sharma (2020), Jeppesen et al. (2018), Zamora-Izquierdo et al. (2019), Saranya and Nagarajan (2020), Liu (2016), Vincent et al. (2019)
Insufficient of user knowledge	6	(Perakis et al., 2020; Laurent et al., 2019; Baseca et al., 2019), Muñoz et al. (2020), Cipolla et al. (2019), Bahri et al. (2020)
The trustworthiness of the data	5	Souza et al. (2020), Chen et al. (2019), Tsipis et al (2020), Subahi and Bouazza (2020), Meena and Sujatha (2019)
The complexity of the implemented application	5	(Baseca et al., 2019), Singh et al. (2020), Vincent et al. (2019), Campos et al. (2019), Bahri et al. (2020)
High-cost of the proposed framework	5	Ampatzidis et al. (2020), Alonso et al. (2020), Pavòn-Pulido et al. (2017), López-Riquelme et al. (2017), Ferrández-Pastor et al. (2016)
Issue of Data collection	5	Liu (2016), Keswani et al. (2019), Silva et al. (2014), Cañadas et al. (2017), Fawcett et al. (2019)
The high variety of data	4	(Baseca et al., 2019), Alonso et al. (2020), Meena and Sujatha (2019), Cipolla et al. (2019)
Poor data analytics result	4	(Laurent et al., 2019), Souza et al. (2020), Salamí et al. (2019), (Lee and Wang, 2020)
The complexity of the nodes network	4	Swain et al. (2020), Li (2019), Triantafyllou et al. (2019), Ferrández-Pastor et al. (2016)
The inefficiency of resources management	4	(Perakis et al., 2020), Silva et al. (2014), Singh et al. (2020), Subahi and Bouazza (2020)
Lack of supporting infrastructure	4	Alonso et al. (2020), Ferrández-Pastor et al. (2016), Tsipis et al (2020), Taneja et al. (2020)
High speed of Data velocity	4	(Perakis et al., 2020), Bendre and Manthalkar (2019), Keswani et al. (2019), Subahi and Bouazza (2020)
Lack of capacity of traditional DBMS	3	Yang et al. (2018), Liu (2016), Subahi and Bouazza (2020)
Limited computational capabilities of sensor nodes.	3	Triantafyllou et al. (2019), Ferrández-Pastor et al. (2016), Chen et al. (2015)
The noise of image data	3	McCarty et al. (2017), Dong et al. (2016), Ampatzidis and Partel (2019)
Network connection failure	3	Kaloxylos et al. (2014), Liu (2016), Tsipis et al (2020)
Limited access to research data	2	(Laurent et al., 2019), Kamilaris et al. (2018)
Time-consuming data processing	2	Ampatzidis et al. (2020), Laurent et al. (2020)
Latency	2	Triantafyllou et al. (2019), López-Riquelme et al. (2017)
Low system performance	2	Bendre and Manthalkar (2019), Laurent et al. (2020)
Credibility and reliability of the sensor reading	2	Tsipis et al (2020), (Lee and Wang, 2020)
Security	1	Alonso et al. (2020)
Lack of interoperability standard	1	Jeppesen et al. (2018)
Data compatibility	1	Jeppesen et al. (2018)
Issue of copyright and ownership data	1	Kaloxylos et al. (2014)
Standardized system interface	1	Kaloxylos et al. (2014)
Issue of data management	1	Popović et al. (2017)
Issue of model integration	1	Sawant et al. (2017)
Data Privacy	1	Sawant et al. (2017)
Unmanageable in-service delivery	1	Singh et al. (2020)
Imbalanced dataset	1	Taneja et al. (2020)
Issue of data transmission	1	Cipolla et al. (2019)
Poor user interface	1	Laurent et al. (2020)

Table 29
The obstacles that are mentioned two times or more in the selected paper and their correlation with publication year.

Obstacles	Year							Total of occurrence in primary studies
	2020	2019	2018	2017	2016	2015	2014	
Vast amounts of data volume	3	4		3				10
The complexity of data processing and analysis	3	2	1		1			7
Insufficient of user knowledge	3	3						6
The trustworthiness of the data	3	2						5
The complexity of the implemented application	2	3						5
High-cost of the proposed framework	2			2	1			5
Issue of Data collection		2		1	1		1	5
The high variety of data	1	3						4
Poor data analytics result	2	2						4
The complexity of the nodes network	1	2			1			4
The inefficiency of resources management	3						1	4
Lack of supporting infrastructure	3				1			4
High speed of Data velocity	2	2						4
Lack of capacity of traditional DBMS	1		1		1			3
Limited computational capabilities of sensor nodes.		1			1	1		3
The noise of image data		1		1	1			3
Network connection failure	1				1		1	3
Limited access to research data		1						2
Time-consuming data processing	2							2
Latency		1		1				2
Low system performance	1	1						2
Credibility and reliability of the sensor reading	2							2

Table 30
The proposed solutions reported in the primary studies.

NO	OBSTACLES AND CHALLENGES	PROPOSED SOLUTIONS	SOURCES
1	Vast amounts of data volume	Using vertical scalability method by enhancing computing/storage of the system	(Perakis et al., 2020)
		Using horizontal scalability method by performing High-Performance Computing resource	(Perakis et al., 2020), Bendre and Manthalkar (2019)
		Performing Big data platform	Yang et al. (2018), Pavòn-Pulido et al. (2017), McCarty et al. (2017), Meena and Sujatha (2019)
		Avoiding traditional database management system	Subahi and Bouazza (2020)
2	The complexity of data processing and analysis	Not mentioned	
3	The high variety of data	Not mentioned	
4	Insufficient user knowledge	Not mentioned	
5	The trustworthiness of the data	Not mentioned	
6	The complexity of the implemented application	Not mentioned	
7	High cost of the proposed framework	Not mentioned	
8	Issue of Data collection	Not mentioned	
9	Poor data analytics result	Not mentioned	
10	The complexity of the nodes network	Implement proper protocol communication	Swain et al. (2020), Ferrández-Pastor et al. (2016)
11	The inefficiency of resources management	Not mentioned	
12	Lack of supporting infrastructure	Not mentioned	
13	High speed of Data velocity	Using specific software to control data movement	(Perakis et al., 2020)
		Performing HPC or parallel computing	(Perakis et al., 2020), Bendre and Manthalkar (2019)
		Real-time processing	Keswani et al. (2019)
14	Lack of capacity of traditional DBMS	Not mentioned	
15	Limited computational capabilities of sensor nodes.	Not mentioned	
16	The noise of image data		
17	Network connection failure	Optimize network gateway transmission	Liu (2016)
18	Limited access to research data	Not mentioned	
19	Time-consuming data processing	Not mentioned	
20	Latency	Making local operation	Kaloxylos et al. (2014), Tsipis et al. (2020)
		Implement proper protocol communication	Swain et al. (2020), Ferrández-Pastor et al. (2016)
21	Low system performance	Not mentioned	
22	Credibility and reliability of the sensor reading	Not mentioned	
23	Security	Not mentioned	
24	Lack of interoperability standard	Not mentioned	
25	Data compatibility	Not mentioned	
26	Issue of copyright and ownership data	Not mentioned	
27	Standardized system interface	Not mentioned	
28	Issue of data management	Not mentioned	
29	Issue of model integration	Not mentioned	
30	Data Privacy	Not mentioned	
31	Unmanageable in-service delivery	Not mentioned	
32	Imbalanced dataset	Not mentioned	
33	Issue of data transmission	Not mentioned	
34	Poor user interface	Not mentioned	

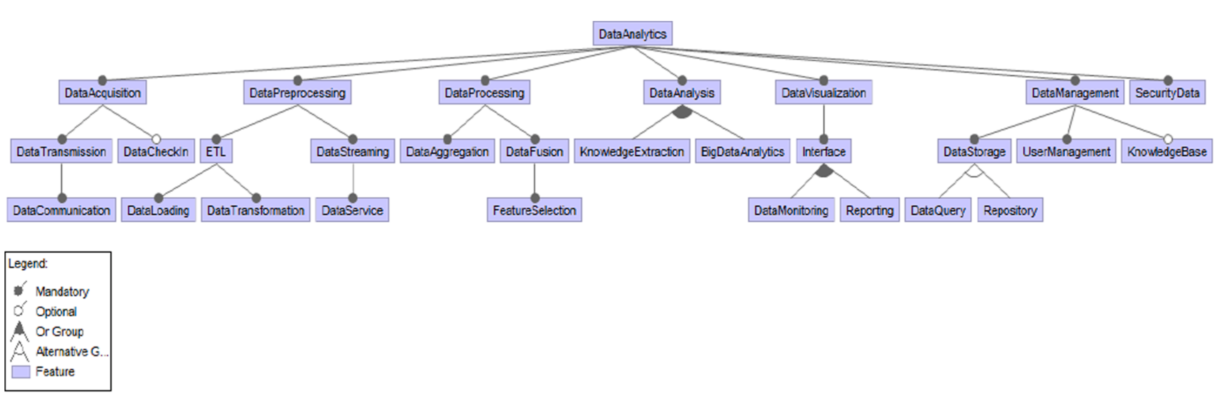

Fig. 23. Features diagram of a data analytics platform.

(Kamilaris et al., 2017). They also analyzed the tools, algorithm and data used, and dimensions of big data employed, and scale of use. The results of this study highlighted several obstacles of big data in agriculture, such as privacy issues, security, accuracy, and limited access to ground truth information. This paper also discussed various solutions to mitigate the obstacles, such as the creation of a regulatory framework on data ownership, and investments in cloud infrastructures. The application of some techniques such as data aggregation, data reduction and proper analysis also can contribute towards more user-friendly platforms.

On top of that, Wolfert *et al.* did a literature survey in the agriculture sector to identify Big Data Application challenges from the socio-economic perspective (Wolfert et al., 2017). They followed a systematic approach to the literature survey from January 2010 to March 2015, which resulted from 20 most relevant articles, 94 relevant articles, nine blogs, nine magazine articles, and 11 additional articles and web-items. This research stated that there are six significant issues in Big Data developments in agriculture: data-ownership (privacy and security issues), data quality issues, intelligent processing and analytics issues, sustainable integration of Big Data sources, business models attractiveness, and openness of platforms.

Weersink *et al.* also surveyed Big data analysis in agricultural sectors (Weersink et al., 2018). The purposes of their research were to review the benefits and drawbacks of Big Data. The focus of this research was on the policy issues that influenced agricultural activities. The Big Data system brings many benefits to help the farmers increase their productions and maintain the quality of the crop yield since the analytics results give the farmers early information about their field, and they could make some anticipated decisions if needed. However, the users encountered several issues to optimize the utilization of the Big Data system, which are data governance, a lack of knowledge and skills in how to handle and interpret the data at the farm level, and issues in the high cost of the system. Furthermore, they also suggested that the policymakers need to determine and release policies to control the Big Data system utilization in agricultural sectors. Coble *et al.* discovered that policymakers' critical role is to produce the rules to solve big data management issues in the smart farming system, including data privacy (Coble et al., 2018). Furthermore, the government should consider technology infrastructures, especially in rural areas that still lack data transfer rates since this drawback leads to the lack of technology tools utilization in rural areas. Finally, since vast amounts of data are still needed to boost Big Data performance, it would be much better if the government could integrate private data, government data, and specific data collection surveys to complement the big data system.

Basnet and Bang (Basnet and Bang, 2018) gave attention to applied sensing systems and data analytics in many agricultural sectors. This survey study provided information about the state-of-the-art of sensing system, and sensor data are raw materials for any data analytics approaches. To narrow down the scope of research in the enormous agricultural area, the authors only reviewed the use of sensors and information and communication technology (ICT) for In-field applications, that is, planting/raising and harvesting. In conclusion, the authors stated that agriculture is becoming more data-intensive, and these technologies, the sensing system, and the data analytics technique helped in the advancement of the area of agriculture. Positive trends also have been identified as the improvements of sensors and data analytics approaches, and it will bring more insights to face a wide variety of agricultural problems.

Avci et al. (Avci et al., 2020) presented the results of an SLR on the adoption of software architectures for big data platforms. The adopted big data software architectures for various domains were analyzed and synthesized. The study shows that big data software architectures are applied in various application domains. Several recurring common motivations were identified for adopting big data software architectures, such as supporting analytics process, improving efficiency, improving real-time data processing, reducing development costs, and enabling new kinds of services, including collaborative work. The previous related studies identified several obstacles and challenges of Big Data application in agriculture. However, none of them reviewed the architecture design that applied to handle and processing Big Data as well as the features that are offered in data analytics platforms. Therefore, this SLR study's primary motivation is the lack of systematic review regarding data analytics architectures and their offered features. On top of that, this is the first SLR that reviews data analytics technical factors, such as the adopted technologies, the inputted data, the applied algorithms, and the data analytics purposes and tasks.

6. Conclusion and future work

To the best of our knowledge, this is the first systematic literature review of 45 primary studies that systematically and explicitly discussed the features and obstacles of data analytics platforms, as well as the state-of-the-art data analytics in agricultural systems. This study only using journal articles as primary studies since it intended to review high-quality academic papers only. This study shows that the data analytics platform in agricultural systems had an increasing development in the last five years. A set of 33 features have been identified in this study, and most of the identified features had a strong correlation to the data management area, where data visualization was the most occurring feature. It implies that data visualization is an essential feature in the data analytics platforms in this review. This result is strengthened by another finding in the applied programming languages, whereby the most mentioned programming languages were JavaScript, PHP, HTML, CSS. Those languages are utilized to build a user-friendly visualization system. The results show that most studies have explicitly described the system architecture patterns where layered pattern is the most popular pattern to design the system from our observation. Most of the studies adopted one or more data analytics tasks and used various machine learning algorithms and statistical analysis to finish the tasks. In addition, Descriptive and Diagnostic analytics types are the most dominant data analytics type applied in the studies.

This study also found about 34 different obstacles reported in the primary studies. Furthermore, we have classified most of the obstacles as big data issues, where the possible solutions to those challenges are discussed. Therefore, we believe that those big data issues would be the rising problem in data analytics platforms. Furthermore, we also consider that this study will pave the way for further research and maturation of data analytics platforms, especially in big data analytics systems for the agriculture domain. We aim to create and build a new reference architecture for big data analytics platforms based on the identified obstacles and features from this study in our future work.

Declaration of Competing Interest

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

Appendix A. Primary studies (sources reviewed in the SLR)

[P1]	K. Perakis, F. Lampathaki, K. Nikas, Y. Georgiou, O. Marko, and J. Maselyne, "CYBELE – Fostering precision agriculture & livestock farming through secure access to large-scale HPC enabled virtual industrial experimentation environments fostering scalable big data analytics," *Comput. Networks*, vol. 168, pp. 1–10, 2020.
[P2]	A. Laurent, P. Kyveryga, D. Makowski, and F. Miguez, "A Framework for Visualization and Analysis of Agronomic Field Trials from On-Farm Research Networks," *Agron. J.*, vol. 111, no. 6, pp. 2712–2723, 2019, https://doi.org/10.2134/agronj2019.02.0135.
[P3]	M. Swain, R. Singh, A. K. Thakur, and A. Gehlot, "A Machine Learning Approach of Data Mining in Agriculture 4.0," *Int. J. Emerg. Technol.*, vol. 11, no. 1, pp. 257–262, 2020.
[P4]	C. C. Baseca, S. Sendra, J. Lloret, and J. Tomas, "A Smart Decision System for Digital Farming," *Agronomy*, vol. 216, no. 9, pp. 1–19, 2019.
[P5]	Y. Ampatzidis, V. Partel, and L. Costa, "Agroview: Cloud-based application to process, analyze and visualize UAV- collected data for precision agriculture applications utilizing artificial intelligence," *Comput. Electron. Agric.*, vol. 174, pp. 1–12, 2020, https://doi.org/10.1016/j.compag.2020.105457.
[P6]	A. Kumar and A. Sharma, "Socio-Sentic framework for sustainable agricultural governance," *Sustain. Comput. Informatics Syst.*, pp. 1–7, 2020.
[P7]	R. S. Alonso, I. Sittón-candanedo, Ó. García, J. Prieto, and S. Rodríguez-gonzález, "An intelligent Edge-IoT platform for monitoring livestock and crops in a dairy farming scenario," *Ad Hoc Networks*, vol. 98, pp. 1–23, 2020.
[P8]	A. Kamilaris, A. Anton, A. B. Blasi, and F. X. Prenafeta-boldú, "Assessing and mitigating the impact of livestock agriculture on the environment through geospatial and big data analysis," *Int. J. Sustain. Agric. Manag. Informatics*, vol. 4, no. 2, pp. 98–122, 2018, https://doi.org/10.1504/IJSAMI.2018.094809.
[P9]	J. Yang, M. Liu, J. Lu, Y. Miao, and M. A. Hossain, "Botanical Internet of Things: Toward Smart Indoor Farming by Connecting People, Plant, Data and Clouds," *Mob. Networks Appl.*, vol. 23, pp. 188–202, 2018.
[P10]	J. H. Jeppesen, E. Ebeid, R. H. Jacobsen, and T. S. Toftegaard, "Open geospatial infrastructure for data management and analytics in interdisciplinary research," *Comput. Electron. Agric.*, vol. 145, pp. 130–141, 2018, https://doi.org/10.1016/j.compag.2017.12.026.
[P11]	N. Pavòn-Pulido, J. A. Ló pez-Riquelme, R. Torres, R. Morais, and J. A. Pastor, "New trends in precision agriculture: a novel cloud-based system for enabling data storage and agricultural task planning and automation," *Precis. Agric.*, vol. 18, pp. 1038–1068, 2017, https://doi.org/10.1007/s11119-017-9532-7.
[P12]	B. Li, "Recommendation System of Crop Planting Books Based on Big Data," *Rev. La Fac. Agron. La Univ. Del Zulia*, vol. 36, no. 4, pp. 1166–1174, 2019.
[P13]	A. Triantafyllou, P. Sarigiannidis, and S. Bibi, "Precision Agriculture: A Remote Sensing Monitoring System Architecture," *Information*, vol. 10, pp. 1–25, 2019, https://doi.org/10.3390/info10110348.
[P14]	J. L. Mccarty, C. S. R. Neigh, M. L. Carroll, and M. R. Wooten, "Extracting smallholder cropped area in Tigray, Ethiopia with wall-to-wall sub-meter WorldView and moderate resolution Landsat 8 imagery," *Remote Sens. Environ.*, vol. 202, pp. 142–151, 2017, https://doi.org/10.1016/j.rse.2017.06.040.
[P15]	M. Bendre and R. Manthalkar, "Time series decomposition and predictive analytics using MapReduce framework," *Expert Syst. Appl.*, vol. 116, pp. 108–120, 2019.
[P16]	M. A. Zamora-izquierdo, J. A. Martı, and A. F. Skarmeta, "Smart farming IoT platform based on edge and cloud computing," *Biosyst. Eng.*, vol. 177, pp. 4–17, 2019.
[P17]	C. P. Saranya and N. Nagarajan, "Efficient agricultural yield prediction using metaheuristic optimized artificial neural network using Hadoop framework," *Soft Comput.*, vol. 24, pp. 12659–12669, 2020, https://doi.org/10.1007/s00500-020-04707-z.
[P18]	J. Dong *et al.*, "Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine," *Remote Sens. Environ.*, vol. 185, pp. 142–154, 2016.
[P19]	A. Kaloxylos *et al.*, "A cloud-based Farm Management System: Architecture and implementation," *Comput. Electron. Agric.*, vol. 100, pp. 168–179, 2014, https://doi.org/10.1016/j.compag.2013.11.014.
[P20]	J. A. López-riquelme, N. Pavón-pulido, H. Navarro-hellín, and F. Soto-valles, "A software architecture based on FIWARE cloud for Precision Agriculture," *Agric. Water Manag.*, vol. 183, pp. 123–135, 2017, https://doi.org/10.1016/j.agwat.2016.10.020.
[P21]	T. Popović, N. Latinović, A. Pešić, ˇarko Zeč́ević, B. Krstajić, and S. Djukanović, "Architecting an IoT-enabled platform for precision agriculture and ecological monitoring: A case study," *Comput. Electron. Agric.*, vol. 140, pp. 255–265, 2017, https://doi.org/10.1016/j.compag.2017.06.008.
[P22]	P. S. S. De Souza *et al.*, "Detecting abnormal sensors via machine learning: An IoT farming WSN-based architecture case study," *Measurement*, vol. 164, pp. 1–7, 2020.
[P23]	J. Liu, "Design and Implementation of an Intelligent Environmental-Control System: Perception, Network, and Application with Fused Data Collected from Multiple Sensors in a Greenhouse at Jiangsu, China," *Int. J. Distrib. Sens. Networks*, vol. 12, no. 7, pp. 1–10, 2016.
[P24]	B. Keswani *et al.*, "Adapting weather conditions based IoT enabled smart irrigation technique in precision agriculture mechanisms," *Neural Comput. Appl.*, vol. 31, pp. 277–292, 2019, https://doi.org/10.1007/s00521-018-3737-1.
[P25]	F. J. Ferrández-pastor, J. M. García-chamizo, and M. Nieto-hidalgo, "Developing Ubiquitous Sensor Network Platform Using Internet of Things: Application in Precision Agriculture," *Sensors*, vol. 16, pp. 1–20, 2016, https://doi.org/10.3390/s16071141.
[P26]	S. Silva, L. Alçada-almeida, and L. C. Dias, "Development of a Web-based Multi-criteria Spatial Decision Support System for the assessment of environmental sustainability of dairy farms," *Comput. Electron. Agric.*, vol. 108, pp. 46–57, 2014, https://doi.org/10.1016/j.compag.2014.06.009.
[P27]	J. Cañadas, J. A. Sánchez-Molina, F. Rodríguez, and I. M. Águila, "Improving automatic climate control with decision support techniques to minimize disease effects in greenhouse tomatoes," *Inf. Process. Agric.*, vol. 4, no. 1, pp. 50–63, 2017, https://doi.org/10.1016/j.inpa.2016.12.002.
[P28]	N. Chen, X. Zhang, and C. Wang, "Integrated open geospatial web service enabled cyber-physical information infrastructure for precision agriculture monitoring," *Comput. Electron. Agric.*, vol. 111, pp. 78–91, 2015, https://doi.org/10.1016/j.compag.2014.12.009.
[P29]	S. Sawant, S. S. Durbha, and J. Adinarayana, "Interoperable agro-meteorological observation and analysis platform for precision agriculture: A case study in citrus crop water requirement estimation," *Comput. Electron. Agric.*, vol. 138, pp. 175–187, 2017, https://doi.org/10.1016/j.compag.2017.04.019.
[P30]	E. Salamí, A. Gallardo, G. Skorobogatov, and C. Barrado, "On-the-Fly Olive Tree Counting Using a UAS and Cloud Services," *Remote Sens.*, vol. 11, no. 3, pp. 1–21, 2019, https://doi.org/10.3390/rs11030316.
[P31]	Y. Ampatzidis and V. Partel, "UAV-Based High Throughput Phenotyping in Citrus Utilizing Multispectral Imaging and Artificial Intelligence," *Remote Sens.*, vol. 11, no. 4, pp. 1–19, 2019, https://doi.org/10.3390/rs11040410.
[P32]	D. Fawcett, B. Azlan, T. C. Hill, L. Khoon, and J. Bennie, "Unmanned aerial vehicle (UAV) derived structure-from-motion photogrammetry point clouds for oil palm (Elaeis guineensis) canopy segmentation and height estimation," *Int. J. Remote Sens.*, vol. 40, no. 19, pp. 1–23, 2019, https://doi.org/10.1080/01431161.2019.1591651.
[P33]	S. Singh, I. Chana, and R. Buyya, "Agri-Info: Cloud Based Autonomic System for Delivering Agriculture as a Service Data Analytics," *Internet of Things*, vol. 9, pp. 1–16, 2020.
[P34]	X. Chen *et al.*, "A Model-Based Real-Time Decision Support System for Irrigation Scheduling to Improve Water Productivity," *Agronomy*, vol. 9, pp. 1–18, 2019.
[P35]	A. Tsipis, A. Papamichail, I. Angelis, G. Koufoudakis, G. Tsoumanis, and K. Oikonomou, "An Alertness-Adjustable Cloud/Fog IoT Solution for Timely Environmental Monitoring Based on Wildfire Risk Forecasting," *Energies*, vol. 13, pp. 1–35, 2020, https://doi.org/10.3390/en13143693.
[P36]	A. F. Subahi and K. E. Bouazza, "An Intelligent IoT-Based System Design for Controlling and Monitoring Greenhouse Temperature," *IEEE Access*, vol. 8, pp. 125488–125500, 2020, https://doi.org/10.1109/ACCESS.2020.3007955.
[P37]	M. Muñoz, J. D. Gil, L. Roca, F. Rodríguez, and M. Berenguel, "An IoT Architecture for Water Resource Management in Agroindustrial Environments: A Case Study in Almería (Spain)," *Sensors*, vol. 20, pp. 1–21, 2020, https://doi.org/10.3390/s20030596.
[P38]	C. Lee and Y. Wang, "Development of a cloud-based IoT monitoring system for Fish metabolism and activity in aquaponics," *Aquac. Eng.*, vol. 90, pp. 1–12, 2020.
[P39]	K. Meena and J. Sujatha, "Reduced Time Compression in Big Data Using MapReduce Approach and Hadoop," *Trans. Process. Syst.*, vol. 43, pp. 1–12, 2019.
[P40]	M. Taneja, J. Byabazaire, N. Jalodia, A. Davy, C. Olariu, and P. Malone, "Machine learning based fog computing assisted data-driven approach for early lameness detection in dairy cattle," *Comput. Electron. Agric.*, vol. 171, pp. 1–15, 2020.
[P41]	S. S. Cipolla, M. Maglionico, M. Masina, A. Lamberti, and I. Daprà, "Real Time Monitoring of Water Quality in An Agricultural Area With Salinity Problems," *Environ. Eng. Manag. J.*, vol. 18, no. 10, pp. 2229–2240, 2019.
[P42]	D. R. Vincent, N. Deepa, D. Elavarasan, K. Srinivasan, S. H. Chauhdary, and C. Iwendi, "Sensors Driven AI-Based Agriculture Recommendation Model for Assessing Land Suitability," *Sensors*, vol. 19, pp. 1–16, 2019.
[P43]	N. G. S. Campos, A. R. Rocha, R. Gondim, T. L. C. da Silva, and D. G. Gomes, "Smart & Green: An Internet-of-Things Framework for Smart Irrigation," *Sensors*, vol. 20, pp. 1–25, 2019, https://doi.org/10.3390/s20010190.
[P44]	

(*continued on next page*)

(*continued*)

	A. Laurent, X. Lyu, P. Kyveryga, D. Makowski, H. Hofmann, and F. Miguez, "Interactive web-based data visualization and analysis tool for synthetizing on-farm research networks data," *Res. Synth. Methods*, pp. 1–22, 2020, https://doi.org/10.1002/jrsm.1440.
[P45]	O. Bahri, A. Mourhir, and E. I. Papageorgiou, "Integrating fuzzy cognitive maps and multi-agent systems for sustainable agriculture," *Euro-Mediterranean J. Environ. Integr.*, vol. 5, pp. 1–10, 2020, https://doi.org/10.1007/s41207-020-0143-8.

Appendix B. Data extraction form

#	Extraction Element	Contents
General Information		
1	ID	
2	Title	
3	Year	
4	Authors	
5	Authors Affiliation	
6	Authors Organization	☐ University ☐ndustry B☐h
7	Repository	
8	Publication Venue	
9	Date of extraction	
10	SLR Category	☐ Include ☐xclude
Description		
11	Targeted Domain	
12	Stakeholders	
13	Motivation for Study	
14	Data Analytic Objective	
15	Data Analytic Platforms	
16	Programming Languages	
17	Applied Databases	
18	Software	
19	Identified Features	
20	Used Libraries	
21	Data Analytic Tasks	
22	Data Analytics Types	☐ Descriptive ☐ Predictive ☐ Diagnostic ☐ Prescriptive
23	Data analytics Algorithms	
24	Architecture Pattern	☐ Layers ☐ Pipes and Filters ☐ Broker ☐ Blackboard ☐ Model-View-Controller (MVC)
25	Data Access type	☐ Public Data ☐ Upon Request ☐ Private Data
26	Data format	
27	Identified Obstacles	
28	Provided Solutions	
Evaluation		
29	Personal Note	
30	Additional Note	
31	Quality Assessment	

References

Avci, C., Tekinerdogan, B., Athanasiadis, I.N., 2020. Software architectures for big data: a systematic literature review. Big Data Anal. 5 (5), 1–53.

Bacco, M., Barsocchi, P., Ferro, E., Gotta, A., Ruggeri, M., 2019. The digitisation of agriculture: a survey of research activities on smart farming. Array 3–4 (November), 1–11. https://doi.org/10.1016/j.array.2019.100009.

Baseca, C.C., Sendra, S., Lloret, J., Tomas, J., 2019. A smart decision system for digital farming. Agronomy 216 (9), 1–19.

Basnet, B., Bang, J., 2018. The state-of-the-art of knowledge-intensive agriculture: a review on applied sensing systems and data analytics. J. Sensors 2018, 1–13. https://doi.org/10.1155/2018/3528296.

Bourne, L., Walker, D.H.T., 2008. Project relationship management and the stakeholder circle. Int. J. Manage. Proj. Bus. 1 (1), 125–130. https://doi.org/10.1108/17538370810846450.

F. Buschmann, R. Meunier, H. Rohnert, P. Sommerlad, M. Stal, Pattern-Oriented Software Architecture: A System of Patterns (Volume 1). United States: John Wiley & Sons, Inc., 1996.

Catal, C., Giray, G., Tekinerdogan, B., 2022. Applications of deep learning for mobile malware detection: a systematic literature review. Neural Comput. Appl. 34 (2), 1007–1032.

Coble, K.H., Mishra, A.K., Ferrell, S., Griffin, T., 2018. Big data in agriculture: a challenge for the future. Appl. Econ. Perspect. Policy 40 (1), 79–96.

T. De Vylder, Feature Modelling: A Survey, a Formalism and a Transformation for Analysis, Belgium, 2011.

Gregor, D., Jaakko, J., Kulkarni, M., Lumsdaine, A., Musser, D., Schupp, S., 2005. Generic programming and high-performance libraries. Int. J. Parallel Program. 33 (213), 145–164. https://doi.org/10.1007/s10766-005-3580-8.

Gurbuz, H.G., Tekinerdogan, B., 2018. Model-based testing for software safety: a systematic mapping study. Softw. Qual. J. 26, 1327–1372. https://doi.org/10.1007/s11219-017-9386-2.

Husamaldin, L., Saeed, N., 2020. Big data analytics correlation taxonomy. Information 11 (17), 1–12.

Kamilaris, A., Kartakoullis, A., Prenafeta-boldú, F.X., 2017. A review on the practice of big data analysis in agriculture. Comput. Electron. Agric. 143, 23–37. https://doi.org/10.1016/j.compag.2017.09.037.

Kitchenham, B., Brereton, O.P., Budgen, D., Turner, M., Bailey, J., Linkman, S., 2009. Systematic literature reviews in software engineering – a systematic literature review. Inf. Softw. Technol. 51, 7–15. https://doi.org/10.1016/j.infsof.2008.09.009.

Köksal, Ö., Tekinerdogan, B., 2017. Obstacles in data distribution service middleware: a systematic review. Futur. Gener. Comput. Syst. 68, 191–210. https://doi.org/10.1016/j.future.2016.09.020.

Laurent, A., Kyveryga, P., Makowski, D., Miguez, F., 2019. A framework for visualization and analysis of agronomic field trials from on-farm research networks. Agron. J. 111 (6), 2712–2723. https://doi.org/10.2134/agronj2019.02.0135.

Laurent, A., Lyu, X., Kyveryga, P., Makowski, D., Hofmann, H., Miguez, F., 2021. Interactive web-based data visualization and analysis tool for synthetizing on-farm research networks data. Res. Synth. Methods 12 (1), 62–73.

Lee, C., Wang, Y., 2020. Development of a cloud-based IoT monitoring system for Fish metabolism and activity in aquaponics. Aquac. Eng. 90, 1–12.

Pathak, H.S., Brown, P., Best, T., 2019. A systematic literature review of the factors affecting the precision agriculture adoption process. Precis. Agric. 20, 1292–1316. https://doi.org/10.1007/s11119-019-09653-x.

Perakis, K., Lampathaki, F., Nikas, K., Georgiou, Y., Marko, O., Maselyne, J., 2020. CYBELE – Fostering precision agriculture & livestock farming through secure access to large-scale HPC enabled virtual industrial experimentation environments fostering scalable big data analytics. Comput. Networks 168, 1–10.

T. Popovic´, N. Latinovic´, A. Pešic´, ˇarko Zecˇevic´, B. Krstajic´, and S. Djukanovic´, "Architecting an IoT-enabled platform for precision agriculture and ecological monitoring : A case study, Comput. Electron. Agric., vol. 140, pp. 255–265, 2017, doi: 10.1016/j.compag.2017.06.008.

Provost, F., Fawcett, T., 2013. Data Science for Business: What You Need to Know About data Mining and data-Analytic Thinking. O'Reilly Media Inc, Sebastopol.

Röling, N., Jiggins, J., Hounkonnou, D., VanHuis, A., 2014. Agricultural research –from recommendation domains to arenas for interaction. Outlook Agric. 43 (3), 179–185. https://doi.org/10.5367/oa.2014.0172.

Tantalaki, N., Souravlas, S., Roumeliotis, M., 2019. Data-driven decision making in precision agriculture: the rise of big data in agricultural systems. J. Agric. Food Inf. 20 (4), 1–37. https://doi.org/10.1080/10496505.2019.1638264.

Tummers, J., Kassahun, A., Tekinerdogan, B., 2019. Obstacles and features of farm management information systems: a systematic literature review. Comput. Electron. Agric. 157, 189–204. https://doi.org/10.1016/j.compag.2018.12.044.

United Nations (UN), Goal 2: Zero Hunger. Sustain. Dev. Goals. https://www.un.org/sustainabledevelopment/hunger/.

Van Geest, M., Tekinerdogan, B., Catal, C., 2021. Computers in Industry Design of a reference architecture for developing smart warehouses in industry 4.0. Comput. Ind. 124, 1–23. https://doi.org/10.1016/j.compind.2020.103343.

Weersink, A., Fraser, E., Pannell, D., Duncan, E., Rotz, S., 2018. Opportunities and challenges for big data in agricultural and environmental analysis. Annu. Rev. Resour. Econ. 10 (1), 19–37.

Wolfert, S., Ge, L., Verdouw, C., Bogaardt, M., 2017. Big data in smart farming – a review. Agric. Syst. 153, 69–80. https://doi.org/10.1016/j.agsy.2017.01.023.

X. Zhou, Y. Jin, H. Zhang, S. Li, X. Huang, A map of threats to validity of systematic literature reviews in software engineering, in: 23rd Asia-Pacific Software Engineering Conference, 2016, pp. 153–160, doi: 10.1109/APSEC.2016.62.