

A functional framework based on big data analytics for smart farming

Loubna Rabhi¹, Noureddine Falih², Lekbir Afraites³, Belaid Bouikhalene⁴

^{1,3}Mathematics and Applications Laboratory, Faculty of Sciences and Technology, Sultan Moulay Slimane University, Beni Mellal, Morocco

^{2,4}Laboratory of Innovation in Mathematics, Applications and Information Technologies, Polydisciplinary Faculty, Sultan Moulay Slimane University, Beni Mellal, Morocco

Article Info

Article history:

Received Feb 9, 2021

Revised Oct 16, 2021

Accepted Oct 25, 2021

Keywords:

Agriculture 4.0

Big data analytics

Digital agriculture

Smart farming

ABSTRACT

Big data in agriculture is defined as massive volumes of data with a wide variety of sources and types which can be captured using internet of things sensors (soil and crops sensors, drones, and meteorological stations), analyzed and used for decision-making. In the era of internet of things (IoT) tools, connected agriculture has appeared. Big data outputs can be exploited by the future connected agriculture in order to reduce cost and time production, improve yield, develop new products, offer optimization and smart decision-making. In this article, we propose a functional framework to model the decision-making process in digital and connected agriculture.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Loubna Rabhi

Mathematics and Applications Laboratory

Faculty of Sciences and Technology, Sultan Moulay Slimane University, Beni Mellal, Morocco

Email: rabhi.lubna@gmail.com

1. INTRODUCTION

Following the exponential development of technologies for collecting, storing, and processing big data [1], agriculture is moving from a classical to measured and smart agriculture. This latter is driven and guided by agricultural data for greater accuracy. Today's agriculture is entering the digital age which is a good opportunity for agriculture to aspire to a brand-new level of development [2].

According to the agriculture innovation (2025) report prepared in 2015 in France, digital agriculture can make a big contribution to the competitiveness of agriculture and respect for the environment. The report highlighted two priorities: on the one hand, the establishment of an agricultural data portal for open innovation (open public data, health and economic data, private data from farmers or other economic actors). On the other hand, the structuring of research on digital transformation in agriculture by creating "living laboratories" to foster open innovation within the territories. Digital agriculture is defined as the way in which digital technologies will be integrated into agriculture to gather, store and analyze agricultural big data so as to restore, in real or near real time, more accurate knowledge which will revolutionize the farming profession [3]. This paper tries to deal with the topic of big data analytics applied to connected agriculture. It begins with an introduction, the transformation from reasoned to digital agriculture is discussed in II, the important role of big data analytics in digital agriculture is treated in III, then, a functional framework of Smart farming based on big data analytics is proposed in IV before the conclusion and future trends in this area.

2. FROM REASONED TO DIGITAL AGRICULTURE

Today’s agriculture is changing paradigm, it's moved from reasoned agriculture, codified by a reference framework taking under consideration environmental protection, consumer health and welfare-be animal, towards digital agriculture measured and guided by the data (Table 1) [4]. Measurement and diagnosis are often done by fixed or mobile sensors, by the eye, by smartphones, or other digital devices. Using strengths, weaknesses, opportunities, and threats (SWOT) analysis, reasoned agriculture has strengths, weaknesses, opportunities, and threats [5].

Table 1. SWOT analyses for reasoned agriculture

Strengths	Weaknesses	Opportunities	Threats
<ul style="list-style-type: none"> - Technical skills - Strategic support (Green Plan) - Quality of the territory (attitude, water resources and soil types) - Proximity to international market - International logistics 	<ul style="list-style-type: none"> - Soft skills - Agricultural management - Mass overproduction (deviation) - Formalization of experience feedback 	<ul style="list-style-type: none"> - Digital farming - Real-time decision support tools - IOT tools - Digital market - Advice and support in soft skills 	<ul style="list-style-type: none"> - Agricultural production in Africa - The evolution of norms and standards

The transformative capacity of digital agriculture will help farmers better manage their parcels even if they are remote. The farmer becomes a real manager, he is in continuous contact with his distributors, his supply, his tractors and his employees. Digital agriculture allows parcels to be surrounded by connected tractors, smartphones, drones, weather station, and connected objects which can collect data (soil, crop, weather data) and through rotating models, the farmer can make smart decisions at the right time (Figure 1) [6].

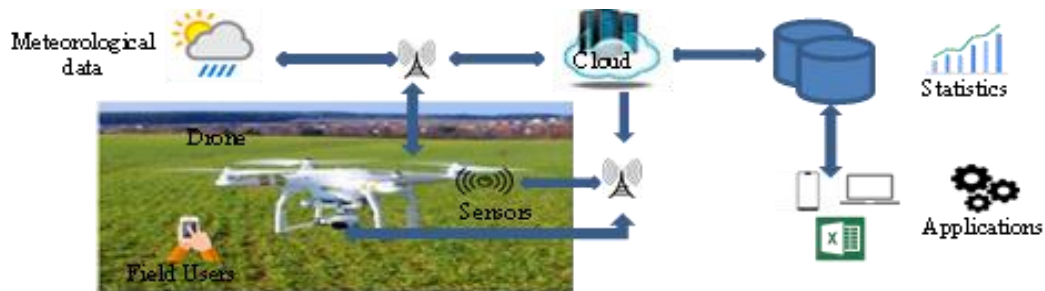


Figure 1. Digital and connected agriculture

Digital agriculture follows the development of newly connected tools. Drones, sensors, connected tractors, weather stations, and connected objects are now everywhere [7]. These connected objects have the particularity of automatically sending their sensor data for analysis by algorithms and transformed them into useful information and advice that concerns breeding, irrigation and others agricultural disciplines [8].

2.1. Breeding

These are links between animals and the parcel. The breeder can manage herds fed by sensors implanted on the animal to interpret animals' heat (cows for example) and send an SMS to the owner when they are in reproduction period. These sensors also allow the recording of the behaviors made by the animals and the automatic updating of the data by connecting directly on specific networks and servers to detect animal health and control pests (Figure 2). The breeder can also equip his herds with wireless sensor networks to know their position at all times. He can locate them in real time by defining a perimeter not to be exceeded, by receiving an SMS alert to warn him in case of overrun [9].

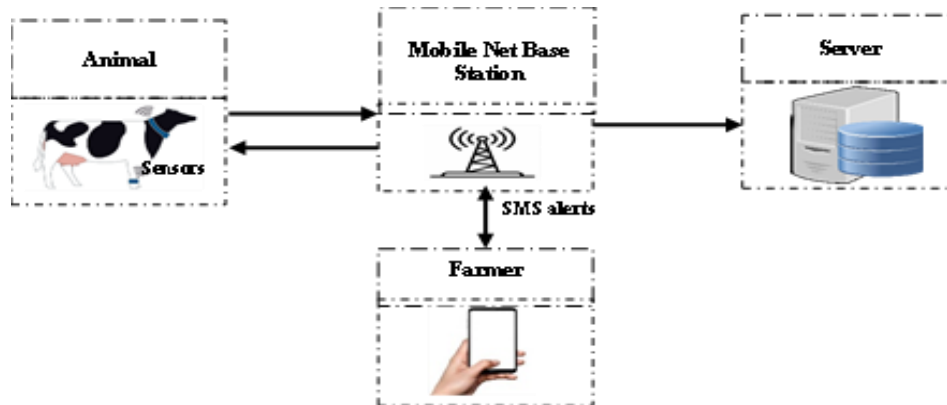


Figure 2. Internet of things (IoT) at the service of breeding

2.2. Irrigation

Classified in connected objects, wireless sensor networks are often found in irrigation to ensure optimal and remote irrigation. They make it possible to determine the driest areas in order to water them first. These internet-connected devices adjust field irrigation schedules based on local weather forecast data (rain, temperature), crop data, and soil data. Taken the case of vines for example, the sap flow sensors, placed directly on the vines, allows to evaluate the transpiration of the vine and the impact of the weather on the plant. Coupled with weather data, these massive data were transmitted via a wireless sensor network (WSN) then analyzed with decision support tools to extract useful knowledge for optimal irrigation (Figure 3) [10].

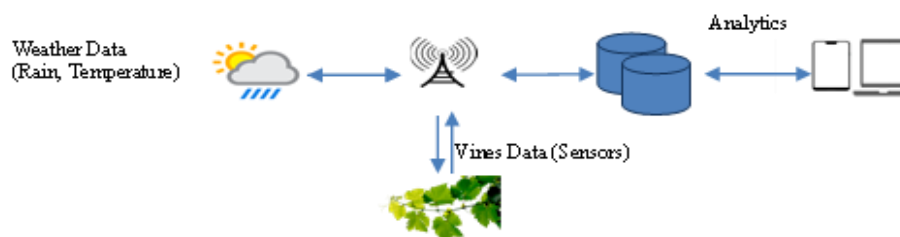


Figure 3. IoT at the service of irrigation-case of vines

2.3. Plowing

With the appearance of connected, autonomous and driverless tractor, everything is becoming under control and at all times. It allows autonomous driving on private roads of operation through automated route planning. It also has techniques for detecting obstacles and automatic or manual deflection of the vehicle. It can even stop and start automatically within the presence of moving obstacles. It is equipped with a console that controls the fertilizer (or plant protection sprayer) and automatically cuts if it passes over an already treated area. With the acquisition of this technique, grain farmers can operate under better conditions and also improve their production. However, the use of automatic GPS-guided and connected tractors remains restricted to a minority of farmers as they are too expensive [11].

2.4. Remote control of parcel

Otherwise, drones are unmanned aircraft that operate either automatically or remotely, they permit the gathering of key data on the condition of parcels thanks to the pictures taken by their sensors. These flying tools make it possible to offer agronomic indicators without having to take samples. By flying over a parcel, the drone records a multitude of geo-reference images to a precision centimetric without it being bothered by the clouds because it flies underneath. Agricultural drones have the particularity not only to film to detect weeds or damage made by pests but also to be equipped with various sensors that facilitate the analysis of a whole range of data: chlorophyll, biomass, moisture level, and water pressure [12]. After collecting the data taken during the flyby, the information still must be analyzed and interpreted to draw concrete recommendations. Agriculture is therefore in the midst of change; the digital revolution is revolutionizing the sector of agriculture. this requires human collaboration, smart machines, the imagination of specially adapted sensors, the acquisition of new knowledge and the guarantee of the safety of uses [13].

3. BIG DATA ANALYTICS: ONE OF THE MAIN PILLARS OF DIGITAL AGRICULTURE

Aware of the importance of big data analytics in driving change and promoting sustainable agriculture, OCP group and the Mohammed VI Polytechnic University initiate, on 23, 24 and 25 April 2019, the second edition of the «Agri Analytics Days» in Morocco. It is a meeting on the agricultural ecosystem titled «Big Data at the service of the farmer». This meeting focused on the importance and impact of the exploitation of agricultural data by «analytics» models to transform agricultural data into decision-making information for actors in the sector. A term “Big data analytics”, refers to advanced technologies designed to work with large volumes of heterogeneous data in order to further improve the traditional data analytics process. it includes: descriptive analytics, predictive analytics and prescriptive analytics [1]:

- Descriptive analytics: Largely based on historic agricultural data. In this technique new insights are developed using probability analysis. It gives information about What happened.
- Predictive analytics: To predict the future outcomes based on historical and current agricultural data. It provides information on what is likely happen in the future and what actions can be taken [14].
- Prescriptive analytics: Helps to derive a best possible agricultural outcome by analyzing the possible outcomes. It can be improved with experience. It provides information on So what? Now What?
- Using big data analytics approach, farmers can [15]:
- Predict and respond to changes in weather conditions: Climate change is a constant global threat, but more so for agriculture. The predictive capabilities of big data can provide accurate forecasts to allow farmers to prepare for unusual weather conditions [16]. Predictive analytics can also provide parameters for other natural factors such as precipitation, temperature, wind speed and cloud cover [17].
- Accurately predict harvest and yields: Agricultural data analytics processes the history of crop yield data to help farmers anticipate the number and status of crop yields even before seed planting. And this by choosing the best blends of agricultural yields [6].
- Predict and identify diseases: There are many risks in agriculture, such as pest infestations and crop productivity diseases that are beyond farmers' control. Prior to big data analytics, experienced farmers could spot signs of crop failure, but it was too late to do anything. With big data analytics, it is possible to track signs of plant disease and even predict if crops are threatened [18].
- Optimize crop prices: Farmers are not the only players in the agricultural sector. There are also suppliers, merchants and purchasing managers who operate in the industry’s marketing sector. Any problems in the agricultural process, such as crop delays and changes in customer supply and demand, can cause many fluctuations in market prices. Fortunately, big data analytics also has the ability to analyze the current state of the market and make price forecasts based on it [19].

4. BIG DATA ANALYTICS FRAMEWORK FOR AN AGRICULTURE 4.0

In this section, we are inspired by the ISO 19440 Enterprise Metamodel which brings a global representativeness of any company to consider a smart farm as a company characterized by four main views: the information view (IoT, sensor data, and database), the organizational view (management rules and real time monitoring), the functional view (agricultural process, data ingestion process, and analytics process) and the resource view (agricultural needs, water, and cloud server) [20], [21].

4.1. Proof of concept (PoC)

We will opt for the Figure 4 to illustrate the concept of digital agriculture by inserting big data analytics tools. Apache Sqoop technology is used to retrieve data from relational databases or simple files (master data management systems (MDM), Excel, and csv). While Apache Flume is used to recover streaming data. The storage can be done on Apache Hadoop distributed file system (HDFS). Analytics and advanced analytics can be realized using Apache Pig, Apache Hive, Spark, and Machine learning [13], [22].

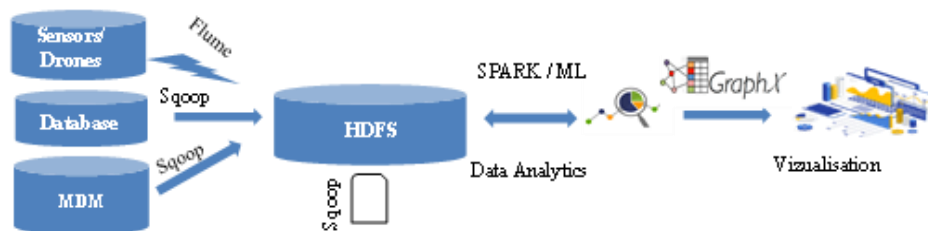


Figure 4. Proof of concept

4.2. Functional big data analytics-based framework

Based on the analysis made, the proposed functional framework is formed by four linked layers. Each layer responds to a function among the functions described in the big data analytics process. Figure 5 detail these layers and the tools used at each layer.

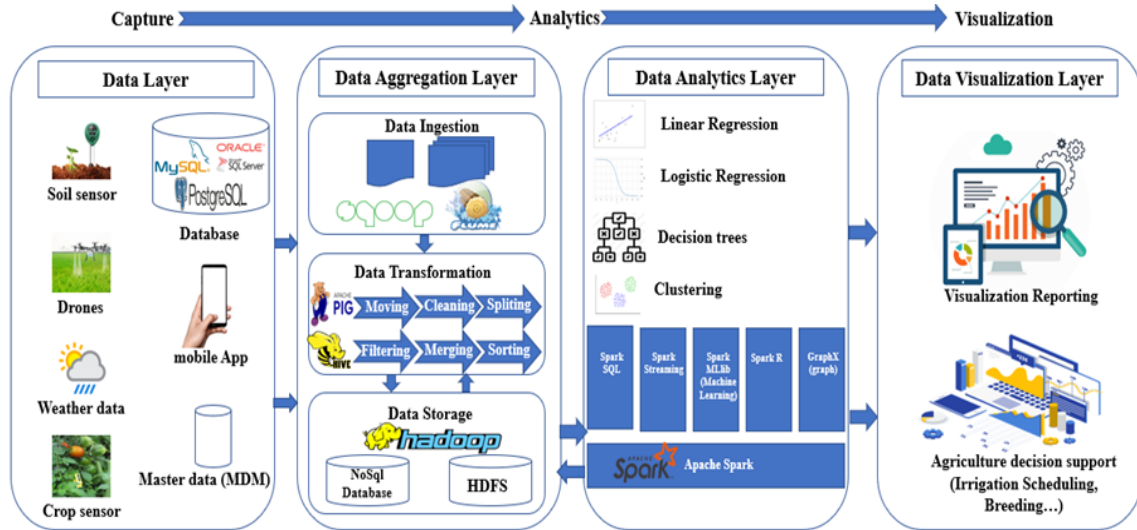


Figure 5. Functional big data analytics-based framework for connected agriculture

- 1) Data layer: Represents the plurality of agricultural data sources: Relational databases, distributed databases, master data management systems (MDM), images (drones), Excel files, and smartphones.
- 2) Data aggregation layer: Groups operations of data ingestion, data transformation and data storage:
 - i. Data ingestion: It is the process of importing, transferring, loading and processing data for later use or storage in database. It requires connecting to various data sources (relational database management system (RDBMS), web-logs, streaming data, and social media), extracting the data, and detecting changed data. Apache Sqoop and Apache flume technologies can be used for data ingestion that involves multiple source ingestion (databases, drones, soil and crop sensors, weather Station), streaming and real time data collection (Figure 6) [23]:
 - Apache Sqoop: It is an analytical tool designed for efficient data transfer, using connectors, between Hadoop and external structured datastores such as common relational databases or NoSQL databases. Sqoop can import data from an external relational database that stocks agricultural data, then it takes the table as an input, it reads row by row, and generates output files placed in HDFS.
 - Apache Flume: It is a distributed data collection service. It is useful in processing log data for getting streaming event data from different sources. So, Apache Flume can move large amount of agricultural log data from many different sources to a centralized data store.

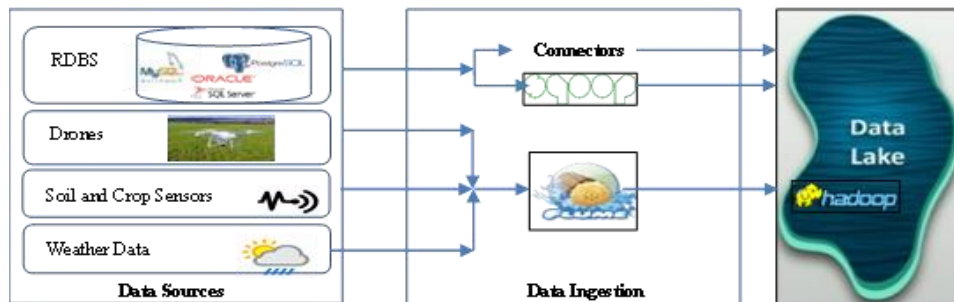


Figure 6. Data ingestion in agriculture context

- ii. Data transformation: After acquiring the data, transformations can be applied to obtain raw information as: Moving, cleaning, filtering, and sorting. These operations can be done using Apache Pig or Apache Hive tools:
 - Apache Pig: It is an open-source data flow language useful for analyzing large data sets. It enables data analysts to write data analysis programs. Pig consists of a series of operations or transformations which are applied to the input data. The language consists of relational operators like join, group, filter, and sort. Pig is for preparing agricultural data to make it suitable for querying using Apache Hive tool.
 - Apache Hive: It's a data warehouse that enables easy data summarization and ad-hoc queries via an SQL-like interface for large datasets stored in HDFS. Hive can access to files stored in HDFS or other mechanisms like HBase. It provides SQL-like declarative language, called HiveQL, to run distributed queries on large volumes of agricultural data. In Hive, tables and databases are created first then agricultural data is loaded into these tables for managing and querying structured data.
 - iii. Data storage HDFS: Data stores are at the core of the big data infrastructure and need to be fast, scalable and highly durable. Data handled can be in different forms: structured, semi-structured, unstructured data and can be stored in distributed file systems i.e. HDFS. These data are sliced and stored in a distributed and redundant parallel file system on separate cluster nodes (servers). Queries are applied (mapped) to each distributed data set [24].
- 3) Data analytics layer: Advanced analytics can be done using Spark and machine learning:
- Apache Spark: It is a fast and general engine for large-scale data processing. Spark is considered as a more comprehensive unified framework for big data analytics by allowing a rich programming APIs like SQL, machine learning, graph processing and streaming, interactive and batch processing to run on clusters of computers. It is developed in Scala language. Apache Spark can be used for process and analyze large-scale of agricultural data in order to make smart decisions [25].
 - Machine learning: It is a computer program that can learn from data. It improves its performance at some tasks through experience. His purpose is to push workload to a self-sufficient machine. We can effectively dress an algorithm for automating smart and optimal agricultural process (Figure 7) [22]. Big data analytics can be done using machine learning technique by following: Data collection to find appropriate agricultural data and merge it into single table, Feature Engineering to preprocessing and cleaning data. Dataset to forming list of clean and structured data items, training data to have a dataset from which we learn, test dataset to have data set on which we check if our model works and gives the correct prediction, algorithm to dressing logical and sequential sequence of steps to solve a given problem, and model to represent a set of numbers (vectors, matrix, parameters) that represent what we have learned.

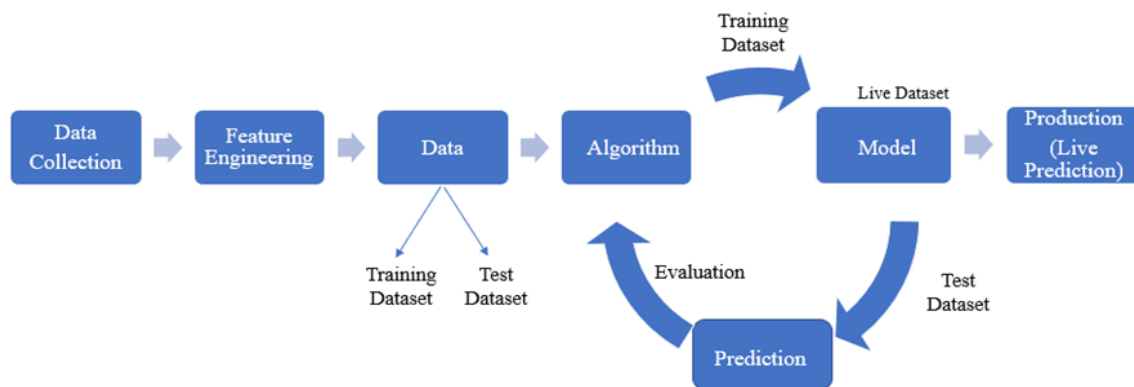


Figure 7. Machine learning blueprint

Once we receive a well-organized and cleaned table, we can model a set of algorithms based on analytical techniques. Then choose according to a survey which one gives the best result. Finally, we can run the chosen algorithm to evaluate the results before deployment (Figure 8).

- 4) Data visualization layer: This layer enables to interactively reconstitute the final result of our advanced processing on agricultural big data. We propose GraphX as a technique of data visualization.
- GraphX: Is a library that belongs to Apache Spark and programmed with Scala language. It is intended for creating charts by including a lot of functionality such as numEdges (the number of edges in the

graph) and numVertex (the number of vertex). GraphX is primarily a graph processing library. However, it doesn't include built-in visualization tools. Libraries such as Gephi or GraphLab are available to help provide visualization after GraphX has done the processing. We can also create views for visualizing data in GraphX. On the side of the farmer, we propose to use a simple mobile application that will help him to make the best decision at the right time.

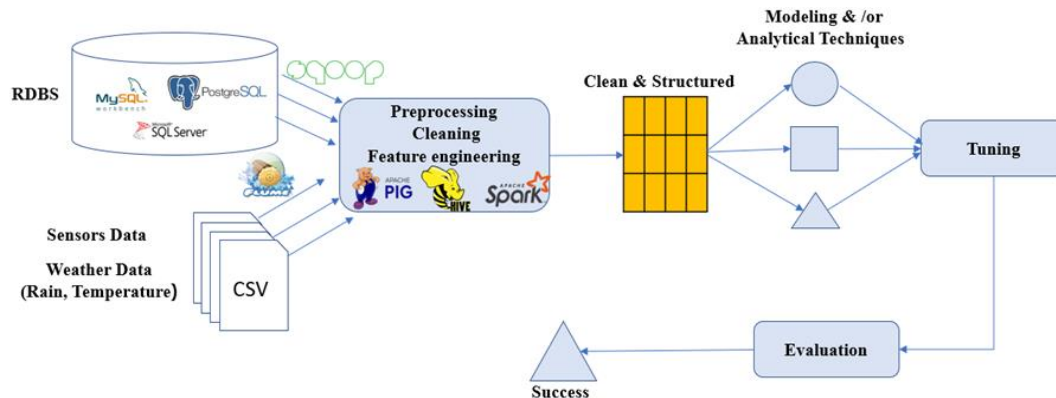


Figure 8. Machine learning workflow

5. CONCLUSION

In this paper, we dealt with a new relevant topic which is big data analytics for agriculture 4.0. In front of intelligent agriculture that embodies the enormous global use of internet of thing (IoT) analysis solutions, agriculture is becoming more precise. It offers new technical skills to the farmer, less hardship on the farm, good health and safety of the farmer (rational use of pesticides and toxic products), water resources for future generations, reduced losses, increased yields and improved product competitiveness. We have proposed a functional framework that describes our particular topic of connected agriculture by followed the approach of big data analytics which consists on agricultural data collection, storage, processing, advanced processing and visualization. Our future work will be focused on two perspectives: first, we will propose a metamodel of big data analytics for a smart farm and we will apply the standard “ISO/DIS 19440 Enterprise Metamodel” on connected farm as a smart company. Second, we will study agricultural use cases (like irrigation process, pests and diseases detection) that treat topic of digital and connected agriculture.

REFERENCES

- [1] L. Rabhi, N. Falih, A. Afraites, and B. Bouikhalene, “Big Data Approach and its applications in Various Fields: Review,” *Procedia Computer Science*, vol. 155, no. 2018, pp. 599–605, 2019, doi: 10.1016/j.procs.2019.08.084.
- [2] V. Saiz-Rubio and F. Rovira-Más, “From smart farming towards agriculture 5.0: A review on crop data management,” *Agronomy*, vol. 10, no. 2, 2020, doi: 10.3390/agronomy10020207.
- [3] R. S. Upendra, I. M. Umesh, R. B. Ravi Varma, and B. Basavaprasad, “Technology in Indian agriculture – A review,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 2, pp. 1070–1077, 2020, doi: 10.11591/ijeecs.v20.i2.pp1070-1077.
- [4] D. C. Rose and J. Chilvers, “Agriculture 4.0: Broadening Responsible Innovation in an Era of Smart Farming,” *Frontiers in Sustainable Food Systems*, vol. 2, no. December, pp. 0–7, 2018, doi: 10.3389/fsufs.2018.00087.
- [5] M. Shepherd, J. A. Turner, B. Small, and D. Wheeler, “Priorities for science to overcome hurdles thwarting the full promise of the ‘digital agriculture’ revolution,” *Journal of the Science of Food and Agriculture*, no. November, 2018, doi: 10.1002/jsfa.9346.
- [6] L. Klerkx, E. Jakku, and P. Labarthe, “A review of social science on digital agriculture, smart farming and agriculture 4.0: New contributions and a future research agenda,” *NJAS - Wageningen Journal of Life Sciences*, vol. 90–91, no. November, p. 100315, 2019, doi: 10.1016/J.NJAS.2019.100315.
- [7] W. Rahman, E. Hossain, R. Islam, Harun-Ar-Rashid, Nur-A-Alam, and M. Hasan, “Real-time and low-cost IoT based farming using raspberry Pi,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 17, no. 1, pp. 197–204, 2019, doi: 10.11591/ijeecs.v17.i1.pp197-204.
- [8] A. Triantafyllou, D. C. Tsouros, P. Sarigiannidis, and S. Bibi, “An architecture model for smart farming,” *Proceedings - 15th Annual International Conference on Distributed Computing in Sensor Systems, DCOSS 2019*, pp. 385–392, 2019, doi: 10.1109/DCOSS.2019.00081.
- [9] A. H. Ali, R. F. Chisab, and M. J. Mnati, “A smart monitoring and controlling for agricultural pumps using LoRa IOT technology,” vol. 13, no. 1, pp. 286–292, 2019, doi: 10.11591/ijeecs.v13.i1.pp286-292.

- [10] A. Vij, S. Vijendra, A. Jain, S. Bajaj, A. Bassi, and A. Sharma, "IoT and Machine Learning Approaches for Automation of Farm Irrigation System," *Procedia Computer Science*, vol. 167, pp. 1250–1257, 2020, doi: 10.1016/j.procs.2020.03.440.
- [11] J. A. Delgado, N. M. Short, D. P. Roberts, and B. Vandenberg, "Big Data Analysis for Sustainable Agriculture on a Geospatial Cloud Framework," *Frontiers in Sustainable Food Systems*, vol. 3, no. July, 2019, doi: 10.3389/fsufs.2019.00054.
- [12] A. Gacar, H. Aktas, and B. Ozdogan, "Digital agriculture practices in the context of agriculture 4.0," *Pressacademia*, vol. 4, no. 2, pp. 184–191, 2017, doi: 10.17261/pressacademia.2017.448.
- [13] S. Nagini, T. V. R. Kanth, and B. V. Kiranmayee, "Agriculture yield prediction using predictive analytic techniques," *Proceedings of the 2016 2nd International Conference on Contemporary Computing and Informatics, IC3I 2016*, pp. 783–788, 2016, doi: 10.1109/IC3I.2016.7918789.
- [14] B. M. Sagar and N. K. Cauvery, "Agriculture data analytics in crop yield estimation: A critical review," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 3, pp. 1087–1093, 2018, doi: 10.11591/ijeecs.v12.i3.pp1087-1093.
- [15] M. Bajceta, P. Sekulić, B. Krstajic, S. Đukanović, and T. Popovic, "A Private IoT Cloud Platform for Precision Agriculture and Ecological Monitoring," *3th IcETRAN*, no. June, 2016.
- [16] N. Tantalaki, S. Souravlas, and M. Roumeliotis, "Data-Driven Decision Making in Precision Agriculture: The Rise of Big Data in Agricultural Systems," *Journal of Agricultural and Food Information*, vol. 20, no. 4, pp. 344–380, 2019, doi: 10.1080/10496505.2019.1638264.
- [17] M. N. I. Sarker, M. S. Islam, H. Murmu, and E. Rozario, "Role of big data on digital farming," *International Journal of Scientific and Technology Research*, vol. 9, no. 4, pp. 1222–1225, 2020.
- [18] S. Chaterji *et al.*, "Artificial intelligence for digital agriculture at scale: Techniques, policy, and challenges," *arXiv*, pp. 1–15, 2020.
- [19] M. E. Sykuta, "Big Data in Agriculture: Property Rights, Privacy and Competition in Ag Data Services," *The International Food and Agribusiness Management Review*, vol. 19, pp. 57–74, 2016, doi: 10.22004/ag.econ.240696.
- [20] N. Falih, B. Jabir, and K. Rahmani, "Systemic approach for optimizing information technology resource as a contribution of information system governance," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 1, p. 135, 2019, doi: 10.11591/ijeecs.v14.i1.pp135-142.
- [21] M. S. Farooq, S. Riaz, A. Abid, K. Abid, and M. A. Naeem, "A Survey on the Role of IoT in Agriculture for the Implementation of Smart Farming," *IEEE Access*, vol. 7, pp. 156237–156271, 2019, doi: 10.1109/ACCESS.2019.2949703.
- [22] Y. Shekhar, E. Dagur, S. Mishra, R. J. Tom, M. Veeramaniandan, and S. Sankaranarayanan, "Intelligent IoT based automated irrigation system," *International Journal of Applied Engineering Research*, vol. 12, no. 18, pp. 7306–7320, 2017.
- [23] A. Suk Oh, "Smart urban farming service model with IoT based open platform," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 1, p. 320, Oct. 2020, doi: 10.11591/ijeecs.v20.i1.pp320-328.
- [24] T. T. Gurmessa, "A Big Data Analytics Framework in Climate Smart Agriculture," *Computer Engineering and Intelligent Systems*, vol. 10, no. 6, pp. 1–6, 2019, doi: 10.7176/ceis/10-6-01.
- [25] H. Cadavid, W. Garzón, A. Pérez, G. López, C. Mendivelso, and C. Ramírez, "Towards a smart farming platform: From IoT-based crop sensing to data analytics," *Communications in Computer and Information Science*, vol. 885, pp. 237–251, 2018, doi: 10.1007/978-3-319-98998-3_19.