**RUFORUM Working Document Series (ISSN 1607-9345), 2018, No. 17 (1): 183-192.**
**Available from *http://repository.ruforum.org***

Research Application Summary

# Dynamically processing agricultural data from controlled legume sites

Oteyo, I. N., [1,2] Kambona, K.,[1] Béra, C.,[1] Toili, M. E. M.,[3] De Meuter,W.,[1] Kimani, S.[2] & Gonzalez Boix, E.[1]

[1]Software Languages Lab, 3Plant Genetics Laboratory, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel – Belgium

[2]School of Computing and IT, Jomo Kenyatta University of Agriculture and Technology, P.O. Box 62000 00200, Nairobi – Kenya

**Corresponding author:** isaacoteyo@googlemail.com; Isaac.Nyabisa.Oteyo@vub.be

## Abstract

Smart agriculture is becoming a discipline with increased data processing requirements due to the need to understand complex agricultural ecosystems. In these ecosystems, modern digital technologies, that extend conventional tools (e.g., rain gauge, thermometer etc.), are being used to continuously monitor the physical environment resulting into large quantities of datasets in an unprecedented pace. In order to get value out of this data, efforts are being harnessed to process the data and use the outcome to make decisions that affect farm management tasks. Most of the existing data processing approaches focus on batch processing with periodic feedback. Even though this is useful, it is more desirable to have the data processed in a timely manner to support farm management decisions. In this paper, we propose our vision for a dynamic data processing platform for smart agriculture called DyPro. The platform allows processing of data as soon as it arrives to the backend and feedback sent to control points for action. To validate our platform, we present a scenario where data is being emitted by sensors into a central server for processing as soon as it arrives, and feedback sent to actuate required measures and controls.

Keywords: Big data, data processing, sensors, smart agriculture, smart farming

## Résumé

L’agriculture intelligente devient une discipline avec des exigences accrues en matière de traitement des données en raison de la nécessité de comprendre les écosystèmes agricoles complexes. Dans ces écosystèmes, les technologies numériques modernes, qui étendent les outils conventionnels (par exemple, pluviomètre, thermomètre, etc.), sont utilisées pour surveiller en continu l’environnement physique, ce qui se traduit par de grandes quantités d’ensembles de données à un rythme sans précédent. Afin de tirer parti de ces données, des efforts sont déployés pour traiter les données et utiliser les résultats pour prendre des décisions qui affectent les tâches de gestion de la ferme. La plupart des approches de traitement des données existantes se concentrent sur le traitement par lots avec rétroaction périodique. Même si cela est utile, il est plus souhaitable que les données soient traitées en temps opportun pour appuyer les décisions de gestion agricole. Dans cet article, nous proposons notre vision d’une plateforme de traitement dynamique des données pour l’agriculture intelligente appelée DyPro. La plate-forme permet le traitement des données dès leur arrivée au backend et les rétroactions envoyés aux points de contrôle pour action. Pour valider notre plateforme, nous présentons un scénario dans lequel des données sont émises par des capteurs vers un serveur central pour être traitées dès leur arrivée et des commentaires sont envoyés pour activer les mesures et contrôles requis.

Mots-clés: Large donnée, traitement des données, capteurs, agriculture intelligente, agriculture intelligente

---

## Introduction

Mobile computing is rapidly spreading into the field of agriculture (Milovanović, 2014; Ninsiima, 2015; Steinfield *et al.,* 2015; Sonka, 2016; Hoeren and Kolany-Raiser, 2018). The technologies supported by mobile computing e.g., smart sensing, have created more opportunities for crop and livestock agriculture (Caria *et al.,* 2017). Smart sensing has been used in previous approaches (Hwang *et al.,* 2010; Sonka, 2016) to monitor environmental conditions resulting in large quantities of datasets generated in an unprecedented pace being conveyed to centralised computers for processing (Sonka, 2016). This has drawn considerable attention from the research community (Bo and Wang, 2011) and efforts are being harnessed to process the data in real-time. Most of the existing data processing approaches focus on processing the data in batches after a given period of time e.g. weekly or monthly (Bennett, 2015; Marjani *et al.,* 2017; Hoeren and Kolany-Raiser, 2018). Although this is useful, in some application scenarios, immediate feedback is more desirable e.g. in crop experiment sites where conditions that have a direct impact on crop development, like temperature and humidity, are continuously monitored and regulated. The need to give immediate feedback, calls for processing the data as soon as it arrives in the central servers to drive real-time operational decisions (Wolfert *et al.,* 2017). Additionally, there is need to develop advanced solutions from data acquisition to integrated analysis and have the entire process performed smoothly in a timely manner to support measuring and monitoring various attributes in controlled experiment sites (Bi and Cochran, 2014).

In this paper, we focus on processing data from closed and controlled legume experiment sites. Research on legumes has drawn considerable attention, especially in developing countries (Protopop and Shanoyan, 2016) because the legumes are affordable and readily available as an alternative to meat and other sources of animal protein. The data is dynamically processed as soon as it arrives. We present, DyPro, a dynamic data processing platform for processing agricultural data from closed and controlled legume sites. The goal of the platform is to aid researchers in agriculture make quick decisions to control environmental conditions in closed experiment sites. The main contributions of this paper are twofold; we present an overview of data processing scenarios in agriculture, comparison of tools and platforms, and DyPro, a dynamic data processing platform.

In the subsequent sections, the paper is organised as follows. We begin by giving the motivation and related work on use cases for agricultural data processing. Then we give an overview on existing smart agriculture data processing platforms and tools and present the envisaged dynamic data processing platform. Finally, we present the conclusion and directions for future work4.

**Motivation and related work**. To motivate the need for a new dynamic data processing platform, use cases for data processing in smart agriculture are presented and explained. This is accompanied with literature study on current platforms for smart agriculture and tools for data processing. The following section describes use cases for data processing in agriculture.

**Use cases for data processing in Agriculture**. Different use cases for data processing in agriculture and focus areas are shown on Table 1. The use cases are based on the kinds of data, focus areas, data sources, and patterns of interest. Typical kinds of data in agriculture include crop, climate, soil, equipment, livestock, and market data. The focus areas include monitoring growth and development of plants, climate and soil conditions, equipment and animal movements, and market conditions. Monitoring growth and development of plants, climate and soil conditions apply both in controlled and outdoor environments. Unlike in open environments, some form of control is achievable in closed environments. This is useful mostly in determining best conditions for growing crops to attain optimal yields. For the legumes, this draws considerable attention from researchers who are mostly interested in identifying suitable varieties for different agro-ecological conditions (Monyo, 2013). The different focus areas show different patterns for different data sources (Bennett, 2015). For crop data, the patterns of interest for analysis include pest attack, disease infestation, nutrient consumption by crops, crop yields, crop development and growth rates. The patterns for climate, soil, equipment, livestock and market data include variation in weather conditions, soil characteristics, GPS locations for farm equipment, livestock movements, and variations in market prices respectively. Most of these patterns are specific to the domain of agriculture.

More patterns are bound to occur as the use of smart machines and sensors scale-up on farms and data from farms grow in quantity and scope. This makes farming processes to become increasingly data driven and data-enabled. This resonates with the crucial role (big) data is playing in all agricultural sub-domains. For example, with the use cases presented on Table 1, sensors can be used to improve on data-driven legume agriculture as a sub-domain of agro-ecological research. Agro-ecological research is concerned with dynamic systems that require detailed information on environmental conditions (Lokers *et al.,* 2016). Therefore, information and communication technologies (ICTs) become necessary to help in capturing environmental conditions like temperature, humidity, soil pH etc., and convey the data into processing platforms for analysis and feedback (Balamurali and Kathiravan, 2015). The subsequent section compares data processing platforms and tools for smart agriculture.

**Data processing platforms and tools for Smart Agriculture**. Table 2 provides an overview of different open source data processing platforms and tools that are applicable to smart agriculture. The features considered include data capturing, processing and analysis, visualisation, machine learning (ML), data mining, prediction and options for automated alerts. Data mining and machine learning aid in making future predictions for models.

**Table 1. Use cases for data processing in agriculture**

| Kind of data | Focus area | Data source | Pattern |
|---|---|---|---|
| Crop data | Pest monitoring and control | Crop scouting data and plant images; disease infestation of plants (Hoeren and Kolany-Raiser, 2018) | Pest attack patterns; disease infestation patterns |
| | Plant growth | Nutrients and growth data; plant images (Hoeren and Kolany-Raiser, 2018) | Nutrient consumption and growth rates |
| | Nutrients and their application | Plant nutrient data (Bennett, 2015; Hoeren and Kolany-Raiser, 2018) | Variations in crop yields |
| | Crop yields | Yield maps (Bennett, 2015) | Variations in yield scales |
| Climate data | Monitoring environmental conditions | Weather conditions (Bennett, 2015; Hoeren and Kolany-Raiser, 2018) | Variations in weather conditions |
| Soil data | Soil characteristics | Soil characteristics (Bennett, 2015; Hoeren and Kolany-Raiser, 2018) | Variations in soil characteristics e.g., moisture content; soil nutrient content |
| Equipment data | Equipment monitoring and control; farm production | GPS tracking (Hoeren and Kolany-Raiser, 2018) | GPS locations of mobile farm equipment |
| Livestock data | Livestock farming | GPS tracking, biometric sensing (Bennett, 2015; Sonka, 2016; Wolfert *et al.,* 2017; Hoeren and Kolany-Raiser, 2018) | Livestock movements |
| Market data | Market conditions | Commodity price (Bennett, 2015; Wolfert *et al.,* 2017) | Variations in market prices |

From Table 2, the data processing tools and some of the smart farming platforms rely on already captured data. None of the tools presented on Table 2 offers support for data capturing and automated alerts. Some of the smart farming platforms offer data capturing and support for automated alerts. In closed and controlled experiment sites, the ideal platform should support most of the features considered in Table 2. Data capturing and configurable support for automated alerts are important features to include in such a platform.

Also, even though most of the tools and platforms presented support data visualisation, there are still research challenges related to visual presentation of data (Bikakis, 2018). The ideal visualisation is expected to handle real-time interactions, on-the-fly processing, visual scalability, user assistance and personalisation. Realtime interactions entail interacting with a large number of datasets. On-the-fly processing supports visualisations over dynamic sets of raw data. And visual scalability provides data abstraction mechanisms to address the problems of overloading visual information (over-plotting). Combining the features not supported by the tools and platforms presented together with the research challenges in data visualisation, the need for a more comprehensive and dynamic agricultural data processing platform becomes apparent. The dynamic data processing platform described in the subsequent section addresses the missing gaps in literature.

**Table 2. Comparison of data processing tools and smart farming platforms**

| | Data capture | Data processing and analysis | Visualisation | ML, data mining and prediction | Alerts |
|---|---|---|---|---|---|
| **Data processing tools** | | | | | |
| Knime | × | ✓ | ✓ | ✓ | × |
| OpenRefine | × | ✓ | × | × | × |
| R | × | ✓ | ✓ | ✓ | × |
| Orange | × | × | ✓ | ✓ | × |
| Gephi | × | ✓ | ✓ | × | × |
| Datawrapper | × | × | ✓ | × | × |
| Tableau public | × | × | ✓ | × | × |
| Google fusion tables | × | ✓ | ✓ | × | × |
| Infogram | × | × | ✓ | × | × |
| RapidMiner | × | ✓ | ✓ | ✓ | × |
| Weka | × | × | ✓ | ✓ | × |
| DataMelt | × | × | ✓ | ✓ | × |
| Rattle | × | ✓ | ✓ | ✓ | × |
| **Smart farming platforms** | | | | | |
| SmartFarmNet | ✓ | ✓ | ✓ | × | × |
| Xively | × | ✓ | ✓ | × | ✓ |
| ThingSpeak | ✓ | ✓ | ✓ | × | × |
| SensorCloud | ✓ | ✓ | ✓ | × | ✓ |
| SeeControlIoT | ✓ | ✓ | ✓ | × | × |
| Oracle IoT cloud | ✓ | ✓ | ✓ | × | × |
| OpenRemote | × | × | ✓ | × | × |
| GroveStream | ✓ | ✓ | ✓ | × | ✓ |
| InfoBright | × | ✓ | × | × | × |
| Plotly | × | ✓ | ✓ | × | × |
| IBM IoT | ✓ | ✓ | × | × | ✓ |
| Ayla IoT Fabric | × | ✓ | ✓ | × | × |
| Exosite | × | ✓ | ✓ | × | × |
| AerCloud | ✓ | ✓ | × | × | × |

***DyPro: Dynamic data processing platfor***m. Figure 1 illustrates the envisaged dynamic agricultural data processing platform, DyPro. The platform is aimed at providing domain-specific tools and features to assist users capture, process, and analyse data from agricultural fields. The key components of the platform are multiple data collections options, centralised aggregation of collected data, analytics, visualisation and notification mechanism for feedback. In order to realise and operationalise these components, the following features drawn from existing literature should to be supported by the platform.

***DyPro features***. Data capture and aggregation: The data coming from multiple sensors and mobile devices, is aggregated in a centralised computer for processing. In Figure 1, data from sensors, e.g., the temperature and humidity sensors in both controlled sites and storage locations, is relayed to a central computer. The assumption is that network connectivity is guaranteed for relaying the data from the sensors to the central server. To support offline, remote locations and other kinds of data that cannot be collected using sensors, options for collecting data using hand held devices like smartphones and tablets will be provided.

***Data processing and analysis***: Although, use of sensors generates large volumes of datasets, it only happens with a large sensor base and numerous data collection points. In contrast, sensor based agro-informatics is often performed in small scale farms that may not warrant use of big data processing platforms or frameworks and tools. However, the data generated has to be processed using platforms and tools with dynamic data processing features. To support dynamic data processing, the platform employs the Apache Spark framework which allows parallelising of tasks on a cluster of computers for improved performance. The data is aggregated into micro-batches and processed dynamically as soon as it arrives.
**Visualisation:** Visualisations are provided to make interpretation and understanding of the processed data easier, and to aid in dynamic monitoring. The platform aims to support

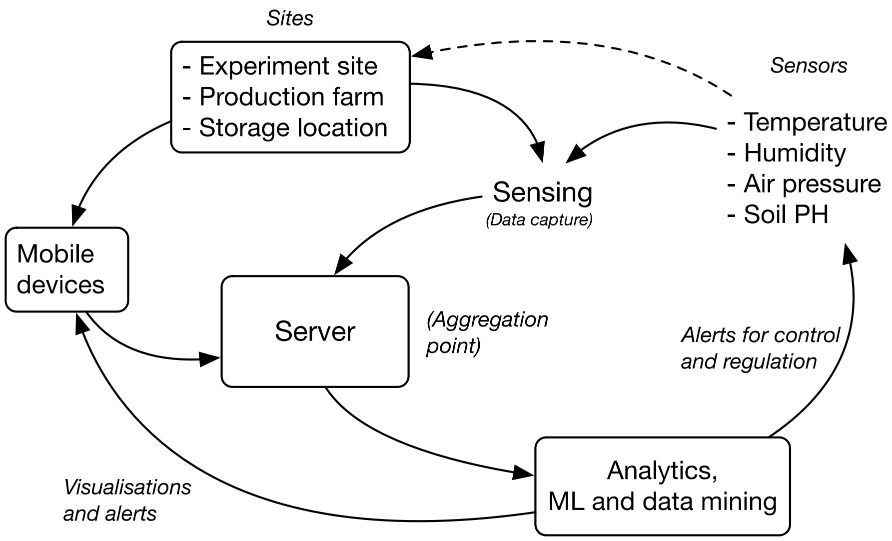


**Figure 1. DyPro - Dynamic data processing platform**

different kinds of visual graphs including line graphs, bar charts, pie charts etc. This gives users multiple options to choose from based on the contextual need. The selected visualisations are generated on-the-fly. Views for the visualised data are supported both for standard display screens and those for hand held devices. This is necessary to make the platform user friendly and flexible to mobility.

**Machine learning, data mining and prediction**: Legume research often entails correlations between different attributes which are better performed by models in machine learning, data mining and prediction. These features will help users see connections between relationships and patterns and be able to make predictions for the future e.g., whether certain environmental conditions will affect crop development and by what magnitude. The output from the machine learning models can be used as input to the data processing and automated generation of alerts.

**Alerts:** To effectively ensure immediate feedback, alerts are sent to sensor actuators or users to signal issues that need attention. It is important to have such alerts if a correction is urgent and human action is required. The alerts are sent to mobile devices e.g., high abnormal temperature or low humidity levels. To provide specialised functionality for different end-users, the immediate feedback mechanism should be customisable. The users should be able to specify their own constraints that determine when an alert is triggered and sent e.g. if the temperature goes beyond their pre-defined threshold.

## DyPro Current Implementation

In this section, we present an initial prototype implementation of DyPro for data processing and visualisation. The current iteration of the prototype exposes a wrapper for Spark that can be used to build further abstractions for components of the platform. During data capture, the data arriving at the central server comes as a stream which cannot be queried directly. The data stream first needs to be pre-processed to allow structured querying during data processing. A schema is defined and is used by DyPro as a reference when performing subsequent transformations over the stream of data. The following code snippet shows how to consume a stream for dynamic processing and visualisation.

Line 1 of the code snippet specifies the source of data for processing. The fileStreamSource specifies

```
1.  val data = DyPro.ds.fileStream("/* fileStreamSource */")
2.  val dataFrame = DyPro.dp.createDataFrame(data, ["/* schemaFields */"])
3.  val timeTempLineData = dataFrame.select($"Time", $"Temp")

                          .map(r=>Temperature(r.getString(0), r.getDouble(1))).collect
4.  val tempLineChart = DyPro.vis.lineChart(timeTempLineData, "/* chartConfig */")
5.  val timeHumidLineData = dataFrame.select($"Time", $"Temp")

                          .map(r=>Humidity(r.getString(0), r.getDouble(1))).collect
6.  val humidityLineChart = DyPro.vis.lineChart(timeHumidLineData, "/* chartConfig *
    /")
```

a location to monitor incoming saved data streams. The data is transformed into a data frame in line 2 which is used to generate visualisations in lines 4 and 6. schemaFields defines the name and type of the expected fields of the incoming data. chartConfig is used to customise the visualisations, such as the labels for the horizontal and vertical axes. Temperature and Humidity are simple abstractions with two fields which are used to retrieve the required values from the data during processing and visualisation.

```
case class Temperature(time: String, temperature: Double)
case class Humidity(time: String, humidity: Double)
```

Figure 2 and 3 show some of the visual graphs generated from the platform. Figure 2 (a) shows the dynamic temperature variation visualised after 20 hours, and (b) shows the same variation extended to 70 hours. Figure 3 (a) and (b) show the variation in humidity over the same time frames. These time-based variations make it easier for end-users to dynamically monitor changing conditions and can further be used to set triggers for automated alerts. The alerts can be triggered immediately when the configured thresholds are surpassed.

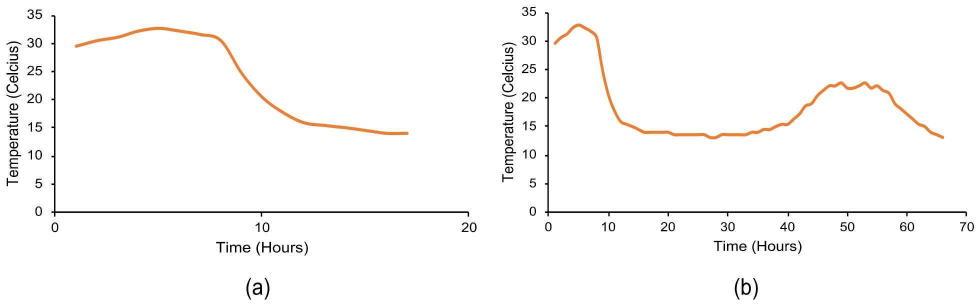


**Figure 2. Temperature variation over time**

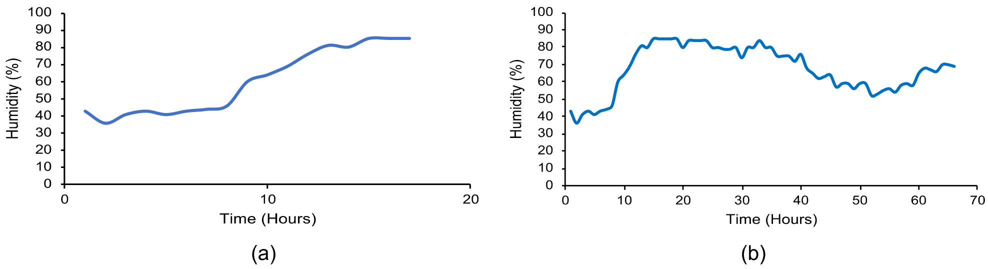


**Figure 3: Humidity variation over time**

## Conclusion

Information and communication technologies (ICTs) play an important role in the agricultural sector by providing new tools for data generation, transformation, visualisation and management. By tying together sensors, hand held mobile devices and data processing capabilities, controlled and closed legume experiment sites become more easier to manage. In this paper we present DyPro, our vision for a platform that supports dynamic processing and visualisation of data from agricultural sites. Currently, DyPro platform is aimed at helping researchers in agriculture successfully monitor conditions in controlled experiment sites. Eventually, DyPro will provide useful insights in the implementation of dynamic and real-time data processing approaches in smart agriculture. Future work includes completing the implementation of all phases of the platform, and to further provide support for collecting offline data from remote locations.

**Acknowledgment**

This work is partially funded by the VLIR-UOS LCEFoNS project and Innoviris project SeCloud. This paper is a contribution to the 2018 Sixth African Higher Education Week and RUFORUM Biennial Conference.

**References**

Balamurali, R. and Kathiravan, K. 2015. An analysis of various routing protocols for Precision Agriculture using Wireless Sensor Network. pp. 156–159. In: Proceedings - 2015 IEEE International Conference on Technological Innovations in ICT for Agriculture and Rural Development .

Bennett, J. M. 2015. Agricultural Big Data : Utilisation to discover the unknown and instigate practice change background. *Farm Policy Journal* 12 (1): 43–50.

Bi, Z., and Cochran, D. 2014. Big data analytics with applications. *Journal of Management Analytics* 1 (4): 249–265.

Bikakis, N. 2018. Big Data Visualization Tools. pp. 1–11. In: Encyclopedia of Big Data Technologies. Springer.

Bo, Y. and Wang, H. 2011. The application of cloud computing and the internet of things in agriculture and forestry. pp. 168–172. In: Proceedings - 2011 International Joint Conference on Service Sciences, IJCSS 2011.

Caria, M., Schudrowitz, J., Jukan, A. and Kemper, N. 2017. Smart farm computing systems for animal welfare monitoring. pp. 152–157. In: 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics.

Hoeren, T. and Kolany-raiser, B. 2018. Big Data on a Farm - Smart Farming. pp. 109–120. In: Hoeren, T. and Kolany-Raiser, B. (Eds.), Big Data in Context. Springer, Cham.

Hwang, J., Shin, C. and Yoe, H. 2010. Study on an agricultural environment monitoring server system using wireless sensor networks. *Sensors* 10 (12):11189–11211.

Lokers, R., Knapen, R., Janssen, S., van Randen, Y. and Jansen, J. 2016. Analysis of Big Data technologies for use in agro-environmental science. *Environmental Modelling and Software* 84:494–504.

Marjani, M., Nasaruddin, F., Gani, A., Karim, A., Hashem, I. A. T., Siddiqa, A. and Yaqoob, I. 2017. Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges. *IEEE Access* 5:5247–5261.

Milovanović, S. 2014. The role and potential of information technology in agricultural improvement. *Economics of Agriculture* 61(2): 471–485.

Monyo, E. S. 2013. Tropical legume farming in Kenya. *Bulletin of Tropical Legumes*1–12.

Ninsiima, D. 2015. Buuza Omulimisa (ask the extension officer): text messaging for low literate farming communities in rural Uganda. pp. 1–4. In: Proceedings of the Seventh International Conference on Information and Communication Technologies and Development - ICTD ’15.

Protopop, L. and Shanoyan, A. 2016. Big data and smallholder farmers: Big data applications in the agri-food supply chain in developing Countries. *International Food and Agribusiness Management Review* 19 (A):173–190.

Sonka, S. T. 2016. Big Data: Fueling the next evolution of agricultural innovation. *Journal of Innovation Management* 4 (1):114–136.

Steinfield, C., Wyche, S., Cai, T. and Chiwasa, H. 2015. The mobile divide revisited: mobile phone use by smallholder farmers in Malawi. pp. 1–9. In: Proceedings of the Seventh International Conference on Information and Communication Technologies and Development - ICTD '15.

Wolfert, S., Ge, L., Verdouw, C. and Bogaardt, M. J. 2017. Big Data in Smart Farming - A review. *Agricultural Systems* 153:69–80.