

## Review

# Agricultural Big Data Architectures in the Context of Climate Change: A Systematic Literature Review

Ania Cravero <sup>1, \*</sup>, Ana Bustamante <sup>2</sup>, Marlene Negrier <sup>2</sup> and Patricio Galeas <sup>1</sup>

<sup>1</sup> Department of Computer Science and Informatics, Centre of Excellence for Modelling and Scientific Computing, Universidad de La Frontera, Temuco, Chile; ania.cravero@ufrontera.cl; patricio.galeas@ufrontera.cl

<sup>2</sup> Department of Computer Science and Informatics, Universidad de La Frontera, Temuco, Chile. Postgraduate Student from UNADE; ana.bustamante@ufrontera.cl, marlene.negrier@ufrontera.cl

\* Correspondence: ania.cravero@ufrontera.cl

**Abstract:** Climate change is currently one of the main problems facing agriculture to achieve sustainability. It causes situations such as drought, increased rainfall, and increased diseases, causing a decrease in food production. In order to combat these problems, Agricultural Big Data contributes with tools that allow improving the understanding of complex, multivariate, and unpredictable agricultural ecosystems through the collection, storage, processing, and analysis of vast amounts of data from diverse heterogeneous sources. This research aims to discuss the advancement of technologies used in Agricultural Big Data architectures in the context of climate change. The study aims to highlight the tools used to process, analyze, and visualize the data and discuss the use of the architectures in the crop, water, climate, and soil management, especially to analyze the context, whether it is in Resilience Mitigation or Adaptation. The PRISMA protocol guided the study, finding 33 relevant papers. Despite the advances in this line of research, few papers were found that mention the components of the architectures, in addition to the lack of standards and the use of reference architectures, which allow the proper development of Agricultural Big Data in the context of climate change.

**Keywords:** big data; architecture; agriculture; climate change; systematic literature review

## 1. Introduction

Climate change is currently harming the agricultural sector, causing situations such as sea-level rise that has encroached on the land and increased precipitation that has led to the extension of floodplains and the reduction of existing croplands [1]. On the other hand, rising temperatures have harmed crop production and sustainability, both nationally and globally [2]. According to Nguyen et al., in the future, warming trends are likely to reduce agricultural production yields unless there is a compelling adaptation [2].

Crop losses will increase due to rainfall variability, extreme heat, and flooding, causing resilience problems in arable areas. Even high rainfall can hinder harvesting and reduce crop quality [3]. Therefore, an adaptation to resilience for agricultural systems should be a policy priority [2]. Furthermore, climate change scenarios suggest that warming will generally reduce the sustainability of major staple crops such as maize and wheat, with higher yield losses in tropical regions [4].

On the other hand, is the growing demand for food and the problems the world faces, like the ability to appropriate new lands, new waters, or new fishing grounds. Furthermore, in addition to the rapid transformation of the earth's natural systems, climate change, associated with the increase in more extreme temperatures and precipitation, can alter the relations between crops, plagues, pathogens, and weeds, exacerbating several trends, like reducing pollinating insects, increasing water shortages and ozone concentrations at ground level, reducing fishing [4].

The pressures of the growing population and urbanization that threaten food safety have a more significant impact in the context of climate change [2]. In light of climate

change, guaranteeing food for the global population is no easy task due to its enormous impact on agricultural production [5].

The research program of the Consultative Group on International Agricultural Research (CGIAR) on climate change, agriculture, and food safety has proposed a framework with four interconnected areas of action for the reconfiguration of food systems: redirecting agricultural trajectories; increasing the resilience of all the agents involved in rapid change; minimizing the environmental footprint of food systems from the point of view of climate change; and realigning the facilitators of change [6]. Through this framework, some synergies between food safety, adaptation, and mitigation are feasible, called "climate-smart agriculture," which is key to this transformation of the food system [6]. The primary goal is to generate actions in agriculture to increase productivity sustainably, improve resilience and, therefore, adaptation; reduce or eliminate greenhouse gases, which is mitigation; and improve the achievement of international food safety objectives [4].

Because of these actions, it is necessary to analyze and better understand the complexities of multivariate and unpredictable agricultural ecosystems need to be analyzed and better understood [7]. The emerging digital technologies mentioned, such as machine learning, the Internet of Things (IoT), and Big Data, contribute to this understanding through the pursuit and continuous measurement of various aspects of the physical surroundings, producing large amounts of data at an unprecedented rate [8]. This involves compiling, storing, processing, modeling, and analyzing enormous amounts of data coming from various heterogeneous sources [7].

The earth is a complex dynamic system like none have ever studied before. Thus, the progress of Big Data in climate science has been slower than the success of Big Data in other fields like biology or advertising [7]. The slow progress has been a stumbling block since climatology has become one of the fields richest in data in terms of volume, speed, and variety [7]. Given that the earth is a vast dynamic system, the learning process is developing, and therefore the understanding its systems is not as advanced as in other fields that use Big Data [9]. Although observation methods originally collected climate data, and the readings had to be corroborated manually, now the data are increasingly younger. In most cases, the data are available in real-time [9]. The use of sensors and satellites has revolutionized the process of climate data collection and study, which is a significant contribution.

E. Hassani et al. [10] present some advances by using Big Data to solve energy efficiency problems, smart agriculture, smart urban planning, weather forecasting, and natural disaster management. Agricultural Big Data addresses various analysis issues in smart agriculture, such as weather forecasting and monitoring crop plagues and animal diseases. Various algorithms, approaches, and diverse techniques are used [11].

Agricultural Big Data is still an area of development that can improve productivity in farming operations by using: precision agriculture, data-based agriculture, smart agriculture, and predictive modeling [12], facilitating intelligent agricultural solutions from a climate point of view in the face of climate change and the effects of natural disasters [13].

One crucial challenge is the design of Big Data architectures, as this is not an easy task [14], [15], and [16]. It will be even more complex to build architectures for Agricultural Big Data in climate change [8], [13]. A. del Pozo et al. propose a multidisciplinary approach, where agronomists, physiologists, molecular biologists, sociologists, economists, and other social scientists must contribute with specific tools to understand complex agricultural systems. This approach can be applied in the different regions of the world, where climate change threatens the future of sustainable agriculture in danger [17]. On the other hand, the dynamism and complexity of climate issues that entail an interdisciplinary approach and the interweaving of various disciplines must not be neglected. According to Sebestyen et al., a System of Systems is urgently required, called "climate computation" [18]. According to the authors, to resolve any of the challenges related to climate change, "it is essential to obtain and integrate knowledge in an entire series of

systems that serve as the basis for the design of solutions that take the complex and uncertain nature of individual systems and their interrelations into account" [18].

In this sense, Agricultural Big Data in climate change is one of the critical answers to these needs. This research aims to discuss the advancement of technologies used in Agricultural Big Data architectures in the context of climate change. The study aims to highlight the tools used to process, analyze, and visualize the data and discuss the use of the architectures in the crop, water, climate, and soil management, especially to analyze the context, whether it is in Resilience Mitigation or Adaptation. We provide a schema that summarizes the data, allowing researchers to decide which Big Data solution to use depending on climate change and agriculture issues. In addition, this document allows us to identify research gaps and opportunities in this area.

Consequently, it serves as a comprehensive basis and facilitator for future research. To this end, a Systematic Literature Review (SLR) was conducted, applying the PRISMA protocol [11]. As a result, a set of 33 articles explaining the use of Agricultural Big Data in the context of climate change was selected.

The paper presents the following sections. Section 2 describes the main characteristics of Agricultural Big Data and climate change in general terms. Section 3 details the methodology and results of the paper selection process. Section 4 discusses the main results of the study. Section 5 includes a discussion of the main challenges of Agricultural Big Data in the context of climate change. Section 6 explains the limitations of the study. Finally, section 7 presents the conclusions, implications, and future research.

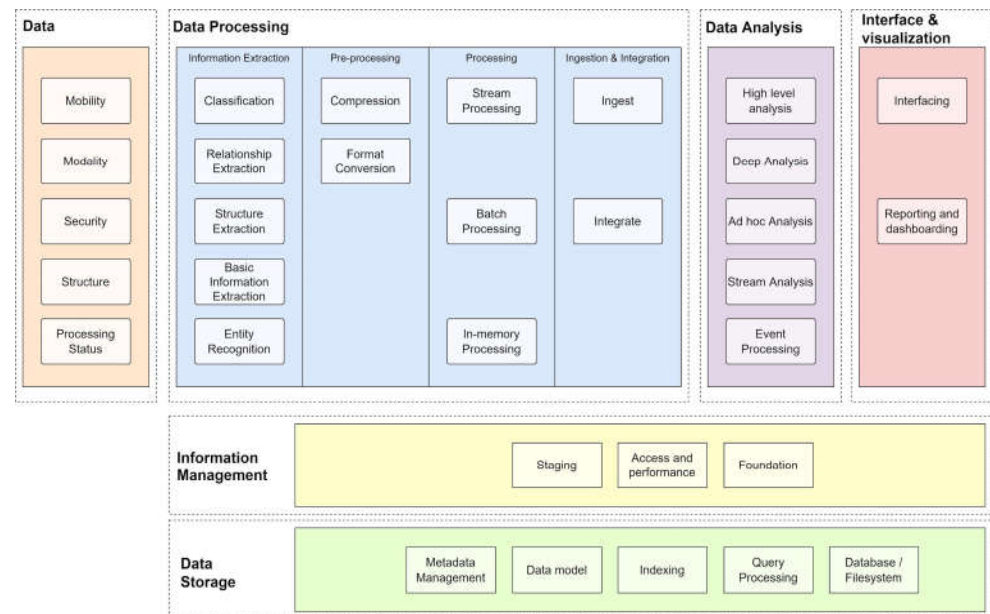
## 2. Background

### 2.1. Agricultural Big Data

Big Data is defined in four dimensions (4 Vs). First, it refers to the enormous volume of data that is generated, stored, and processed. Second, it also refers to the high velocity of data transmission in interactions, and the rates at which data are generated, collected and exchanged. Thirdly it relates to the variety of data formats and structures (structured, semi-structured, and unstructured) which result from the heterogeneity of data sources [19]. Finally, the fourth dimension is veracity, which refers to the ability to validate the quality of the data used in the analyses.

Big Data analysis tools enable data scientists to discover correlations and patterns by analyzing massive quantities of data from different sources. In recent years the science of Big Data has become an essential modern discipline for data analysis [20]. It is considered an amalgam of classic fields like statistics, artificial intelligence, mathematics, and informatics with its sub-disciplines, including database systems, ML, and distributed systems [21].

The Big Data Ecosystem handles the evolution of data, models, and supports infrastructure throughout its life cycle; it is a whole set of components, or architecture, storing, processing, and visualizing data and delivering results to guide applications [22], [23]. The Framework Architecture of Big Data in Figure 1 includes data storage, information management, data processing, data analysis, and interface and visualization components.



**Figure 1.** The General architecture of Big Data.

As shown in Figure 1, the Big Data process starts with identifying the sources from which valuable data are extracted [24]. Next, the data are stored in one of the designed data models, depending on whether the data are structured. In the following step, the data are classified and filtered according to the type of analysis required. In the Processing stage, it is defined whether it will be by Batch or Stream, in addition to the memory-based storage [25]. The classified data are analyzed using appropriate tools, for example DL [26], Ad hoc analysis [25], and data science in general [27]. The data obtained must be presented through some visualization tool. Finally, the data are analyzed by the decision-makers [23].

Big Data has been used to improve various aspects of agriculture, such as knowledge about weather and climate change, land, animal research, crops, soil, weeds, food availability and security, biodiversity, farmers' decision-making, farmers' insurance and finance, and remote sensing [7]. It is also used to create platforms which allow the actors of the supply chain access to high quality products and processes; tools to improve yields and predict demand; and advice and guidance to farmers based on the response capacity of their crops to fertilizers, leading to better fertilizer use.

## 2.2. Climate Change

Climate change is one of the most significant challenges globally at present. It is defined as the significant changes in the mean values of weather elements, like precipitation and temperature, which have been calculated for a long time [28]. Recent decades indicate that significant global climate changes resulted from an increase in human activities that altered the composition of the global atmosphere [29]. The concentration of greenhouse gases, such as methane (CH<sub>4</sub>), carbon dioxide (CO<sub>2</sub>), and nitrous oxide (N<sub>2</sub>O), has increased 150%, 40%, and 20%, respectively since 1750 [30]. CO<sub>2</sub> emissions, representing the maximum proportion of greenhouse gases, increased to 36,140 million metric tons in 2014 from the 22,150 million metric tons in 1990. The mean global temperature has increased at an average rate of 0.15-0.20°C per decade since 1975, and for 2021 between 1.4 and 5.8 °C. Greenhouse gas (GHG) emissions, in particular, CO<sub>2</sub> from fossil fuel combustion and GHG not related to CO<sub>2</sub>, such as N<sub>2</sub>O, CH<sub>4</sub>, and CFC, contribute to global warming [5].

In the 21st century, climate change will become a severe problem and both developed and developing countries will face negative externalities [5]. Agriculture is the sector most vulnerable to climate change due to its enormous size and sensitivity to weather parameters, which has an enormous economic impact [5]. The changes in climate phenomena,

such as temperature and precipitation, significantly affect crop yields and its sustainability. The effect of increasing temperatures, varying precipitation and CO<sub>2</sub> fertilization varies by crop, location and magnitude of parameter change.

According to the influence of climate variables, crop productivity depends on several variables: the type of crop and its ability to adapt, the climate scenario, and the effect of fertilization with CO<sub>2</sub> [31].

Nguyen et al. conducted a study on the effects of climate change perceived by farmers in different countries [2]. The authors report effects such as the changes in precipitation patterns and atmospheric temperatures as being most frequently perceived by farmers in Asia. On the other hand, the authors indicate that farmers have noted a trend in increasing temperature and changes in precipitation. Some farmers perceive a decrease, whereas others show an increase in precipitation. In some places, the changes in precipitation have been described as more erratic, unpredictable, and untimely. The authors conclude that farmers' perceptions on climate change have been consistent with the scientific data observed. In addition, the expected impacts of climate change identified were frequent droughts and floods, in addition to cyclones, heat waves, droughts, cold snaps, and soil erosion [2].

According to Malhi, the increased intensity and frequency of precipitation also affects soil erosion and will have more adverse consequences if GHG increases. The precipitation anomalies have detrimental effects on agriculture, mainly in developing countries. In addition to affecting crop yields, they significantly influence crop surfaces [5].

The studies identified by Nguyen et al. showed that farmers had used different technologies to adjust their land use options and land management practices to climate change. The main adaptation technologies are used for soil management, water management, crop management, and changes in land use options [2]. Some examples of these technologies include crop residues such as topsoil, minimization of plowing and reduction of tillage as soil conservation strategies, deep plowing during the rainy season, rain-water collection for use during the dry season, and others.

According to the Intergovernmental Panel on Climate Change [30], adaptation is adjusting to current or projected climate change and its consequences to minimize damage or take good opportunities. On the other hand, resilience is a process of recovery of the previous condition after facing any adverse effect or the adaptive, absorptive, and transformative capacity of a social unit, such as an individual community or state, to cope with any natural disaster. The resilience perspective is an understanding of the adaptive capacity of a system [32]. Adaptation in agriculture is crucial as it is highly dependent on climate, and some farmers have already started to employ some strategies [2]. On the other hand, climate change mitigation means avoiding and reducing GHG emissions, which trap heat into the atmosphere to prevent the planet from becoming extremely hot. The more we mitigate climate change now, the easier it will be to adapt to changes we can no longer avoid.

The potential of Big Data can be a crucial approach to managing the adverse effect of climate change globally, as it enables the development of context-specific adaptation strategies to improve resilience to climate change and appropriate analytics to assist with mitigation [32]. With Big Data, it is possible to display information regarding upcoming problems, current ones, and the stages of recovery from the adverse effect of climate change. On the other hand, it is a tool for policymakers, managers, and stakeholders to take necessary actions during and after a disaster, such as an early warning system, weather forecasting, emergency evacuation, immediate responses, and aid distribution [32].

To maximize the benefit of Big Data applied to climate change, the authors conclude that there are still challenges to solve, such as data collection, architecture design, ethical and political anomalies, poor team coordination, privacy, and accuracy. They recommend implementing adequate infrastructure, technologies, tools, and knowledge to ensure the proper use of Big Data in climate change [32].



### 3. Methodology

The research method used for this paper was SLR. For the selection of articles, the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) method was applied, which contains four stages: identification, screening, eligibility, and inclusion [11]. To define the objectives and research questions (RQ), we used the methodology proposed by Kitchenham and Charters [33]. The analysis stems from a series of RQ such as (1) what kind of problems are solved using Agricultural Big Data in the context of climate change, (2) which components are used in Agricultural Big Data architectures, (3) what technological tools are used to implement agricultural Big Data architecture, and (4) what are the challenges and trends to implement Agricultural Big Data.

The search string was constructed as follows: (a) from the RQs we obtained keywords; (b) we applied the population–intervention–comparison–results–context criterion to frame the RQs (PICOC [33]) criterion. According to Kitchenham and Charters [33], the population corresponds to an application area, namely, Agricultural Big Data. On the other hand, the intervention deals with climate change in agriculture. In our case, it is not appropriate to apply the comparison. The results obtained are the identified problems, the corresponding solutions, and the technologies used to implement Agricultural Big Data.

Moreover, we accessed Scopus, Springer, ACM, IEEE, MDPI, and Web of Science. The search strings used in all data sources were “Big Data”, “climate change” OR “climate-change”, AND “agriculture OR farm”. In order to include all relevant studies, a second search string was used, which includes the keywords, “global-change” OR “Land use change” OR Drought AND “Big Data” AND “farm OR agriculture”. We looked for these strings in the articles’ titles, keywords, and abstracts. In addition, we considered only scientific articles and conferences in English published from 2015 to 2021.

We identified 5108 potential articles in ACM, 1000 in Springer, 357 in the Web of Science database, 38 in IEEE, and 2 in MDPI. We excluded books, book chapters, working papers, and press articles. This final selection resulted in 33 relevant articles.

In the identification phase, 6505 articles were examined. This examination was followed by a screening where the duplication criterion was applied. This screening resulted in 5842 identified articles. The abstracts of these articles were then reviewed, and we checked whether or not they contained the nexus between agriculture, Big Data, and climate change. After eliminating the non-relevant articles, 356 articles remained for detailed analysis. This analysis consisted in reviewing the entire article to ensure that it included a description of the Agricultural Big Data process and climate change. Most of the excluded articles had theoretical, technological, or experimental issues. Some excluded articles poorly discussed the nexus between agriculture, Big Data, and climate change. Finally, the SLR was based on 33 relevant articles. Figure 2 summarizes the steps for relevant article selection, and Figure 3 presents the selected papers by year.

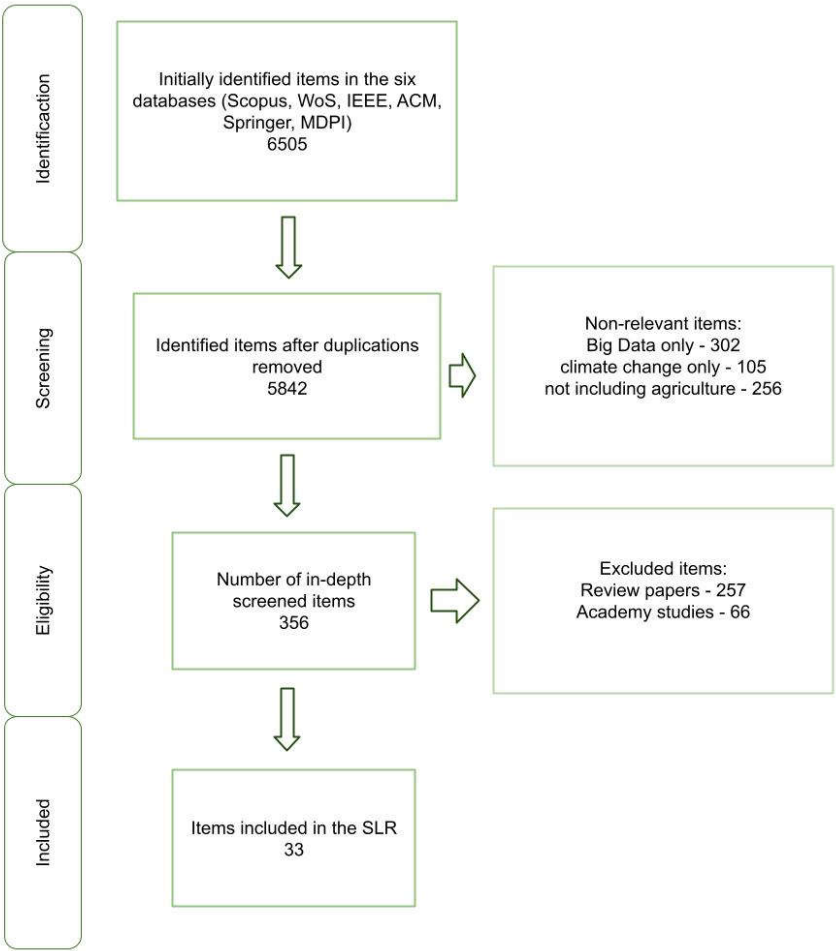


Figure 2. The Selection process of relevant papers.

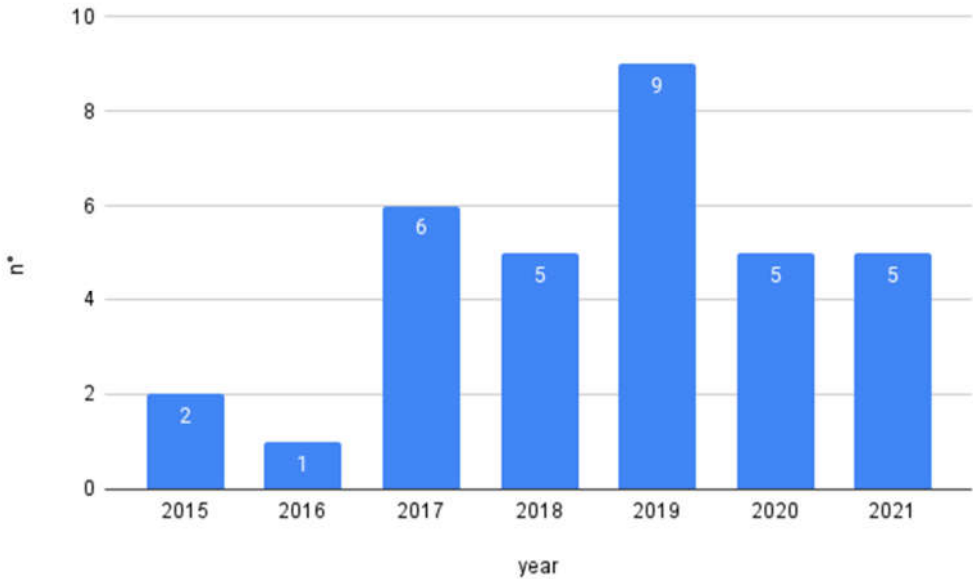


Figure 3. Results by year.

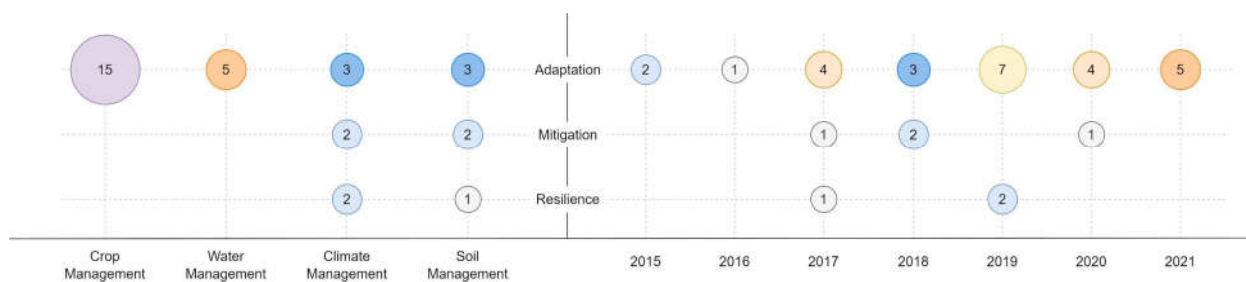
## 4. Results

### 4.1. Agricultural Big Data in the Context of Climate change

This section describes the proposals to solve problems that farmers face in water management, climate management, crop management, and soil management. The issues have been ordered according to the type of action declared by the IPCC. Figure 3 shows the number of papers that contribute to systems in Agricultural Big Data in the context of climate change for each type of problem identified. Thus, 15 people present a solution for crop management, and adaptation is the action carried out. On the other hand, only one paper presents a Big Data solution for soil management, and the action is resilience.

The following sub-sections present a summary of the proposals.

On the other hand, it is observed that since 2015, the concern about adaptation has been constant. The same does not occur with mitigation. Therefore it should be a concern in research topics. In the case of resilience, many adaptation proposals help in this regard.



**Figure 4.** Map of selected papers by type of problem vs IPCC action.

The following subsections present a summary of the proposals.

#### 4.1.1. Big Data for water management

(a) Adaptation: According to Figure 4, five papers selected describe actions for Adaptation in water management. In Coro et al., they use greenhouse effect data flows from a Big Data system that comes from data collections from AquaMaps and NASA. They also use environmental parameters processed through a computation platform in the cloud to generate intermediate scenarios and new informative parameters to show climate change patterns in 10 marine areas. The authors conclude that the Mediterranean Sea may have an independent response to climate change compared to other areas, affecting fishing and other inputs [34]. Senay et al. used Landsat and meteorological Big Data to quantify and understand the relative importance of water management and climate variability on crops. They also used this data to make decisions in water management, hydrological processes, and economic transactions [35].

Alencar et al. describe the decision-making support tools the private sector is being developed in close collaboration with all levels of government, including hydrographic river basin management authorities that provide environmental data. For example, the cumulative effects on water could be seen as reducing flow rates in a concrete river due to climate change and the irrigation and extraction of municipal and industrial water. Generally, Big Data linked the analytical capabilities to help promote a paradigm change in the modeling, analyzing, and managing of hydrographic river basins. As a result, solutions shift from reactive to proactive [36].

On the other hand, Kandekar et al. conclude that the change in surface water resources is of great importance for climate change and the development of agriculture in the semi-arid region. The authors used a cloud-based platform that processes data and geoscience analysis. Google Earth Engine (GEE) supports the processing of remote sensing images with many excellent algorithms for remote sensing image processing. From the Big Data system, vector data on the water mass is obtained, which provides a base for the later extraction of hydrological elements [37].



Finally, Reddy and Sureshababu tested an early prediction BigData system for precipitations. Farmers who live in regions prone to natural disasters like floods use the system to make decisions in managing their crops and water, producing high levels of benefits and production for these farmers [38].

#### 4.1.2. Big Data for climate management

(a) Mitigation: Seven papers selected describe actions of Adaptation, Mitigation, and Resilience in climate management. For Mitigation, the work by Semlali et al. uses Big Data to analyze harmful gas emissions and monitor air quality using remote sensing techniques [39]. Ma and Nie developed a smart weather service model based on Big Data that includes the collaboration of resources, techniques, and services to optimize weather forecast processes, helping farmers in decision-making [40].

(b) Resilience: For the case of Resilience, Mousannif and Zahir looked for solutions to quantify the impact of climate change on farmers' quality of life to reinforce the climate resistance and improve farmers' livelihoods. They applied advanced data analyses to provide a deep understanding of the interrelated climate, environmental, social, cultural, economic, institutional, and political factors that exacerbate the differentiated impacts of climate change. As a result, the system can decide about irrigation, types of crops to plant, and actions to improve crop production to make the most vulnerable communities more resistant to climate change [41]. In addition, Li and Liao developed a Big Data platform for early warning of agricultural weather disasters, including floods, droughts, warm and dry winds, low temperatures, typhoons, hail, and continuous rain [42].

(c) Adaptation: For the case of the Adaptation, Jain and Jain explored several application domains that could benefit from weather forecasting using Big Data analyses, in addition to the challenges to be faced when taking advantage of the weather forecast in these domains [43]. On the other hand, Gad and Manjunatha developed a climate data storage model for Big Data to identify climate patterns useful for agricultural fields, climate change studies, and contingency plans in extreme climate conditions [44]. Mangal and Rajesh developed a Big Data framework to anticipate the potential effects of many climate scenes, including the social, economic, and geographic diversity of the parties involved and the long-term time horizons and irreversible effects [45].

#### 4.1.3. Big Data for crop management

(a) Adaptation: Fifteen papers contribute with actions in Adaptation. First, Gulzar et al. determined the optimal health of crops and analyzed the impact of climate change on food safety. The authors use a taxonomy to provide recommendations [46]. Second, Delgado et al. developed a WebGIS framework to connect generators of specific local data from crop sites called smart farms to a regional and global view of agriculture. They use artificial intelligence, IoT, drones, robots, and Big Data as the foundation for a global "Digital Twin" that contributes to developing specific conservation and management of crop sites [47]. Third, Lopez et al. propose a strategy of merging multi-domain data to support data analysis tasks in agricultural contexts because many crops will no longer adapt to the conditions of a specific territory [48]. Finally, Manogaran and Lopez offer a climate change detection algorithm based on spatial autocorrelation to control changes in the seasonal climate to prevent diseases.

On the other hand, Fenu and Mallocci used Big Data Analytics and Machine Learning to identify potato late blight, considered one of the most devastating diseases anywhere. They use weather parameters provided by ARPAS weather stations [11]. Alex and Kanavalli use Big Data to prove and measure the harvest growth in the agricultural field. That helps predict weather events, and it is also used to calculate the resources sufficient for the crop in the climate change context [49]. Lokers et al. developed an infrastructure that enables efficient discovery and unified consultation of agriculture and forest resources using data connected through semantic technologies. The Big Data system made it possible to semantically process, structure, and consult metadata [50]. Tseng et al. and

Lambrinos used IoT and Agricultural Big Data to monitor the environmental factors on a farm [51], [52]. In Moon et al., they use Big Data to extract the information from each tiny region, such as the types of crops present, the types of food crops, fruit types, the types of medicinal crops, vegetable types, and the changing trends in the amount of production of each crop, to increase the future benefits of agricultural households by providing information [53].

Several authors used satellite remote sensing. For example, Lu followed the growth rate and characteristics of the crop, collecting customized data on the lands to indicate where and how to plant [54]. Sarvia et al. propose a service based on free remote sensing data to support the strategies of insurance companies. The service integrates MODIS and Sentinel-2 data to describe the mean trends of crop yields at the regional level [55]. Ye et al. designed a Big Data RDCRMG architecture to monitor dryness [56].

#### 4.1.4. Big Data for soil management

(a) Adaptation: For the case of the Adaptation, Bestelmeyer et al. developed Big Data to analyze landscapes, using temporal data sets in a grid or at a regional level together with local information on changes in the pastures that grow on the farms. They discovered that there are places where the pastures have decreased, leading to livestock problems [57]. Hu et al. use Big Data to include the recent historical changes in land cover and its interrelations with environmental processes to better support the land use management strategies in climate change [58].

(b) Mitigation: For the case of Mitigation, Fernández\_Getino et al. use Big Data to analyze the connectivity values associated with healthy ecosystems and with high levels of biodiversity, which can ensure better soil protection, favoring sustainability, implying sustainability, and a more significant C sequestration potential in the soil. Furthermore, they use vulnerability and risk maps to identify the areas most prone to habitat degradation and fragmentation [59]. On the other hand, Kamilaris et al. developed the software AgriBigCAT, which uses geophysical information, geospatial analysis, and Big Data analysis to evaluate the impact of the agricultural livestock sector on the environment by examining possible contamination of the land, water, and air due to manure [60].

(c) Resilience: Concerning Resilience, Li et al. examine the land-based ecological and environmental change on analyses of change trends of land cover and the normalized difference vegetation index. The authors indicate that ecological restoration is undergoing ever-greater challenges due to increasing human activities and the fragile environment [61].

#### 4.2. Agricultural Big Data architectures

Nine of the selected papers describe the Agricultural Big Data architecture structure, technical characteristics, and tools used.

A summary of these is presented below.

(a) Climate-smart Agriculture architecture: The architecture presented by Lopez et al. is based on a previous architecture designed by the authors for smart agriculture, which allows detecting favorable conditions for planting and crop production in the Cauca river basin. The updated architecture fuses and analyzes data from different dimensions in climate-smart agriculture (CSA) contexts, applying Big Data, Data Mining, and Machine Learning methods. They apply different algorithms for the fusion process [48]. Figure 5 represents the components considered in the 4-layer architecture.

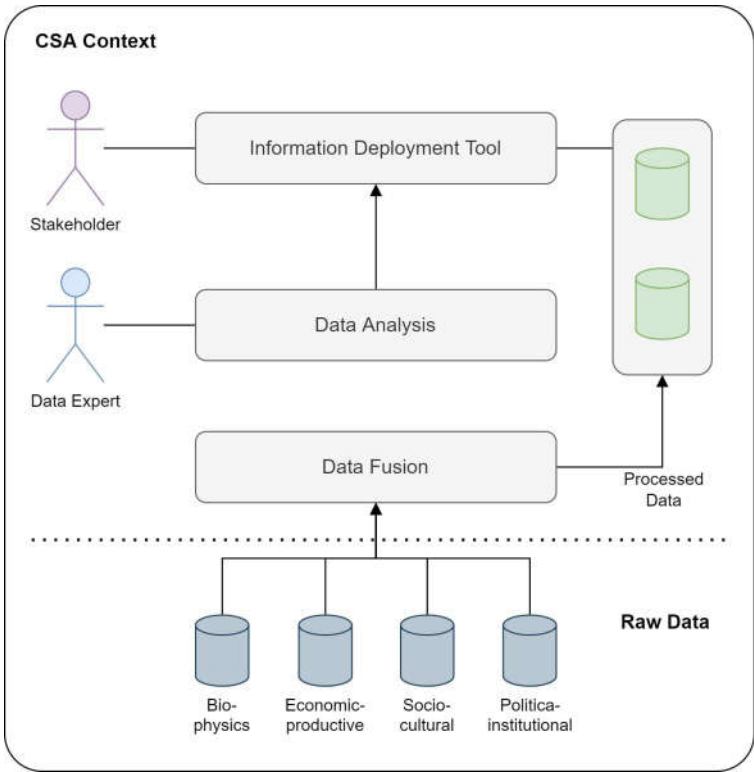


Figure 5. CSA architecture [48].

The architecture of Figure 5 is complemented by the structure of a Data Lake, as presented in Figure 6. The Data Lake enables data ingestion, storage, transformation, processing, and governance processes.

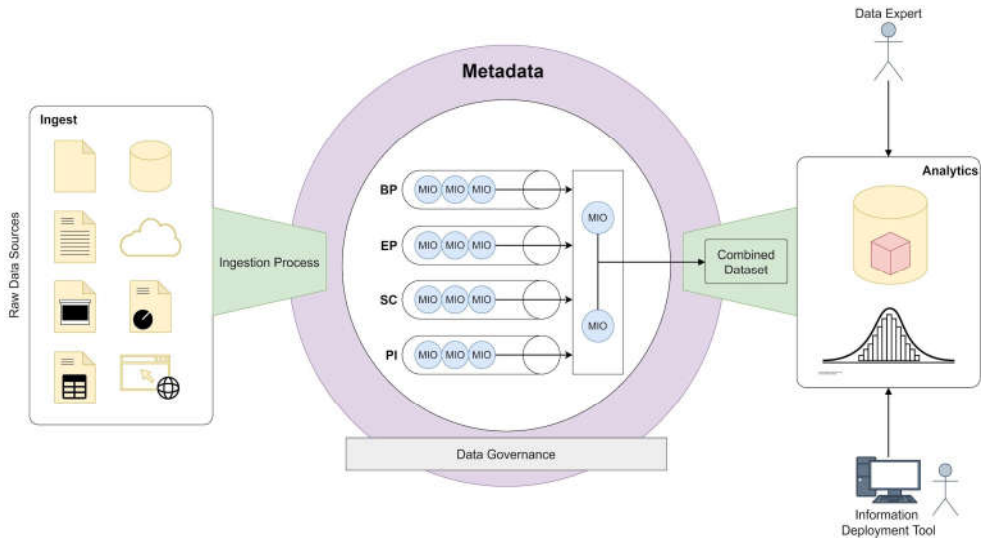
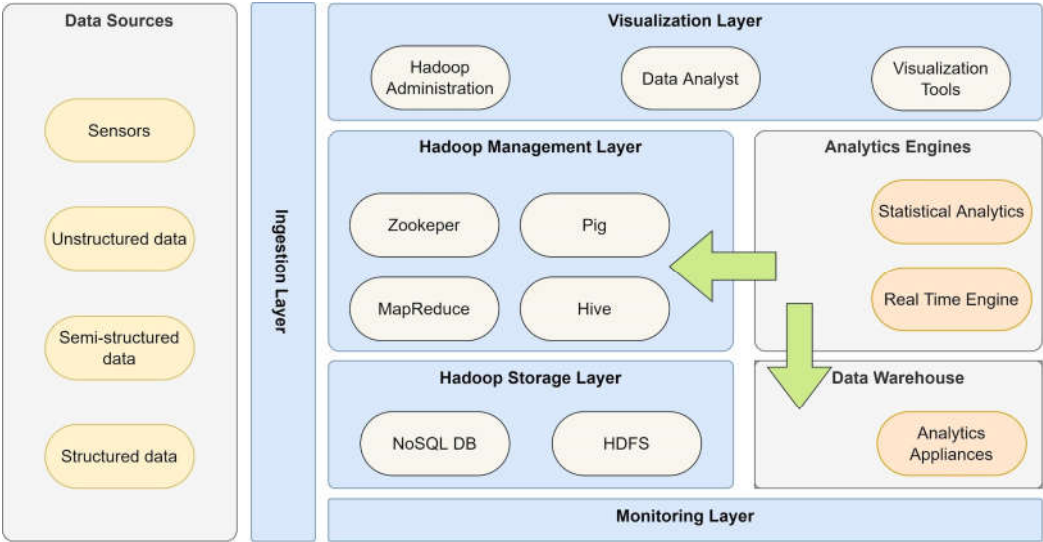


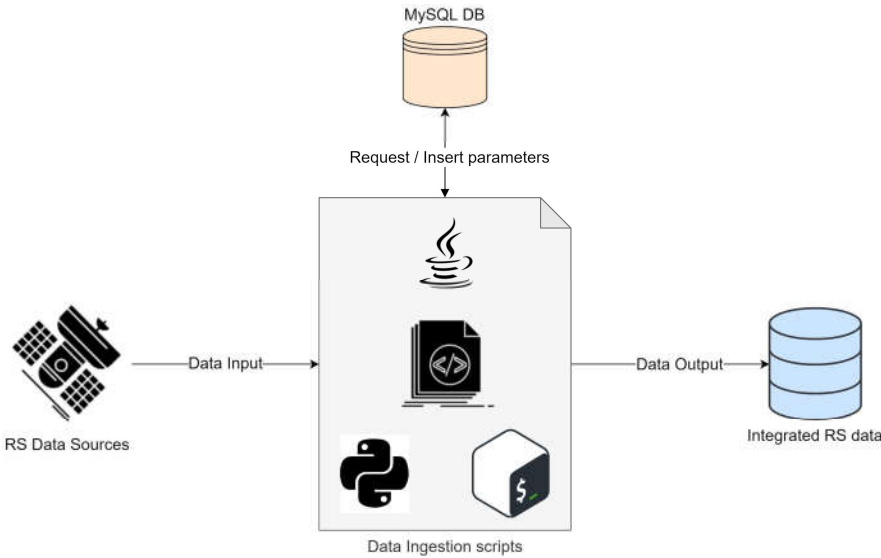
Figure 6. Data Lake is used by CSA [48].

(b) Big Data architecture for atmospheric composition monitoring: Semlali et al. propose a six-layer Big Data architecture: Data source layer, ingest layer, Hadoop storage, management, infrastructure, security, and monitoring layer. Big Data enables the processing of pollutant gas emissions data from various sources such as industry, transportation, and agriculture. In addition, the system performs continuous monitoring of atmospheric composition through remote sensing techniques, providing satellite data on a global scale [39]. Figure 7 represents the proposed architecture.



**Figure 7.** Proposed Big Data architecture for atmospheric composition monitoring [39].

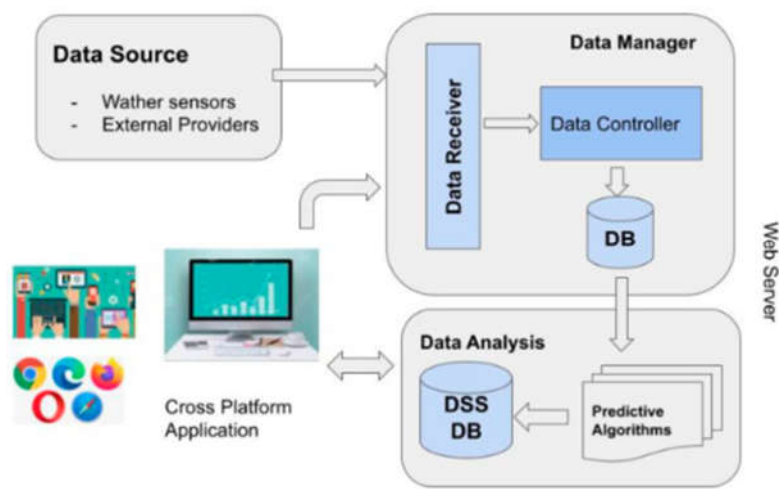
In this architecture, a MySQL database stores the processed data from the various data sources. Then, they use Java, Python, and BASH scripts to read the raw data stored in Hadoop for the ingest process. Figure 8 represents the data flow.



**Figure 8.** Ingest process for Big Data atmospheric composition monitoring system [39].

(c) DSS LANDS Architecture for Forecasting Crop Disease: Fenu and Mallocci develop a prototype agricultural DSS in collaboration with Laore Sardinia Agency [62]. The DSS is designed to help Laore technicians and Sardinian farmers in decision-making. The purposes of LANDS are: (1) to optimize resource management by reducing specific inputs, for example, chemical and natural resources, etc.; (2) to predict crop risk situations (for example, diseases, weather alerts, etc.); (3) increase the quality of decisions for field management (4) reduce environmental impact and production costs. To address the objectives, LANDS is divided into three steps: (1) collects, organizes and integrates a large amount of data from different sources; (2) analyzes and interprets the information; and (3) uses the analysis to recommend the best action to take. The authors structured the DSS LANDS into three components: (1) an integrated system for monitoring the crop components and storing their data; (2) a data analysis modules system that performs through several mathematical and forecasting models across and dynamic analysis of different types of data; (3) a cross-platform application used by farmers to upload crop data collected during the

field survey and to visualize the up-to-date information for managing the cultivation. Figure 9 shows the DSS LANDS Architecture for Forecasting Crop Disease.



**Figure 9.** DSS LANDS Architecture for Forecasting Crop Disease [62].

(d) Big data architecture for crop yield prediction: In Alex et al, they use a Big Data architecture to predict crop health risks due to excess fertilizer use, based on average rainfall, soil moisture, and soil nutrients, among others [49]. The data they use for this are nitrogen, phosphorus, Calcium, Magnesium, and Sulfur. The Big Data architecture contemplates the following stages: data enrichment, data clustering, data classification, and recommendations. Figure 10 represents the architecture design.



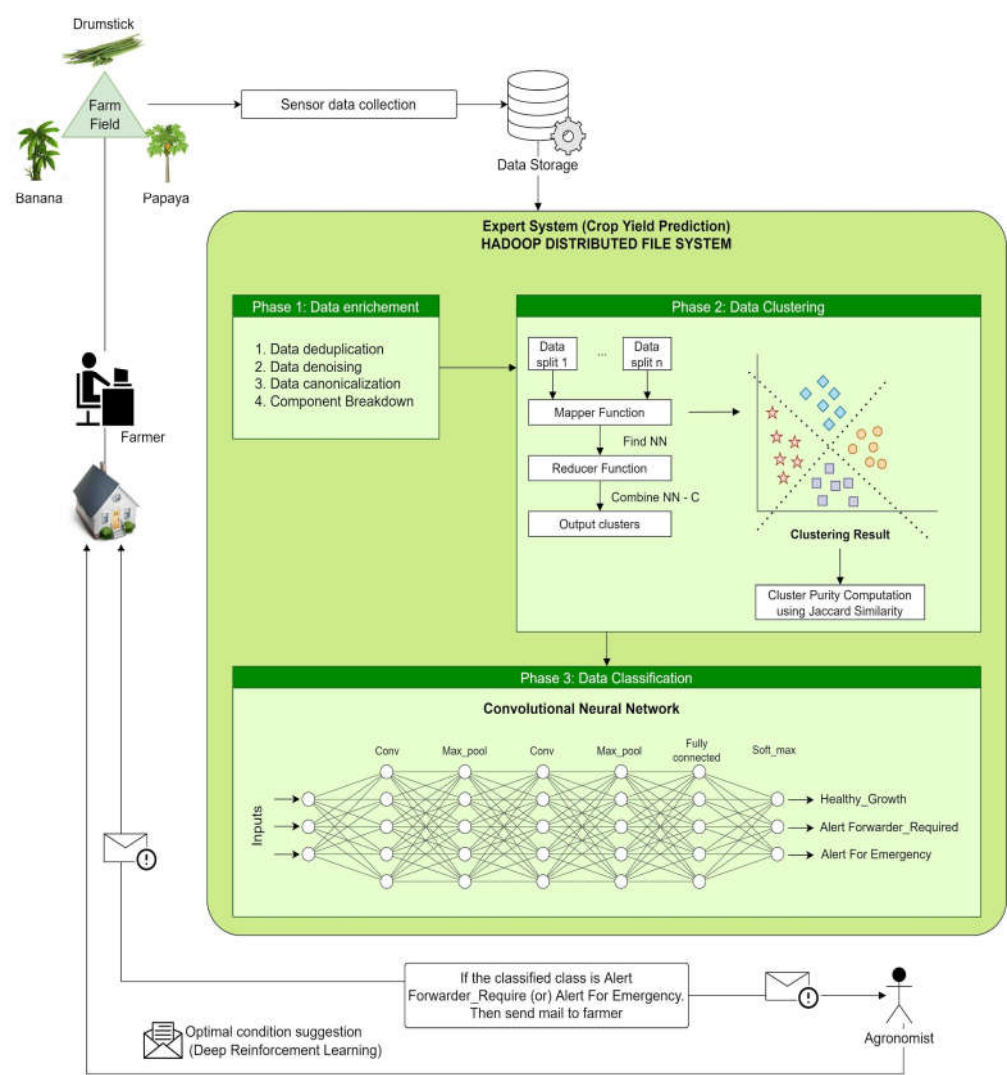


Figure 10. Big Data architecture for crop yield prediction and fertilizer management [49].

(e) Climate Big Data architecture: Gad et al, integrate technologies used in data warehousing with Big Data to process NCDC weather data. Climate observations are taken daily or hourly, although the collection of temperature and precipitation, wind speed, wind direction, sunshine hours, and dew condensation were initiated in 1901 and are published annually, monthly, daily, and hourly [44].The architecture used is presented in Figure 11.

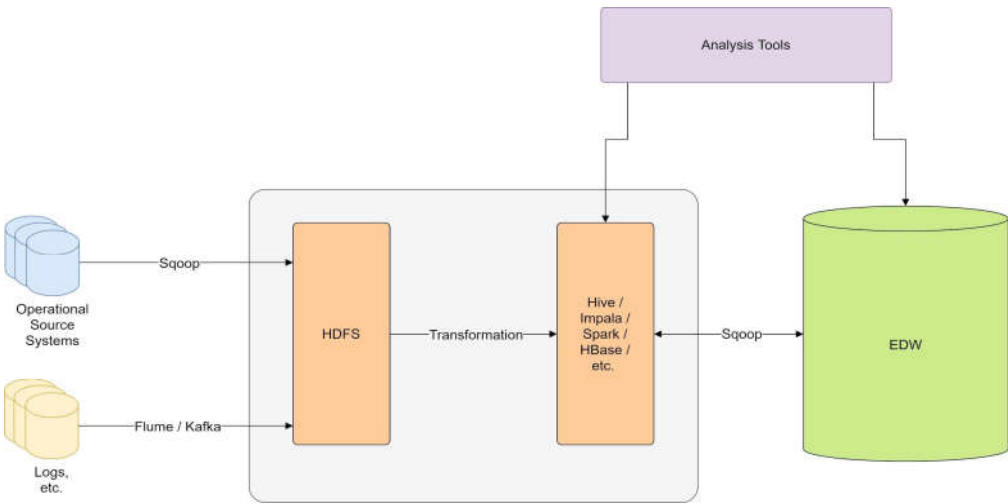


Figure 11. Climate Big Data architecture [44].

(f) Big Data architecture and IoT: Lambrinos uses a Big Data architecture to integrate climate and water data through IoT technology [52]. The system allows the farmer to adapt to weather conditions and assess new conditions due to climate change (See Figure 12).

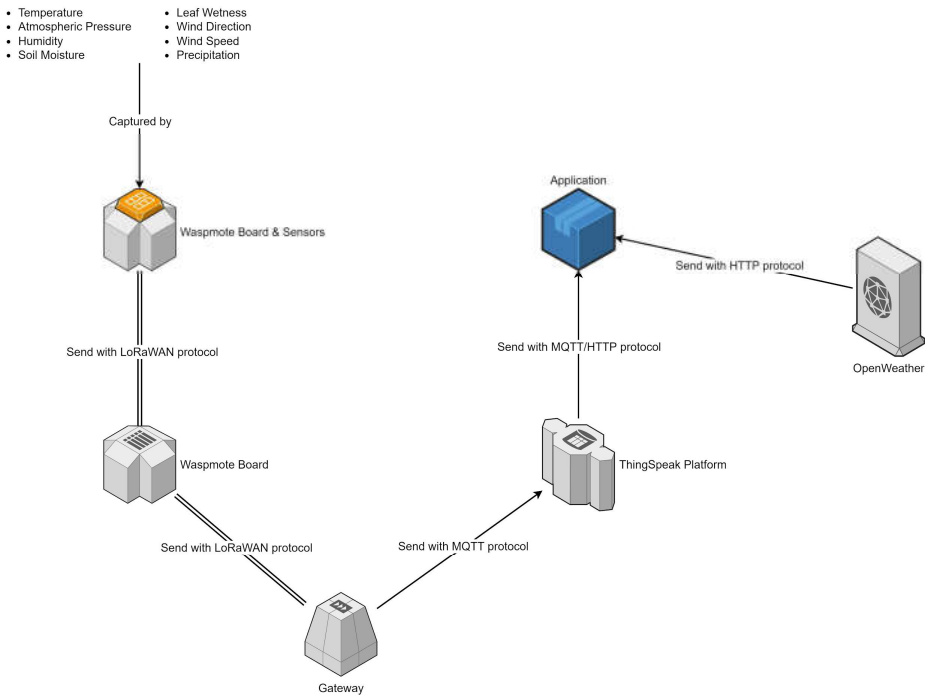
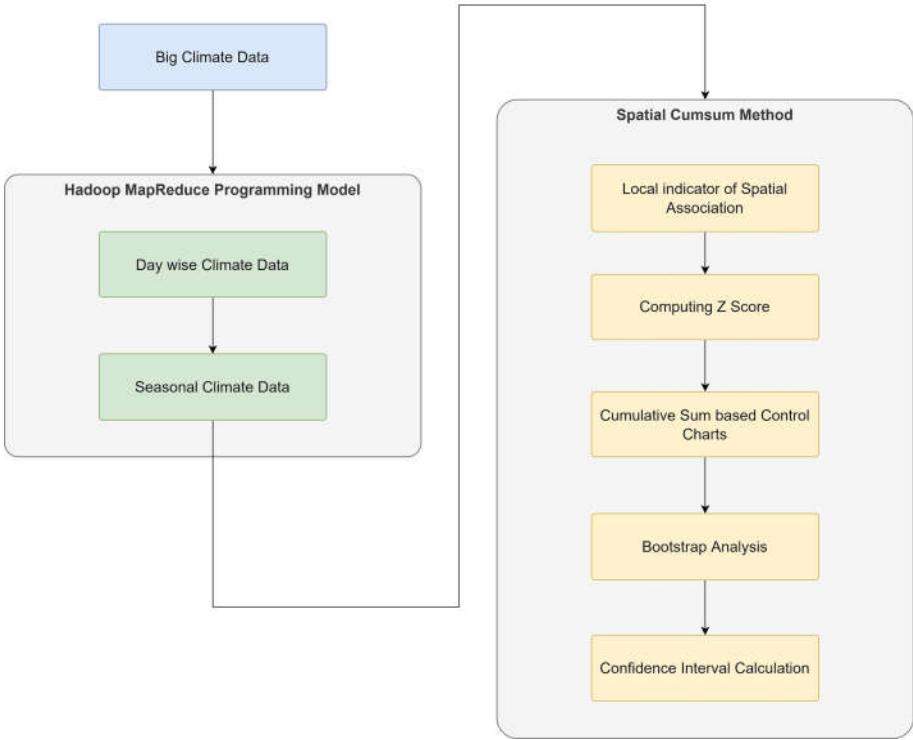


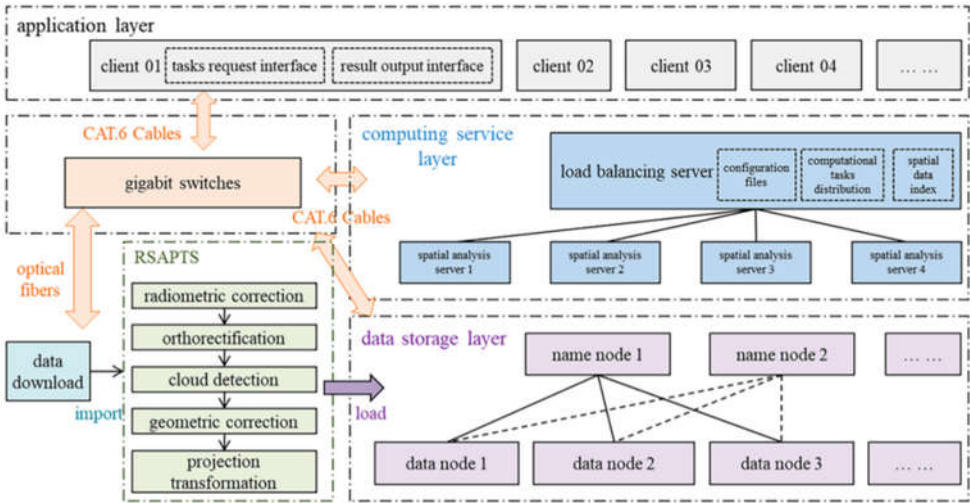
Figure 12. Big Data Architecture and the IoT [52].

(g) Big Data architecture for climate change detection: Manogaran and Lopez present a scalable data processing framework with a novel change detection algorithm [63]. They propose the spatial cumulative sum algorithm to monitor seasonal changes in climate data. The MapReduce algorithm is used to create a table in Apache HBase with the help of Apache Hive. Finally, the authors propose a climate change detection algorithm based on spatial autocorrelation to monitor changes in a seasonal climate. The Big Data architecture used is shown in Figure 13.



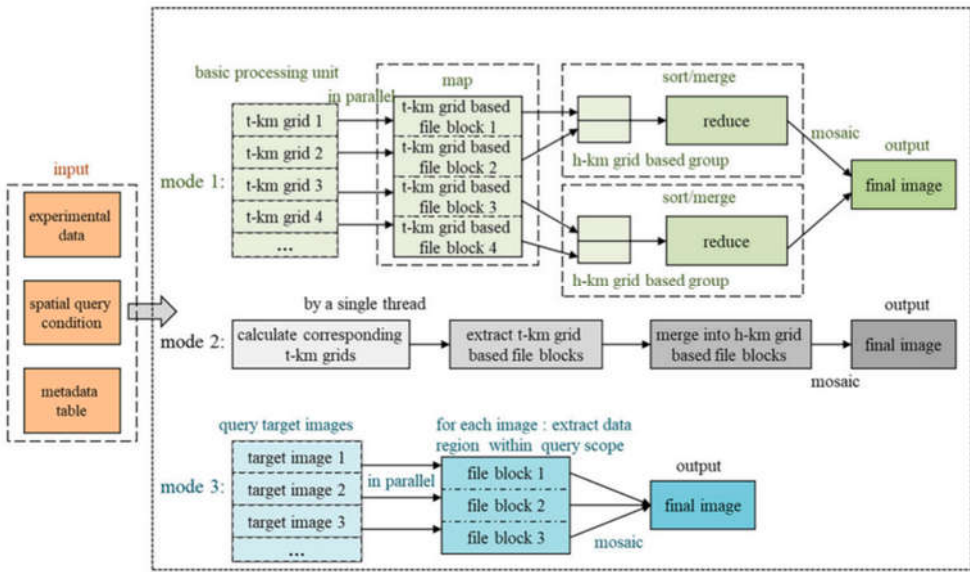
**Figure 13.** Proposed Big Data Architecture for climate change detection [63].

(h) Big Data architecture for monitoring vegetation dryness: Ye et al. developed an RDCRMG architecture for monitoring remotely sensed dryness. The authors use spectral reflectance to improve monitoring accuracy by comparing R-NDVI and O-NDVI percentage H-information data [56]. The Big Data architecture is shown in Figure 14; it includes the data storage layer, network transport layer, computing service layer and application layer. A computing cluster with five computers was established for parallel data mining and processing in the computing service layer. In addition, an RSAPTS automatic data pre-processing system is available to support near real-time updating of GF-1 Remote Sensing data. It includes automatic radiometric correction, orthorectification, cloud detection, geometric correction, and projection transformation.



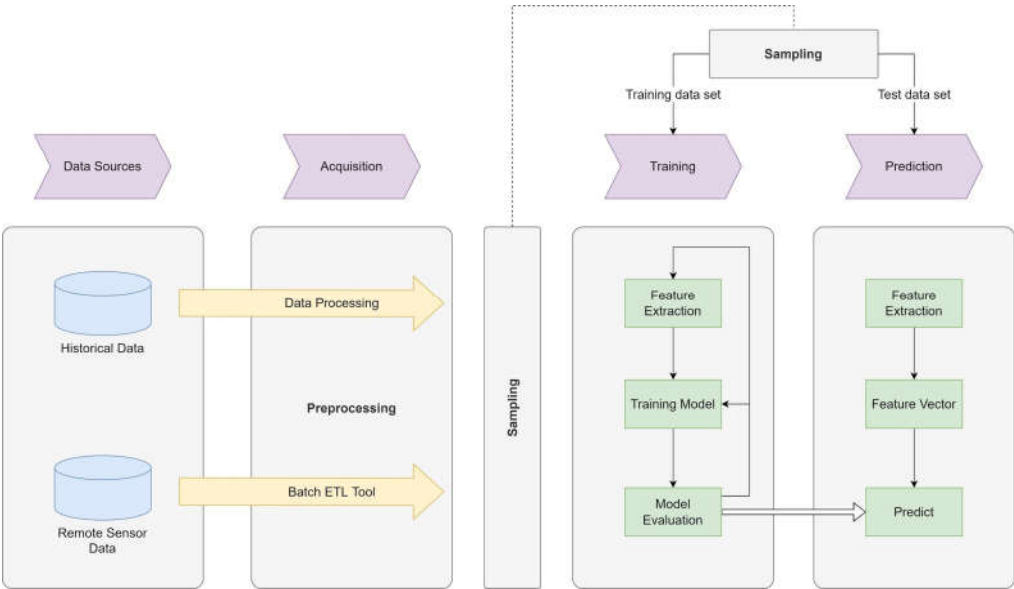
**Figure 14.** Big Data architecture to monitor dryness [56].

The authors used three nodes with different algorithms to make the data extraction process fast. Figure 15 shows the processes.



**Figure 15.** Process of Three Modes for data extraction.

(i) Big Data architecture for vegetation monitoring: Sumalatha and Akila, propose an architecture to analyze the land data and build a regression-based model that fits the observations [64]. The data is obtained daily from satellite imagery and historical data through the government portal "indiawaterportal.org." The designed architecture is 4-layered, as shown in Figure 16.



**Figure 16.** Big Data architecture for vegetation monitoring [64].

(j) Architecture for monitoring the ocean: In Coro et al., they use a three-layer architecture to analyze data from Aqua-Maps and NASA archives. In the data preparation layer, six-time environmental parameters from AquaMaps were processed to produce a structured and standard representation as NetCDF files by R-based algorithms using a Map-Reduce approach via DataMiner. Next, two additional parameters were used in the processing layer as the differences between the annual average near-surface air temperature and sea surface temperature (SAT-SST). Finally, in the data analysis layer, they represent ocean areas to reflect global average climate change trends [34].

Figure 18 shows the technologies used in the selected papers, which were used to capture the data for the analysis.





**Figure 18.** Cloud of concepts of data sources used.

For crop management, data from weather stations, agroforestry resources, sensors, remote sensing images, GPS positioning, satellite information engines such as Google Earth, and external private/public data providers were used [54], [46]. For such action, in some cases as in Colombia, local data systems such as AVA CDKN, CVC, IDEAM, Agronet MinAgriculture, Datos Abiertos Colombia MinTIC, SAC, FAO, DANE, DNP were used [48]. Each data source provides temperature, humidity, illumination, atmospheric pressure, soil electrical conductivity, soil and vegetation moisture content, and soil salinity [51]. On the other hand, genomic data related to flood tolerance was found [65].

In Moon et al., the data were provided by non-standard technological actors, such as the Agricultural Technology Center, Jeollanamdo Institute of Agricultural Technology, and Jeju Island Institute of Agricultural Technology, mainly linked to the Rural Development Administration [53].

For the case of water management, the most frequently used data sources are satellites, which provide historical information on temperature, humidity, and precipitation levels. On the other hand, annual reports from institutions such as the Water Department were also used over a characteristic range of years [35]. Other data are mainly provided by federal, provincial and municipal authorities (governments in general), also commercial and research organizations [36]. The Google Earth Engine data catalog is also a viable alternative and climate data from Meteorological Departments [38].

In the case of climate management, a fairly widely used option is satellite remote sensing [39] and data from meteorological services. Another real source of data is provided by government ministries, communities and farmers [66]. In one of the cases, interviews and surveys were also applied by to experts obtain information for the analysis [40].

Data from soil maps, remote sensing image data, meteorological centers, and statistical data from various stakeholders are used [42]. Similarly, some data came from geo-

physical information from various sources and satellite imagery [64]. Some of the attributes used in the data sources were precipitation, temperature (minimum, maximum, and average), cloudiness, vapor pressure, frequency of wet days, and ground frost [64].

#### 4.2.3. Data storage and processing

Figure 19 shows the main tools used for data storage and processing. It is observed that the most used storage tools are Hadoop and MySQL. Hadoop, specifically HDFS is used when it is required to store unstructured data such as satellite images, videos, and spatial data, among others. MySQL is a relational database engine that allows storing structured data such as project data, sensor data, data from other databases, etc. It is also noted that the most used processing tools are Map-Reduce, Flume, and Python. Map-Reduce allows mapping and reduces the memory space to generate a higher processing speed. Flume is responsible for collecting, aggregating, and moving volumes of data from various sources. Python is a programming language that allows creating data processing according to specific needs.



**Figure 19.** Data storage and processing tools.

Coro et al. used R-based algorithms in conjunction with Map-Reduce to transform textual data into NetCDF files [34]. The authors had to extract metadata from AquaMaps reference files, e.g., size, description, data provider, etc.; and then make the metadata conform to NetCDF (Network Common Data Form) conventions. On the other hand, Lopez et al. used Hadoop tools, such as HDFS, Sqoop, Flume, and Kafka, to store and process the data in a Data Lake, which allows storing raw, semi-processed, and processed data in different storage areas [48]. A similar case is that of Gad et al., in which they use Hadoop technologies to process and store data from a data warehouse. SQOOP they used it to extract data from RDBMS (in this case, Mysql and Oracle) in Hadoop; Hive to create, join and partition tables and databases in Hadoop, in addition, to performing aggregation operations; SQOOP they used it to export the results of data warehouse aggregation [44]. On the other hand, besides Hadoop technology, Bendre et al. use cloud technology such as Amazon Elastic Computing Cloud (EC2) and Google File System to achieve parallel and distributed processing. The authors use Map-Reduce to process data that is not stored in memory to be stored in Hadoop [67]. Manogaran et al. [63], Lopez et al. [48], and Reddy and Sureshbabu [38], also use Map-Reduce to process the data and then be stored in Hadoop. They use Apache Hbase and Apache Hive [63], [38].

Semlali et al., process data obtained from ground and satellite sensors to capture data from the sun and data from repositories such as MDEO, NASA, NOAA, ESA, and some

MGS. At this stage, the Big Data technologies used are: Raspberry Pi, BUFRextract (Python), pybufr\_ecmwf (python), wget, curl; and as data repositories, the MySQL database engine for structured data, Hadoop HDFS for unstructured data; in addition, processing algorithms developed in Java, Python, and Bash processes are used [39].

#### 4.2.4. Data Analysis

Figure 20 shows the main tools used for data analysis. It can be seen that the most used analysis tools are R language, Python, Storm, and ArcGIS. R is a language widely used in statistics and data mining. Python is a general programming language, but it efficiently allows data analysis thanks to its functions and ease of use. Storm offers enormous features for situations requiring real-time analysis, machine learning, and continuous operations monitoring. Finally, ArcGIS is a tool for geospatial data analysis.



**Figure 20.** Data analysis tools.

Various data analysis tools are described in the papers. For example, Coro et al. use the DataMiner tool to analyze ocean data [34]. Delgado et al.; uses the DND, Comment-Farm, CEEOT, and NTT tools to analyze agricultural data [47]. Li et al., use ArcGIS, MATLAB, and Excel tools to analyze land cover data [42]. Senay et al., use the SSEBop (Operational Simplified Surface Energy Balance) model to calculate real total daily evapotranspiration to analyze meteorological data [35]. Lokers et al. use rdf4j, SOAP, and SparQL to perform agro-environmental analysis [50]. The authors developed an infrastructure based on available metadata and semantics to process data and integrate them so that they can be downloaded in a format known as a NetCDF formatted file [50]. Finally, Gad et al., uses OLAP (Online Analytical Processing) technology to analyze meteorological variables [44].

The R programming analysis tool is used in the works of [65], [38], [68] and [34].

On the other hand, the analysis tools that come from the Hadoop ecosystem are used in different papers. For example, Lopez et al. use HIVE, XPath, and XQuery [48]. On the other hand, Semlali et al. use Map-Reduce with Spark and Storm [39]. Finally, Bendre et al. use Mahout, Drill, and Storm [67].

Some authors used programming languages for data analysis. For example, this is the case of Fenu et al. [62] and Tseng et al. [51], who use Phyton. Javascript is used by Kandekar et al. [37].

#### 4.2.5. Data Visualization

Figure 21 shows the main tools used for data visualization. It can be seen that the most used tool is the websites developed by the institutions themselves. On the other hand, GEE, GADA, and D4Science platforms are also used.



**Figure 21.** Data visualization tools.

Few papers mention the data visualization tools they use in the Big Data system. For example, Coro et al. use the D4Science tool that provides services of a distributed e-Infrastructure that provides spaces for collaborative experimentation and sharing of experimental parameters [34]. In addition, D4Science adds open-source and free-to-use web services to publish and access the results produced, either openly or tailored to a group of people focused on a domain, e.g., on climate change.

Delgado et al. use GADAS, the United States Department of Agriculture's Global Agriculture and Disaster Assessment System [47]. GADAS represents a first-generation system that could be the basis for a broader public-private partnership to connect agricultural data generators to an interconnected network for sustainability. It brings together a variety of spatial data through a GIS to illustrate the impacts of climate, water, crop conditions, land use, etc., to provide a comprehensive assessment of agriculture. In this regard, the U.S. federal government recognizes the need to govern geospatial data from a global point of view.

Lopez et al. use the Shiny tool, which allows users to interact with the analytical system through an Information Display Tool, accessed from any computer or mobile device. The GUI implementation is independent of the development environment. Furthermore, shiny allows direct compatibility with R language models and easy visualization of such information [48].

Semlali et al. use Kibana, which is a customized dashboard that displays, respectively, the Vertical Mixing Ratio (VMR) with the unit of Part per Million or Billion (PPMv-PPBv) of Ozone (O<sub>3</sub>), and Carbon Monoxide (CO), to perform a pollution analysis [39].

Kandekar et al. use the Google Earth Engine (GEE), an open cloud platform developed by Google for geoscience data processing and analysis. It supports online programming and interactive visualization and can process remote sensing data online without the need to download imagery. Its mighty computing power supports PB-level remote sensing data processing, suitable for large-scale and long-term research series. In addition, the rich API documents of GEE absorb many superior remote sensing image processing algorithms, which are very friendly for people who focus on data processing rather than programming [37].

In Kamilaris et al., visualizations of datasets and geospatial analysis are performed using ArcGIS and its API for JavaScript. In ArcGIS, specific geospatial analysis scenarios are run through the platform in the form of maps. Datasets imported into ArcGIS are abstracted as layers, which become part of a large-scale geodatabase for Spatio-temporal analysis. The layers can also be combined to gain insights into the locations of livestock farms and nitrate-vulnerable areas [60].



Finally, Alencar et al. [36], Tseng et al. [51], and Lambrinos [52] use a web platform for data visualization.

#### 4.2.6. Artificial Intelligence Algorithms

It is striking that few authors use AI to perform data analysis. Fenu and Mallocci [62], and Lu [54] used Neural Networks and SVM. On the other hand, CNN was used by Alex and Kanavalli [49]. Linear regression and multiple linear regression were used by Bendre et al. [67], Sumalatha and Akila [64], and Reddy and Sureshbabu [38]. Modified Random forest and SoyNet algorithms were used by Lai et al. [65]. K-means was also used by Sumalatha and Akila [64].

### 5. Discussion

#### 5.1. Main Challenges and Trends

According to del Pozo et al., there is a wide range of technological approaches to improve the adaptation of crops and agricultural systems to climate change. However, technological demands and extension and training approaches need to be further evaluated to understand better the role of technology in adaptation [17]. In this sense, Agricultural Big Data plays a vital role in developing climate change adaptation. Of the 33 papers selected, 26 used Big Data to adapt agriculture to the farmer's needs, which can be through water, climate, soil, and crop management. Of the papers analyzed, 15 present adaptations in crop management, either to monitor plant growth, disease prediction, quality monitoring, etc.

According to Sebestyen et al., data and systems science tools can play a crucial role in recognizing climate challenges and mitigation opportunities by integrating heterogeneous data and models and exploring the relationship between environmental and social factors. This integrated thinking lays the foundation for promising future trends in climate informatics [18]. For the case of Agricultural Big Data, few papers are applied in mitigation situations. Two papers include architectures for minimizing climate risk situations due to GHGs and two for soil. We believe that the use of Big Data in this crucial area is still incipient.

Mitigating climate change impacts and successful adaptation require effective strategic climate change planning by countries worldwide, whose decision-making requires complex models and information sources [18]. The Big Data toolkit enables the systematization, processing, and evaluation of heterogeneous data and information sources, which are infeasible with traditional analysis tools. Harmonization of the constant scientific knowledge and data sources related to climate change may be one of the most researched areas in the future [18]. Thirty-three papers using Big Data to analyze and adapt agriculture processes due to problems caused by climate change were reviewed. Despite the promising results, researchers still need more collaboration to have open access to data sources. Repositories are required to integrate information (Data Lake) and access for the different levels of researchers: farmer, data scientist, statistician, and the data processing engineer.

According to Lopez et al., intelligent tools for decision-making represents an advantage in different fields, especially in agriculture. Although the development of mechanisms to merge heterogeneous data sources involves a high degree of complexity, current technological advances make it possible to implement these tasks with more excellent reliability. Data lakes represent a promising alternative in terms of different aspects such as flexibility, reusability, scalability, and access protocols. These aspects allow expert personnel to generate data analysis, update and reuse analytical models that support the information needs of different users [48]. However, only one paper of those analyzed (33 in total), uses a Data Lake as a data repository. Although data lakes have been successful in other projects, such as banks, retail, etc., in agriculture, it is a developing technology. This type of tool is crucial to achieving flexible architecture designs in Agricultural Big Data, adapting the data for analysis according to local and global needs in climate change.



Of the 33 papers analyzed, only nine describe the Big Data architecture design used. All the designs were created to solve specific problems, such as analysis of climate data or atmospheric composition, soil data such as dryness, and crop data analysis through sensors, among others. Only one architecture was designed to monitor climate change as a general concept. On the other hand, the use of appropriate methodologies for designing Big Data architectures was not observed. According to Salma et al., a Big Data architecture should be designed using three possible cases: (1) adopt a reference architecture, (2) follow an architectural design methodology, and (3) use a reference model (which is often created from a reference architecture [69]. In this sense, a reference architecture in Agricultural Big Data could be used as a guide for the work team, with the possibility to analyze different configurations and thus improve the quality of the architectures. The use of a reference architecture would make it possible to move forward with appropriate solutions to the problem of mitigating the effects of climate change.

Another important aspect is the opportunity for governments. According to Seles, Big Data can help governments develop smart regulations that drive proactive responses to climate change and improve organizational performance [70]. So also, farmers can leverage Agricultural Big Data and assess risks based on data to analyze the consequences of climate change on their environment [12]. In this sense, we have found few papers that include projects with government participation. Most use Big Data to solve problems from a market point of view and not from a people's point of view. There is still a lack of open-access systems for any farmer who needs them since the papers analyzed use self-developed visualization platforms. An example of an open access system is GADAS, used by both the public and private sectors.

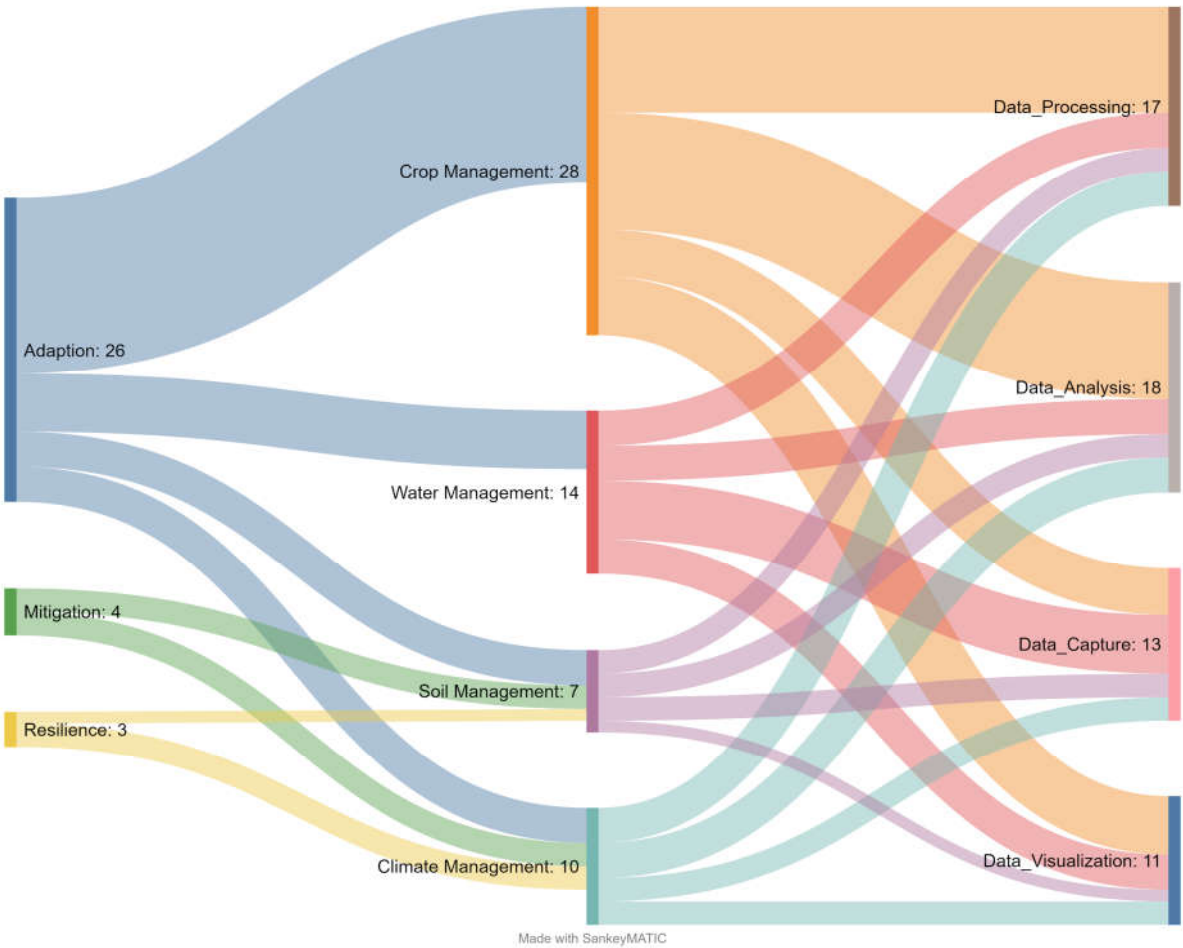
The problems of climate change will increase, so the proper development of Agricultural Big Data will be essential if we want to move forward with practical solutions. For this technology to deliver what we need, it will be urgent to have open repositories, achieve interoperability between different systems, and design reference architectures, considering all the possibilities of data sources, processing, types of analysis, and visualization types. In this sense, it is necessary to advance in modeling tools that allow flexibility. An example of this is the Feature Model and Ontologies [69], [71], [72].

After analyzing the various selected papers, we can identify the following challenges in designing architectures for Agricultural Big Data in the context of climate change: (1) use of various components and various technologies, which shows that there are still no standards; (2) architecture designs are created for particular cases and do not allow scaling for other cases; (3) some architectures only take care of data volume and not variability and processing speed; (4) no design is general and therefore no reference architectures are available; (5) the proposals do not point out the methodology for the design of architectures, and finally; (6) the use of reference models is not identified. With all these challenges, it is difficult to identify the problems that can be solved versus the right components for designing a Big Data architecture for agriculture. Furthermore, all the proposals do not use a general basis for their construction, starting from a reference architecture or model.

## 5.2. General diagram

Figure 22 summarizes the main layers of the Agricultural Big Data architectures in the context of climate change. The figure shows that the Agricultural Big Data architectures were used for crop management, climate management, water management, and soil management for the adaptation case. In addition, the architectures use components for data capture, processing, analysis, and visualization stages.

On the other hand, it is observed that Agricultural Big Data was only used for soil management and climate management in the case of mitigation. Despite the incipient development of solutions for this action, papers described the components used in the architectures. Unfortunately, no papers were found describing solutions for energy efficiency or advancing in improving crop processes and improving sustainability.



**Figure 22.** Main layers of architectures in Agricultural Big Data in the context of climate change.

**6. Limitations of the study**

Although we aimed to provide a comprehensive literature review to contribute to the improvement of future proposals for Agricultural Big Data solutions in the context of climate change, some limitations should not be overlooked. The use of search engines to retrieve scientific articles, filtering the available literature through specific keywords, has risks; and must be treated with care to draw valid conclusions. Furthermore, our approach resulted in reviewed English-language academic articles from 2015 without covering other languages. This approach led us to retrieve articles from various journals and conferences, most of academic origin.

However, filtering studies according to time of publication and English language could lead to chronological and geographical discriminations, respectively [73].

**7. Conclusions, Implications and Future Research**

A Systematic Literature Review was carried out, applying the PRISMA protocol. Thirty-three relevant articles explaining the use of Agricultural Big Data in the context of climate change were selected. The use of Big Data architectures for climate management, water management, crop management, and soil management. The advancement of technologies used in the different layers of the architectures, which are data capture, processing, analysis, and visualization, was discussed.

In climate management, only five papers included a description of the Big Data architecture used. The problems attempted to be solved by analyzing pollutant gas emissions, analyzing favorable conditions for planting, and observing climate. The leading technologies used are MapReduce, Phyton, Machine Learning, Remote Sensing, Data Lake, MySQL, and Hadoop technologies such as Hbase and Hive.

In crop management, only four papers described the components of the Big Data architectures used. The problems attempted to be solved by risk analysis for the crop, plant disease analysis, crop analysis when there is drought, and soil monitoring. The main technologies used are Sensors, Remote Sensing, Hadoop, Machine Learning, MapReduce, and R.

Many publications use Agricultural Big Data to systematically process heterogeneous information from sensors, satellite images, and climate change repositories, which is unfeasible with traditional analysis tools. Despite the advances in Agricultural Big Data related to climate change, there are still many challenges. Some of them are (1) collaboration among researchers with open access to data sources; (2) repositories with access for different researchers (farmer, data scientist, statistician, and data processing engineer); (3) reference architectures that allow collaborative work, (4) the analysis of solution alternatives, and (5) the documentation of experiences.

In future work, we hope to conduct an extensive literature review to identify all the components necessary to design a reference architecture for Agricultural Big Data, especially in climate change.

**Author Contributions:** A.C. contributed to the organization and direction of the SRL and paper writing. A.B. contributed to the planning of the SRL, paper writing, and formatting. M.N. and P.G. contributed to the methodological support and expert judgement and figures and tables. All authors have read and agreed to the published version of the manuscript.

**Funding:** Universidad de La Frontera, Project DI21-0020, funded this research.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Acknowledgments:** The authors acknowledge the technical support provided by Sebastian Pardo. They also thank the Universidad Americana de Europa and its Doctoral Program in Computer Science.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. X. Zhang and X. Cai, "Climate change impacts on global agricultural land availability," *Environ. Res. Lett.*, vol. 6, no. 1, p. 14014, 2011.
2. T.-H. Nguyen, O. Sahin, and M. Howes, "Climate Change Adaptation Influences and Barriers Impacting the Asian Agricultural Industry," *Sustainability*, vol. 13, no. 13, p. 7346, 2021.
3. M. Gaál, S. Quiroga, and Z. Fernández-Haddad, "Potential impacts of climate change on agricultural land use suitability of the Hungarian counties," *Reg. Environ. Chang.*, vol. 14, no. 2, pp. 597–610, 2014.
4. R. B. Zougmore, P. Läderach, and B. M. Campbell, "Transforming food systems in Africa under climate change pressure: Role of climate-smart agriculture," *Sustainability*, vol. 13, no. 8, p. 4305, 2021.
5. G. S. Malhi, M. Kaur, and P. Kaushik, "Impact of climate change on agriculture and its mitigation strategies: A review," *Sustainability*, vol. 13, no. 3, p. 1318, 2021.
6. A. Steiner et al., "Actions to transform food systems under climate change." 2020.
7. F. X. Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, "A review on the practice of big data analysis in agriculture.," *Comput. Electron. Agric.*, vol. 143, pp. 23–37, 2017.
8. A. Cravero, S. Pardo, S. Sepúlveda, & L. Muñoz, "Challenges to Use Machine Learning in Agricultural Big Data: A Systematic Literature Review.," *Agron. MDPI*, vol. 12, no. 3, p. 748, 2022.
9. P. Mangal, A. Rajesh, and R. Misra, "Big data in climate change research: opportunities and challenges," in 2020 International Conference on Intelligent Engineering and Management (ICIEM), 2020, pp. 321–326.
10. E. Hassani, H., Huang, X., & Silva, "Big data and climate change.," *Big Data Cogn. Comput.*, vol. 3, no. 1, p. 12, 2019.
11. D. Moher et al., "An Application of Machine Learning Technique in Forecasting Crop Disease," *Proc. 2019 3rd Int. Conf. Big Data Res.*, vol. 143, no. 76–82, pp. 23–37, 2019.
12. M. N. O. Sadiku, T. J. Ashaolu, and S. M. Musa, "Big data in agriculture," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 9, no. 4, pp. 70–72, 2019.
13. G. Talari, E. Cummins, C. McNamara, and J. O'Brien, "State of the art review of Big Data and web-based Decision Support Systems (DSS) for food safety risk assessment with respect to climate change," *Trends Food Sci. Technol.*, 2021.

14. F. BenSaid and A. M. Alimi, "Online feature selection system for big data classification based on multi-objective automated negotiation," *Pattern Recognit.*, vol. 110, p. 107629, 2021.
15. Y. Guo, B. Zhang, Y. Sun, K. Jiang, and K. Wu, "Machine learning based feature selection and knowledge reasoning for CBR system under big data," *Pattern Recognit.*, vol. 112, p. 107805, 2021.
16. S. R. Tiwari and K. K. Rana, "Feature selection in big data: Trends and challenges," *Data Science and Intelligent Applications*. Springer, pp. 83–98, 2021.
17. A. del Pozo et al., "Climate change impacts and adaptation strategies of agriculture in Mediterranean-climate regions (MCRs)," *Sustainability*, vol. 11, no. 10, p. 2769, 2019.
18. V. Sebestyén, T. Czvetkó, and J. Abonyi, "The applicability of Big Data in climate change research: the importance of system of systems thinking," *Front. Environ. Sci.*, vol. 9, p. 70, 2021.
19. I. SASSI, S. OUAFTOUH, and S. ANTER, "Adaptation of Classical Machine Learning Algorithms to Big Data Context: Problems and Challenges," 2019 1st Int. Conf. Smart Syst. Data Sci. (ICSSD). IEEE, pp. 1–7, 2019.
20. R. Elshawi, S. Sakr, D. Talia, and P. Trunfio, "Big Data Systems Meet Machine Learning Challenges: Towards Big Data Science as a Service," *Big data Res. Elsevier*, pp. 1–11, 2018.
21. B. D. Haig, "Big data science: A philosophy of science perspective," *Am. Psychol. Assoc.*, 2020.
22. A. De Mauro, M. Greco, and M. Grimaldi, "A formal definition of Big Data based on its essential features," *Libr. Rev. Emerald Gr. Publ. Ltd.*, vol. 65, no. 3, pp. 122–135, 2016.
23. Y. Demchenko, C. De-Laat, and P. Membrey, "Defining architecture components of the big data ecosystem," *Int. Conf. Collab. Technol. Syst.*, pp. 104–112, 2014.
24. M. Santos et al., "A big data analytics architecture for industry 4.0," *World Conf. Inf. Syst. Technol.* Springer, pp. 175–184, 2017.
25. C. A. Salma, B. Tekinerdogan, and I. N. Athanasiadis, "Chapter 4 - Domain-Driven Design of Big Data Systems Based on a Reference Architecture," *Software Architecture for Big Data and the Cloud*. Morgan Kaufmann, pp. 49–68, 2017, doi: <https://doi.org/10.1016/B978-0-12-805467-3.00004-1>.
26. R. Sowmya and K. Suneetha, "Data mining with big data," 2017 11th Int. Conf. Intell. Syst. Control (ISCO). IEEE, pp. 246–250, 2017.
27. M. Torky and A. E. Hassanein, "Integrating blockchain and the internet of things in precision agriculture: Analysis, opportunities, and challenges," *Comput. Electron. Agric.*, vol. 178, p. 105476, 2020, doi: <https://doi.org/10.1016/j.compag.2020.105476>.
28. G. WMO and G. OMM, "International meteorological vocabulary," 1966.
29. xC. W. Team, "Contribution of working groups I, II and III to the fourth assessment report of the intergovernmental panel on climate change," *IPCC 2007 Clim. Chang. 2007 Synth. Rep.*, vol. 104, 2007.
30. I. P. I. de Cambio Climático, "Cambio climático," *Base Defic. física*. Suiza IPCC, 2013.
31. V. Karimi, E. Karami, and M. Keshavarz, "Climate change and agriculture: Impacts and adaptive responses in Iran," *J. Integr. Agric.*, vol. 17, no. 1, pp. 1–15, 2018.
32. M. N. I. Sarker, B. Yang, Y. Lv, M. E. Huq, and M. M. Kamruzzaman, "Climate change adaptation and resilience through big data," *Int J Adv Comput Sci Appl*, vol. 11, no. 3, pp. 533–539, 2020.
33. B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," *Thechnical Rep. EBSE/07*, 2007.
34. G. Coro, P. Pagano, and A. Ellenbroek, "Detecting patterns of climate change in long-term forecasts of marine environmental parameters," *Int. J. Digit. Earth*, vol. 13, no. 5, pp. 567–585, 2020.
35. G. B. Senay, M. Schauer, M. Friedrichs, N. M. Velpuri, and R. K. Singh, "Satellite-based water use dynamics using historical Landsat data (1984–2014) in the southwestern United States," *Remote Sens. Environ.*, vol. 202, pp. 98–112, 2017.
36. P. Alencar, D. Cowan, F. McGarry, and R. M. Palmer, "An Open and Big Data Platform for Cumulative Environmental Analysis and Management," in 2015 3rd International Conference on Future Internet of Things and Cloud, 2015, pp. 412–417.
37. V. Kandekar et al., "Surface water dynamics analysis based on sentinel imagery and Google Earth Engine Platform: a case study of Jayakwadi dam," *Sustain. Water Resour. Manag.*, vol. 7, no. 3, pp. 1–11, 2021.
38. P. C. Reddy and A. Sureshababu, "An adaptive model for forecasting seasonal rainfall using predictive analytics," *Int. J. Intell. Eng. Syst.*, pp. 22–32, 2019.
39. B.-E. B. Semlali, C. El Amrani, and G. Ortiz, "Hadoop paradigm for satellite environmental big data processing," *Int. J. Agric. Environ. Inf. Syst.*, vol. 11, no. 1, pp. 23–47, 2020.
40. L. Ma and F. Nie, "A smart meteorological service model based on big Data: A value creation perspective," in 2017 International Conference on Grey Systems and Intelligent Services (GSIS), 2017, pp. 299–305.
41. Y. O. Sayad, H. Mousannif, and H. Al Moatassime, "Predictive modeling of wildfires: A new dataset and machine learning approach," *Fire Saf. J.*, vol. 104, pp. 130–146, Mar. 2019, doi: [10.1016/j.firesaf.2019.01.006](https://doi.org/10.1016/j.firesaf.2019.01.006).
42. J. Li and S. Liao, "Quality Control Framework of Big Data for Early Warning of Agricultural Meteorological Disasters," in *Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science*, 2019, pp. 74–78.
43. H. Jain and R. Jain, "Big data in weather forecasting: Applications and challenges," in 2017 International conference on big data analytics and computational intelligence (ICBDAC), 2017, pp. 138–142.
44. I. Gad, B. R. Manjunatha, and others, "Hybrid data warehouse model for climate big data analysis," in 2017 International Conference on Circuit, Power and Computing Technologies (ICCPCT), 2017, pp. 1–9.
45. P. Mangal and A. Rajesh, "Addressing climate change-making the case for big data-based decision making framework," *Int. J. Public Sect. Perform. Manag.*, vol. 6, no. 2, pp. 205–216, 2020.



46. M. Gulzar, G. Abbas, and M. Waqas, "Climate smart agriculture: a survey and taxonomy," in 2020 International conference on emerging trends in smart technologies (ICETST), 2020, pp. 1–6.
47. J. A. Delgado, N. M. S. Jr, D. P. Roberts, and B. Vandenberg, "Big data analysis for sustainable agriculture on a geospatial cloud framework," *Front. Sustain. Food Syst.*, vol. 3, p. 54, 2019.
48. I. D. López, J. F. Grass, A. Figueroa, and J. C. Corrales, "A proposal for a multi-domain data fusion strategy in a climate-smart agriculture context," *Int. Trans. Oper. Res.*, 2020.
49. S. A. Alex and A. Kanavalli, "Intelligent computational techniques for crops yield prediction and fertilizer management over big data environment," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, pp. 3521–3526, 2019.
50. R. Lokers, Y. van Randen, R. Knapen, S. Gaubitzer, S. Zudin, and S. Janssen, "Improving access to big data in agriculture and forestry using semantic technologies," in *Research Conference on Metadata and Semantics Research*, 2015, pp. 369–380.
51. F.-H. Tseng, H.-H. Cho, and H.-T. Wu, "Applying big data for intelligent agriculture-based crop selection analysis," *IEEE Access*, vol. 7, pp. 116965–116974, 2019.
52. L. Lambrinos, "Internet of things in agriculture: A decision support system for precision farming," in 2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCCom/CyberSciTech), 2019, pp. 889–892.
53. Y.-J. Moon, W. W. Cho, J. Oh, J. M. Kim, S. Y. Han, and K. H. Kim, "Forecasting cultivable region-specific crops based on future climate change utilizing public big data," *Adv. Multimed. Ubiquitous Eng.*, pp. 399–404, 2017.
54. X. Lu, "Evaluation of agricultural climate and regional agricultural economic efficiency based on remote sensing analysis," *Arab. J. Geosci.*, vol. 14, no. 10, pp. 1–16, 2021.
55. F. Sarvia, S. De Petris, and E. B. Mondino, "Multi-scale remote sensing to support insurance policies in agriculture: From mid-term to instantaneous deductions," 2020.
56. C. Ye, S., Liu, D., Yao, X., Tang, H., Xiong, Q., Zhuo, W., ... & Song, "RDCRMG: A raster dataset clean & reconstitution multi-grid architecture for remote sensing monitoring of vegetation dryness," *Remote Sensing*, MDPI, vol. 10, no. 9, p. 1376, 2018.
57. J. Bestelmeyer, B. T., Spiegel, S., Winkler, R., James, D., Levi, M., & Williamson, "Assessing sustainability goals using big data: Collaborative adaptive management in the Malpai borderlands," *Rangel. Ecol. Manag.*, vol. 77, pp. 17–29, 2021.
58. X. Hu, J. S. Næss, C. M. Iordan, B. Huang, W. Zhao, and F. Cherubini, "Recent global land cover dynamics and implications for soil erosion and carbon losses from deforestation," *Anthropocene*, vol. 34, p. 100291, 2021.
59. M. I. Fernández-Getino, A. P., Alonso-Prados, J. L., & Santín-Montanyá, "Challenges and prospects in connectivity analysis in agricultural systems: Actions to implement policies on land management and carbon storage at EU level," *L. Use Policy.*, vol. 71, pp. 146–159, 2018.
60. F. X. Kamilaris, A., Anton, A., Blasi, A. B., & 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.
61. Y. Li, Z. Cao, H. Long, Y. Liu, and W. Li, "Dynamic analysis of ecological environment combined with land cover and NDVI changes and implications for sustainable urban–rural development: The case of Mu Us Sandy Land, China," *J. Clean. Prod.*, vol. 142, pp. 697–715, 2017.
62. F. M. Fenu, G., & Mallocci, "An application of machine learning technique in forecasting crop disease," *Proc. 2019 3rd Int. Conf. Big Data Res.*, pp. 76–82, 2019.
63. G. Manogaran and D. Lopez, "Spatial cumulative sum algorithm with big data analytics for climate change detection," *Comput. & Electr. Eng.*, vol. 65, pp. 207–221, 2018.
64. M. Sumalatha, M. R., & Akila, "Real Time Big Data Analytics for Agricultural Land Hotspot Prediction," 2019 Int. Conf. Comput. Intell. Knowl. Econ. IEEE., pp. 411–416, 2019.
65. C. F. Lai, M. C., Lai, Z. Y., Jhan, L. H., Lai, Y. S., & Kao, "Prioritization and Evaluation of Flooding Tolerance Genes in Soybean [*Glycine max* (L.) Merr.]," *Front. Genet.*, p. 1865, 2021.
66. J. Mousannif, H., & Zahir, "AgriFuture: A new theory of change approach to building climate-resilient agriculture," *Int. Conf. Adv. Intell. Syst. Sustain. Dev. Springer, Cham.*, pp. 88–97, 2018.
67. V. R. Bendre, M. R., Thool, R. C., & Thool, "Big data in precision agriculture: Weather forecasting for future farming," 2015 1st Int. Conf. Next Gener. Comput. Technol. IEEE., pp. 744–750, 2015.
68. T. Moon, S. Hong, H. Y. Choi, D. H. Jung, S. H. Chang, and J. E. Son, "Interpolation of greenhouse environment data using multilayer perceptron," *Comput. Electron. Agric.*, vol. 166, p. 105023, 2019.
69. C. A. Salma, B. Tekinerdogan, and I. N. Athanasiadis, "Feature Driven Survey of Big Data Systems," in *IoTBD*, 2016, pp. 348–
70. A. M. T. Seles, B. M. R. P., de Sousa Jabbour, A. B. L., Jabbour, C. J. C., de Camargo Fiorini, P., Mohd-Yusoff, Y., & Thomé, "Business opportunities and challenges as the two sides of the climate change: Corporate responses and potential implications for big data management towards a low carbon society," *J. Clean. Prod.*, vol. 189, pp. 763–774, 2018.
71. B. S. Jyothi and S. Jyothi, "A study on big data modelling techniques," *Int. J. Comput. Networking, Wirel. Mob. Commun.*, vol. 5, no. 6, pp. 19–26, 2015.
72. B. K. Padhi, S. Nayak, and B. Biswal, "Machine Learning for Big Data Processing: A Literature Review," *Int. J. Innov. Res. Technol.*, vol. 5, pp. 359–368, 2018.
73. A. Mongeon, P.; Paul-Hus, "The Journal Coverage of Web of Science and Scopus: A Comparative Analysis," *Scientometrics.*, vol. 106, pp. 213–228, 2016.