PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
Ministry of Higher Education and Scientific Research
University of Mohamed Khider – BISKRA
Faculty of Exact Sciences, Science of Nature and Life
Computer Science Department

Ordre N° : IA /M2/2020

Thesis

Submitted in fulfilment of the requirements for the Masters degree in

Computer science

Option : **Artificial Intelligence**

A data analysis system for monitoring the smart farming

By :

SENOUSSI CHEMS EDDINE

Members of the jury :

Full Name	Grade	President
MERIZIG Abdelhak	MCB	Supervisor
Full Name	Grade	Member

Session 2020

Acknowledgements

First and foremost, praises and thanks to Allah the most graceful and merciful for his showers of blessings throughout my research work to complete it successfully.

I would like to express my deep and sincere gratitude to my research supervisor **Dr. Abdelhak Merizig** for his immense knowledge, precious guidance and helping during our work. I really admire his support and understanding all along the way. I extend my appreciation to all teachers of the **Computer Science Department** who helped me in my education.

I'm extremely grateful to my parents for their love, care, encouragements and prayers. Special thanks to all my family members, sisters, brother and uncles for contributing by their love and support.

My thanks also go to my precious friends, classmates, and dear members of the **Biskra Computer Science Club**, having such great people around me had a huge impact on my life and pushed me forward to always give my best.

To everyone i love and care about, thank you.

Senoussi Chems Eddine

ملخص

تعتبر الزراعة الذكية (Smart Farming) هي مستقبل المجال الزراعي. يعتمد المزارعون في الوقت الحاضر على التكنولوجيا لإكمال مهامهم الميدانية اليومية وتحسين جودة المحاصيل. من أهم جوانب الزراعة ملاءمة الأرض ، والتي تصف مدى ملاءمة أو عدم ملاءمة الأرض لنمو النباتات. عادة ما يتطلب تحديد صلاحية الأرض خبراء للتنبؤ بها بشكل صحيح. ومع ذلك، أثبت الذكاء الاصطناعي (Artificial Intelligence) في العصر الحديث أنه أداة فعالة للتنبؤات واتخاذ القرارات. مدعومًا بإنترنت الأشياء (Internet of Things) والكميات الهائلة من البيانات المجمعة (Big Data)، يستطيع الذكاء الاصطناعي التعامل مع مثل هذه المهمة وتخفيف العبء على المزارعين والخبراء. في هذا العمل ، نستخدم إنترنت الأشياء مدعومة بالذكاء الاصطناعي لتطوير نظام قوي وموثوق في نفس الوقت لتحليل البيانات المستنبطة من الحقل والتنبؤ بمدى ملاءمة الأرض بناءا على حالة الطقس و التربة. تتم العملية بنشر أجهزة استشعار في الحقل الزراعي، والتي ترسل معلومات الطقس والتربة إلى النظام. يستخدم النظام خوارزمية التعلم العميق بالذاكرة (Long-Short Term Memory) للتنبؤ بمدى ملاءمة الأرض. و قد أظهر النموذج المقترح نتائج واعدة قد يكون لها تأثير على الزراعة في الجزائر من أجل تحسين و زيادة المحاصيل الزراعية من خلال حسن إختيار واستغلال الأراضي.

الكلمات المفتاحية: الزراعة الذكية، الزراعة، ملاءمة الأرض، تحليل البيانات، البيانات الضخمة، إنترنت الأشياء، التعلم العميق، الشبكة العصبية المتكررة الإصطناعية، الذاكرة طويلة المدى الإصطناعية.

Abstract

Smart farming (SF) is considered to be the next step for the agricultural field. Farmers nowadays rely on technology to complete their daily field tasks and improve crops quality. One of the most important aspects of agriculture is land suitability, which describes how suitable or unsuitable the land is for plants to grow. The decision concerning land suitability usually requires experts and mathematical tools to predict it correctly. However, in modern era Artificial Intelligence proved to be an efficient tool for predictions and decision making. Empowered by the Internet of Things (IoT) and the collected huge amounts of data (Big Data), AI is capable of handling such task and ease the burden on farmers and experts. In this work, we make use of Internet of Things combined with Artificial Intelligence to build a system that is both robust and reliable for applying data analytics and predict land suitability based on weather data. We deploy sensors on the farming field, that report back weather and soil information to the system. The system then uses Deep Learning algorithm (Long-Short Term Memory model) to predict land suitability. Our model showed promising results that may have an impact on agriculture in our country.

Keywords: *smart farming, agriculture, land suitability, data analytics, big data, internet of things, deep learning, recurrent neural network, long-short term memory.*

Résumé

L'agriculture intelligente (Smart Farming) est considérée comme la prochaine étape pour le domaine agricole. De nos jours, les agriculteurs comptent sur la technologie pour accomplir leurs tâches quotidiennes sur le terrain et améliorer la qualité des cultures. L'un des aspects les plus importants de l'agriculture est l'aptitude des terres, qui décrit dans quelle mesure la terre est appropriée ou inadaptée à la croissance des plantes. La décision concernant l'aptitude des terres nécessite généralement des experts et des outils mathématiques pour la prédire correctement. Cependant, à l'ère moderne, l'intelligence artificielle s'est avérée être un outil efficace pour les prédictions et la prise de décision. Forte de l'Internet des objets (IoT) et des énormes quantités de données collectées (Big Data), l'IA est capable de gérer cette tâche et d'alléger le fardeau des agriculteurs et des experts. Dans ce travail, nous utilisons l'Internet des objets combiné à l'intelligence artificielle pour construire un système à la fois robuste et fiable pour appliquer l'analyse de données et prédire l'aptitude des terres en fonction des données météorologiques. Nous déployons des capteurs sur le terrain agricole, qui rapportent les informations météorologiques et pédologiques au système. Le système utilise ensuite un algorithme d'apprentissage en profondeur (modèle de mémoire à long-court terme) pour prédire l'aptitude des terres. Notre modèle a montré des résultats prometteurs qui pourraient avoir un impact sur l'agriculture dans notre pays.

Mots clés : *agriculture intelligente, agriculture, adéquation des terres, analyse de données, big data, Internet des objets, apprentissage en profondeur, réseau de neurones récurrents, mémoire à long-court terme.*

List of Figures

List of source code

List of Tables

Contents

Chapter 1

General Introduction

1.1 General context

For many years, agriculture represents the main source of food for the human being and animals due to its importance. Government and countries look always to enhance their profit from it in order to increase the incoming resources and to cover the population needs. With the technological development, we could enhance more and more the farming field. In fact, the incorporation between traditional agriculture and technology produces a new concept untitled *Smart agriculture*. Smart agriculture is an idea that stands on using the Internet of Things (IoT) technologies in traditional agriculture. This combination will make field monitoring much easier than ever, especially in the production chain process. In the traditional vision of agriculture, farmers cover almost every operation including aggregation, crops gathering, and so on. Besides, experts suffer from limits on making studies on some phenomena or diseases that appear on the crops itself.

The new vision of agriculture stands on the idea of installing some devices on the field to work on different scenarios including those that depend on the climate. The generated data from the devices could help experts and data analysts in the decision-making process. Also, it could report on the circumstances about any phenomena whether related to the seeds or the studied area.

1.2 Problematic and Objectives

One of the major issues in agriculture, and which farmers strive the most for, is land suitability prediction. In order to have the best quality of crops, farmers need to know exactly when and where to plant their seeds, considering different factors and criteria of water, weather, soil and fertilization. Achieving the correct prediction often requires hard studies from experts and a lot of historical data. For this matter, technology comes in help. With the latest tools and devices, farmers can finally know more about their lands and crops with less efforts than traditional ways. Thus, our work is about delivering an AI system with IoT technology that provides studies about the farming field, and makes prediction about land suitability.

1.3 Outlines

Our work is organized as follows:

In the 2nd chapter, we talk about state of the art of Smart Farming. We start by giving some definitions of smart farming and Internet of things, its architecture and some devices and sensors used in IoT. We also discuss some important concepts such as Farm process, Big data, Data analysis/analytics and Data streaming.

In the 3rd chapter, we discuss some related works in smart farming. People have been working on this field since the dawn of IoT, to solve different issues that farming field suffered from in the past.

The 4th chapter is mainly about Design and Contribution. We present our proposed architecture to solve the issue of land suitability prediction, the proposed architecture is explained along with UML diagram and the used deep learning algorithm.

Implementation and results are discussed in the 5th chapter. We present the development tools and frameworks, with screenshots of the system (the web application). Then we discuss and analyse the obtained results.

Finally, we conclude our work in the 6th chapter

Chapter 2

Smart Farming state of the art

2.1 Introduction

In our modern era, some nations are still suffering from hunger, the estimated number of people who go to sleep hungry is 795 million. Despite the decrease it witnessed in the last decade by 167 million, the situation still requires more attention for Agricultural Deve et al. [2015]. As we are facing a great increase on global demand in the next 40 years, we are required to improve our crop production by 60% Shankar et al. [2016]. And this seems to be a challenge facing humanity, which needs modern and latest technologies. Smart Farming seems to be the necessary alternative to traditional farming, that delivers sustainability and improves productivity to meet people needs in the most efficient way.

2.2 Smart farming Definition

As defined in Krintz et al. [2016]: *"Smart farming is hybrid cloud technology designed to enable small-holder growers and other agricultural professionals, researchers, and students to use analytics to improve environmental sustainability and efficiencies in food production"*.

In more technical words, smart farming is defined as the combination and application of modern Information and Communication Technologies (ICT) into agriculture, providing the agricultural field with an infrastructure to leverage advanced technology, including cloud and its services, big data, robotics, automation and the Internet of Things (IoT) for tracking, monitoring, automating and analyzing operations Gorli [2017].

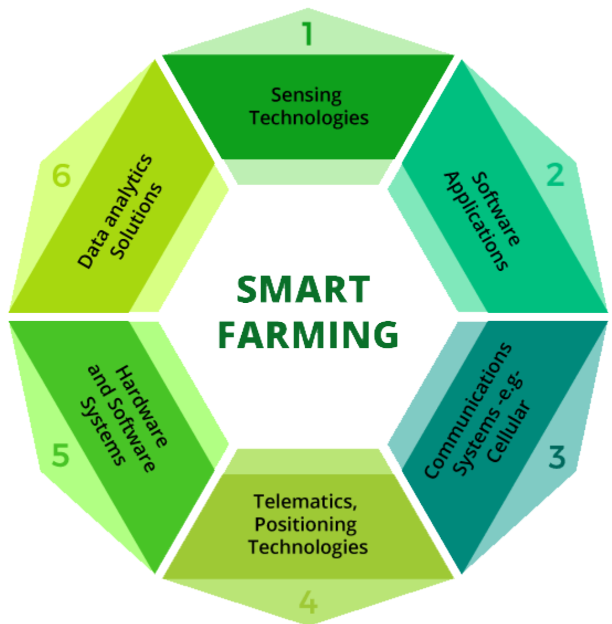

Figure 2.1: Technologies involved in Smart Farming

Applying all these technologies helps on establishing a machine-to-machine communication for deriving data. The derived data will be fed to a decision making or decision support system that will provide farmers with a clear monitoring over their fields to a very detailed level. or instance, measuring variations within a field in a precise way and adapting the strategy accordingly, this can help farmers in increasing the effectiveness of fertilizers and pesticides and use them wisely. Smart farming also helps farmers in observing animals and their needs, to provide the necessary nutrition and monitor their health to prevent diseases. SF is widely used in the USA, almost 80% of farmers adopted its techniques. However, that's not the case in Europe, only 24% of them are using it Gorli [2017].

Smart farming is mainly powered by IoT. Connecting machines and sensors to make farming processes and tasks carried with data and automatation.

2.3 IoT definition and architectures

Human life went through different waves which are agriculture, industry and information technology Jabraeil Jamali et al. [2020]. However, a new wave seems to appear in our life. Nowadays, new technology allows everything to be connected to everyone in any place at anytime, that technology is known as the *"Internet of things"*. In this section we elaborate what IoT is and its architecture.

2.3.1 Definition

According to Rayes and Salam IoT can be defined as: *"the network of things, with clear element identification, embedded with software intelligence, sensors, and ubiquitous connectivity to the Internet"* Rayes and Salam [2017].

For a simpler definition IoT can be considered as the intersection and interoperability of things, internet, and data Jabraeil Jamali et al. [2020]. The internet has evolved from a network of computers only to a network of devices of all kinds and sizes, smart phones, vehicles, different home appliances all connected and communicating & sharing information based on specific protocols, positioning, tracing, safe & even personal real time online monitoring, online upgrade, process control & administration Patel et al. [2016].

IoT is now considered to be an essential part of almost every intelligent system, due to its huge impact on different fields. This claim is powered by McKinsey's report on the global economic impact of IoT, the annual economic impact of IoT in 2025 would be in the range of $2.7 to $6.2 trillion Manyika et al. [2013].

2.3.2 Architectures

Different technologies are merged into the Internet of Thing such as sensors, actuators, cloud services and IoT protocols. Thus IoT architecture consists of different layers which serve to illustrate the communication and information exchange between the various technologies. In this section we discuss IoT architecture and the functionality of each layer.

IEEE standard association considers three layers for IoT architecture Zouai et al. [2019], as explained below:

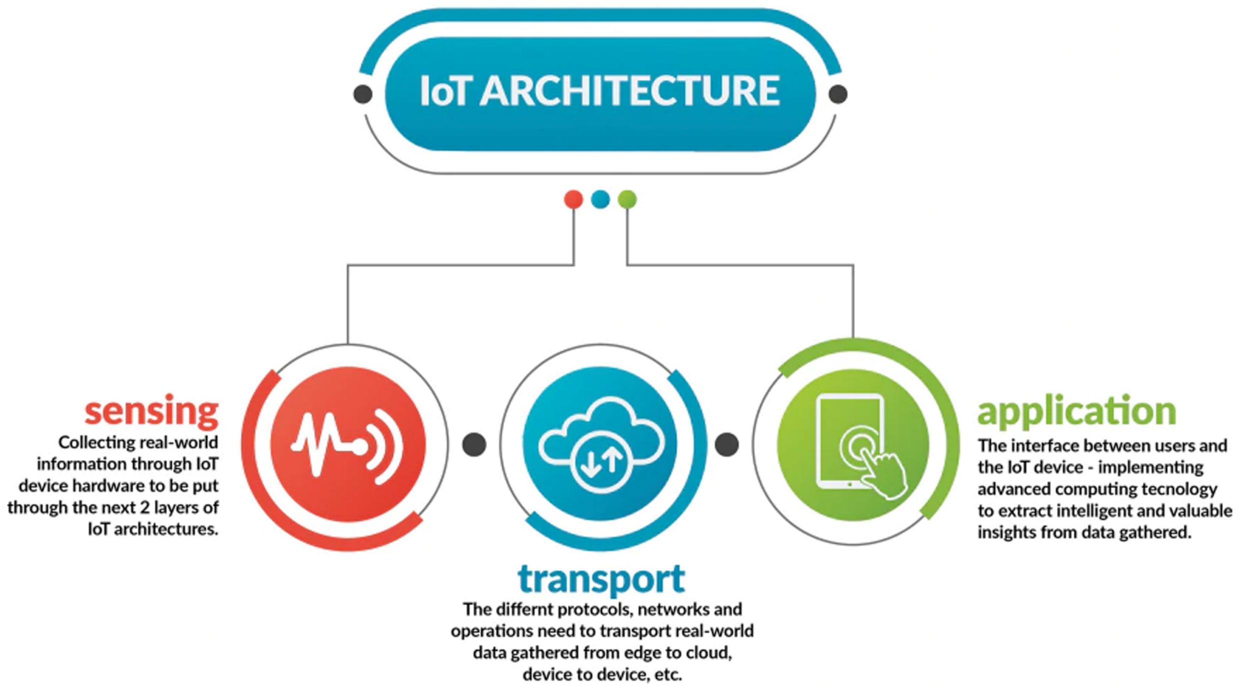

Figure 2.2: Internet of Things architecture

- **Perception layer:** sometimes referred to as Sensing layer, is the lowest level on IoT architecture Patel et al. [2016]. Perception layer is responsible for real-time sensing that allows data collection from the surrounding environment for further process, analysis and sharing. Environment sensing is achieved through smart objects integrated with small electronic devices called *sensors* Patel et al. [2016]. Sensors require connectivity to the upper layer, and this is established through Local Area Networks (LAN) such as Ethernet and Wi-Fi connections, or Personal Area Networks (PAN) such as ZigBee, Bluetooth and Ultra Wideband (UWB).

- **Network layer:** perception layer produces massive volume of data from the physical environment Patel et al. [2016]; transferring and forwarding this data requires efficient connectivity and protocols. The network layer ensures the secure transmission of information from the sensing layer to the application layer Jabraeil Jamali et al. [2020]. This layer manages packets forwarding and routing using different technologies such as wired, wireless and satellite connections which are built to support the communication requirements for latency, bandwidth, security and delivering information to the addressed device or application Patel et al. [2016].

- **Application layer:** this is considered as the top layer of the IoT architecture Patel et al. [2016]. As its name indicates, it contains the software part of the system Zouai et al. [2019]. This layer offers different services that are destined for the user; they may vary depends on the needs from health monitoring (E-healthcare) to transportation and so on.

2.4 Type of sensors

Sensors are small electronic devices that can detect and acquire information from physical objects on the environment Merizig et al. [2018] and transform it into an electrical signal. The acquired data is then transferred to a computer. Most sensors take analogue inputs and deliver digital outputs Rayes and Salam [2017], sensing inputs can come from a variety of sources such as light, temperature, motion and pressure, etc.

Some sensors that are used in the Internet of Things:

- **Temperature Sensors:** these sensors are used for measuring the amount of heat energy in a source, changes are then converted into data and reported to the system. Machinery used in manufacturing often requires environmental and device temperatures to be at specific levels. Similarly, within agriculture, soil temperature is a key factor for crop growth.

- **Humidity Sensors:** these sensors allow the measuring of the amount of water vapor in the atmosphere of air or other gases. They are planted on different devices: air conditioning (AC), heating...etc. Humidity sensors are used in both industrial and residential domains, meteorology stations and hospitals.

- **Pressure Sensors:** these kind of sensors are responsible for detecting changes in gases and liquids, then they report the change to the system. Pressure sensors are used in leak testing, water manufacturing systems as they can sense fluctuations in pressure.

- **Proximity Sensors:** they allow the detection of non-contact objects that are near the sensor. These sensors are found in parking lots in companies, malls and airports for monitoring available parking slots. Another field of use is security, proximity sensor are useful for detecting movements in an area.

- **Level Sensors:** Level sensors are used to detect the level of substances including liquids, powders and granular materials. Many industries including oil manufacturing, water treatment and beverage, and food manufacturing factories use level sensors. Waste management systems provide a common use case as level sensors can detect the level of waste in a garbage can or dumpster.

2.5 Traditional farm and smart farm

Traditional farm relied more on physical human efforts. Farmers used to handle all tasks manually and manage everything themselves. Fortunately, the advancement technology had seen in the last decade, brought so much to the field. In this section we explore the impact of technology on the agricultural domain.

2.5.1 IoT in Agriculture

Agriculture plays an important role in production and livelihoods. In order to increase productivity and enhance crops quality, IoT technology was adopted and merged into agro-industrial and agriculture Talavera et al. [2017]. IoT can add value to all areas of farming and agriculture from growing crops to forestry. Applications may surpass ground floor automation to decision making in the domain of agriculture Vincent et al. [2019].

The big spheres where IoT systems can revolutionize agriculture are: Precision farming, Agriculture drones, Live stockmonitoring and Smart greenhouse.

Precision farming

Precision farming is the process of making agriculture more accurate, precise and controlled for animals husbandry and crop production. This is achieved through the use of IoT devices such as sensors, autonomous vehicles, automated hardware, control systems, and robotics.

In this sector IoT offers services like soil moisture probs and optimization of VRI (Variable Rate Irrigation), etc. This had huge impact on improving crop production.

Agriculture drones

Drones are used to control the field and plant growth, by taking images that can be useful for information collection about the entire farm land area. Combining the collected data from drones and sensors will create detailed digital maps of specific field areas Hoeren and Kolany-Raiser [2017].

Figure 2.3: Drones in Agriculture

Live stockmonitoring

Monitoring, tracking, and controlling farm animals (cows, goats, chickens, etc.) in open grasslands or indoor locations such as cages or stables. IoT is also used to monitor animal toxic gas levels, study ventilation, and warn on air quality to protect farm animals from harmful gases emitted from excrements Rayes and Salam [2017].

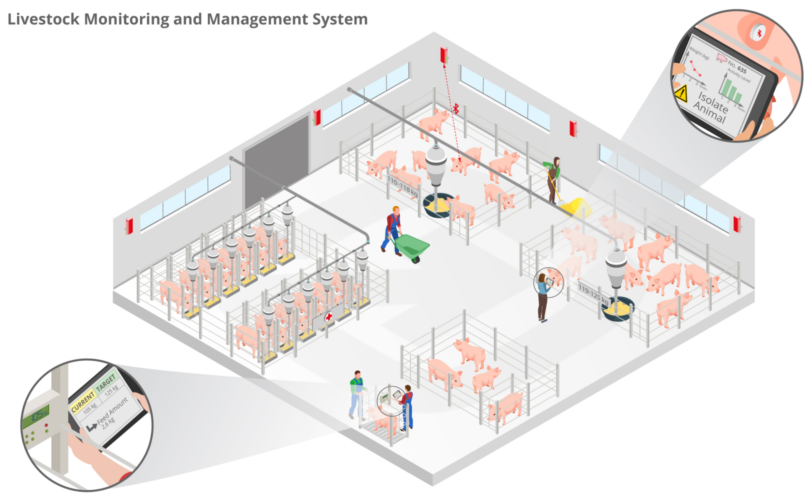

Figure 2.4: Livestock monitoring

Smart greenhouse

Smart greenhouse helps farmers to carry out the work in a farm automatically without the use of much manual inspection and the least of human interventions Kodali et al. [2016]. Tasks can be automated such as irrigation and water controlling, regulating temperature and weather conditions Raviteja and Supriya [2020].

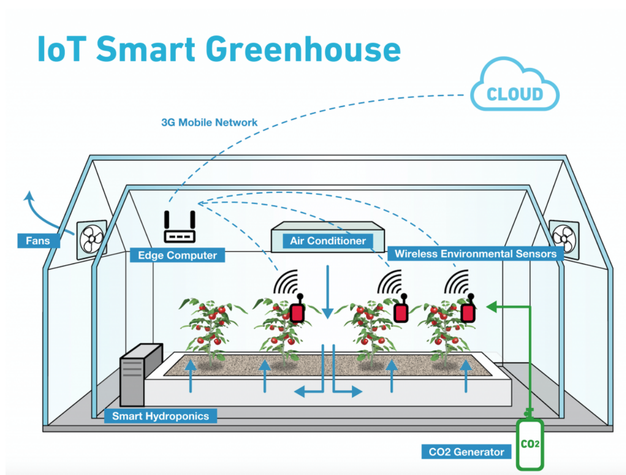

Figure 2.5: IoT in greenhouse (Smart Greenhouse)

2.5.2 Smart farm architectures in literature

Researchers have worked on smart farming, and proposed different architectures. In Vincent et al. [2019], the author proposed the following architecture for monitoring the field, which consists of 3 layers as illustrated in Figure2.5.2:

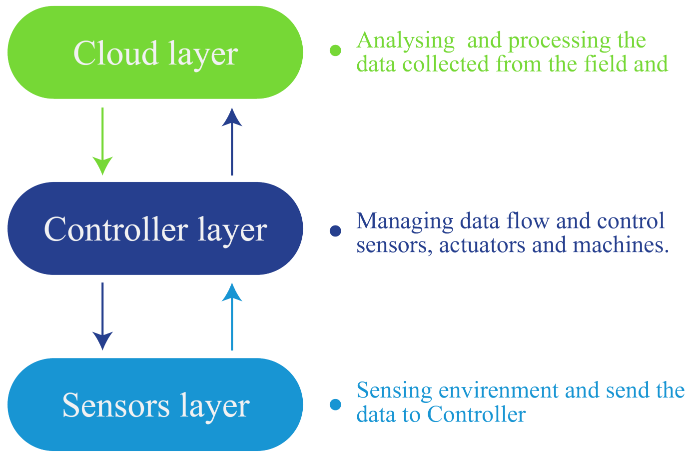

Figure 2.6: Smart Farming architecture

- **Sensors Layer:** deploying sensors on the field for data collection (pH, soil moisture, salinity, electromagnetic...).
- **Controller layer:** In this layer, we find the entity that is responsible for controlling the field and manages the data, and then sends it to the next layer for further processing and analysis. We distinguish different types of controllers: Raspberry Pi 3, Arduino, NodeMCU, etc.
- **Cloud layer:** This layer is responsible for performing the cloud tasks such as: data warehousing, data processing and analysis, and then delivers insight and results to the user.

2.5.3 Comparison

Due to the population growth, traditional farm couldn't fulfill the needs anymore, and the agricultural sector suffered from different issues Ziska et al. [2016]:

- Productivity was going slower, as almost everything relied on farmers and traditional machines.
- Climate change effects productivity, since agriculture is highly dependent on the climate.
- The need for water and a better controlling and management system for irrigation.
- Fewer people were going into the industry, due to the impact of urbanization on the rural labor supply.
- Livestock difficulties such as the absence of full time monitoring, which also may reduce the quality of their food supply.
- Shortage in arable lands due to the limited availability.
- High energy consumption and the inefficient approaches used in the domain.

Figure 2.7: Traditional farming vs. Smart farming

Dealing with the aforementioned issues using the traditional ways is difficult. Therefor, IoT and smart farming concepts are important Hoeren and Kolany-Raiser [2017], for tackling the challenges in the way of agricultural productivity Kamilaris et al. [2017].

2.6 Farm process

Business process was defined in Wolfert et al. [2017] as a set of logically related tasks performed to achieve a defined business outcome. Business processes are organized under two sub-categories: primary and supporting business processes. Primary Business Processes consist of creating the product, marketing and finally delivering to the buyer (client). In the other hand, Supporting Business Processes facilitate the development, deployment and maintenance of resources needed in primary processes.

In farming, business processes vary between different types of production, e.g. livestock farming, arable farming and greenhouse cultivation. A common feature is that agricultural production is depending on natural conditions, such as climate (day length and temperature), soil, pests, diseases and weather.

2.7 Farm management

Management or control processes ensure achieving business process objectives regardless of any disturbance. The process is based on a controller that measures system behavior and corrects if measurements are not compliant with its objectives Wolfert et al. [2017]. Thus, the process implies a feedback loop in which a norm, sensor, discriminator, decision maker are present.

2.8 Big Data in smart farming

Big Data technologies are playing an essential, reciprocal role in this farming industry, machines are equipped with all kind of sensors that measure data in their environment that is used for the automation of machines' behavior Wolfert et al. [2017]. Therefor, devices in smart farming generate huge amounts of data everyday, hence producing big data as shown in the Figure2.8 below.

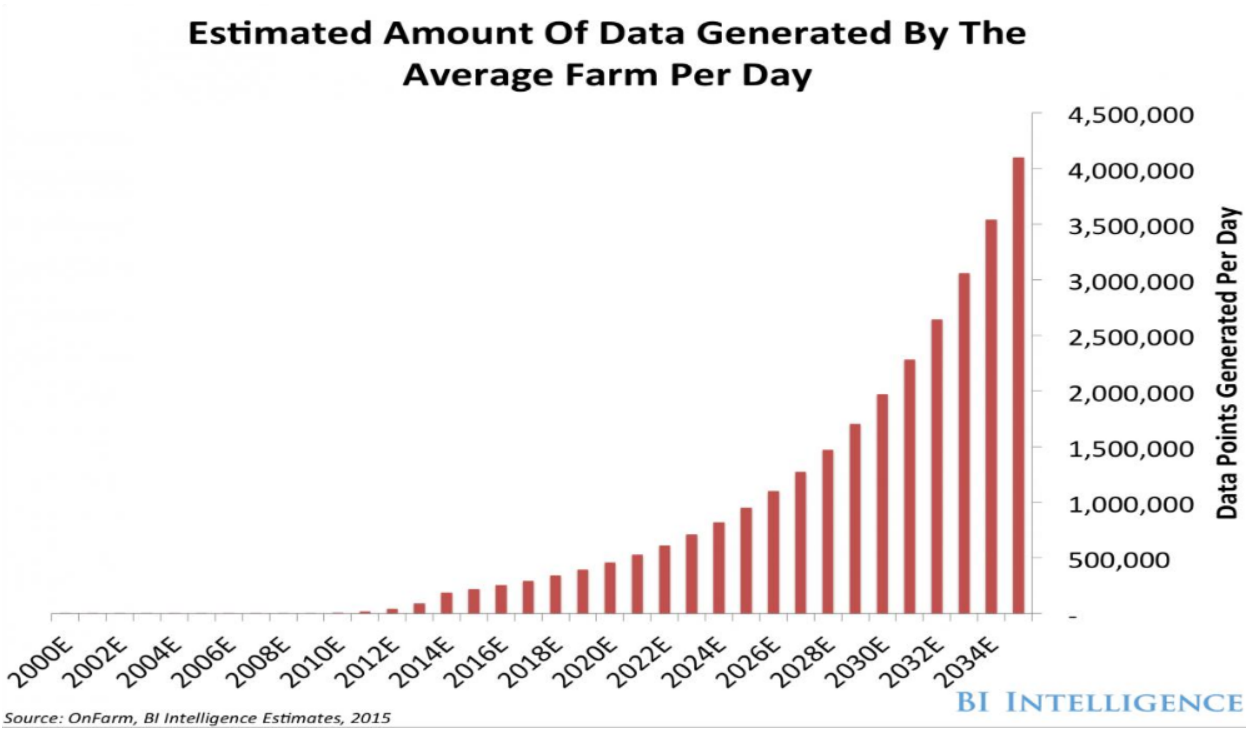

Figure 2.8: Data generated from farming

Chen et al, defined Big data in Chen et al. [2014] as: *"the datasets that could not be perceived, acquired, managed, and processed by traditional IT and software/hardware tools within a tolerable time."* Chen et al. [2014].

In Chi et al. [2016], Chi characterized big data with the 5 dimensions, as shown in figure 2.9

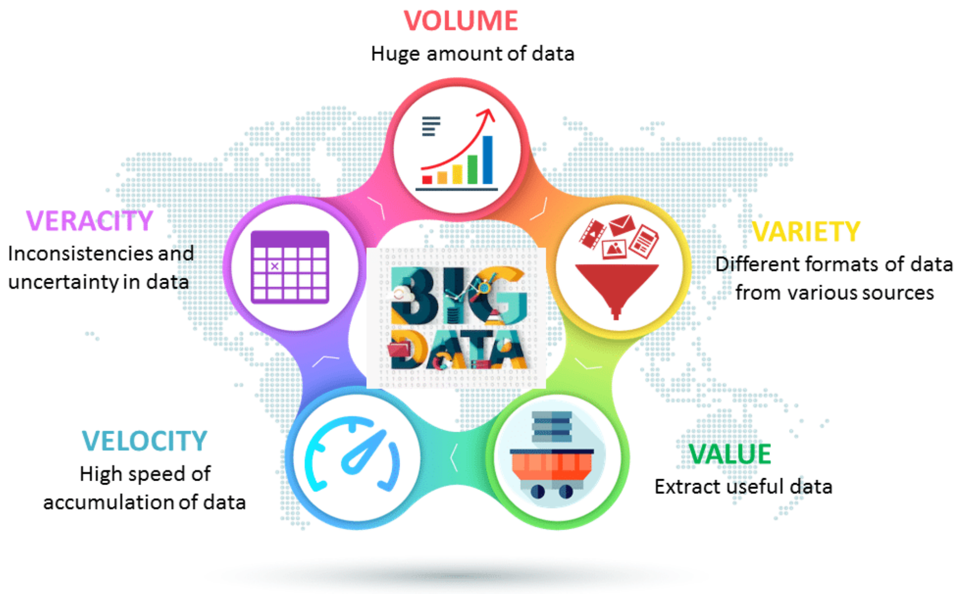

Figure 2.9: 5 Vs of Big Data

- **Volume:** the size of the collected data, from terabytes (TB) to petabytes (PB) and also exabytes (EB) Chi et al. [2014].
- **Variety:** big data can be collected from different sources, at different times and data varies from one discipline to another depending on the application field Chi et al. [2016].
- **Velocity:** the rapidity is not only restricted to data generation from sources, it also involves the efficiency of data processing and analysis Chi et al. [2016]. Sometimes data must be analyzed at real time, e.g., critical systems, driverless cars and medical fields.
- **Veracity:** it represents the quality, accuracy and how reliable the data is Kamilaris et al. [2017].
- **Value:** Although data is being produced in large volumes, it would be of no use if we don't extract knowledge and valuable information using data analysis methods Kamilaris et al. [2017].

However, IoT data is quiet different from the general big data. Accoridng to Mohammadi, in Mohammadi et al. [2018] he characterized IoT data as following:

- Large-Scale Streaming Data, the enormous number of sensors deployed on the field and the continuous streaming, results on producing huge volumes of data.
- High noise data due to its nature. Transmitting huge number of small pieces of data may lead to errors and noise.
- Time and space correlation, as data is labeled with a timestamp and a location according to the sensor.
- Heterogeneity due to the variety of IoT devices.

After collecting the big data, it requires a set of specific techniques, technologies and analytical methods for its transformation into value Hashem et al. [2015], in order to reveal insights from data sets that are diverse, complex, and of a massive scale.

2.9 Data Streaming and Data Analytics

2.9.1 Data Streaming

As mentioned earlier, the data must be collected, analyzed and then processed through various techniques and methods Hoeren and Kolany-Raiser [2017], which enable large-scale data processing based on real-time streams of data coming from a variety of sources.

Data streaming has become an essential part of all data based systems. Modern data is generated by an infinite amount of sources whether it's from hardware sensors, servers, mobile devices, applications, web browsers, internal and external and it's almost impossible to regulate or enforce the data structure or control the volume and frequency of the data generated. Relying on old ordinary methods can lead to failures and data loss. Applying data streaming new techniques and technologies is beneficial in such scenarios, where data is generated on a continual basis. Initially, applications process data streams to produce simple reports, and perform simple actions in response. Eventually, those applications perform more sophisticated forms of data analysis, like applying machine/deep learning algorithms, and extract deeper insights from the data.

Data Streaming examples

- **Sensor based systems and environments:** IoT and sensors can be found on transportation vehicles, industrial equipment, and farm machinery (Smart Farming). Those machines are responsible for collecting data and forwarding it to the application for further processing.

- **Online streaming services:** Streaming services are taking over most of the internet traffic Cullen [VP of Global Marketing], while providing different applications such as: Video streaming (Netflix & YouTube) and Gaming (online gaming & streaming) . They also collect (upload) data about their users and feed it back to the platform for data analysis and enriching users data base.

New methods and frameworks have been developed for applying Data streaming on different fields, as it is considered as the first step toward an efficient Data analytics process. Delivering a complete data sets with fault tolerance systems will produce more accurate results and insights.

2.9.2 Data Analytics

Analytics and its related, more recent term, data science, are key factors by which Big Data capabilities can actually contribute to improved performance in the agricultural sector, which aim to extract information and insights, that could not be easily achieved previously Kamilaris et al. [2017]. Thus, improving productivity and reducing environmental footprint Kamilaris et al. [2016].

In Manyika et al. [2013], the author applied data analysis on the collected data set, he divided the process into 4 steps as follows:

- **Data pre-processing:** the captured data may be incomplete or inconsistent, which may affect the results. Thus, this step is required to enhance the quality of data and to improve accuracy and efficiency.

- **Data reduction:** Reduce the data to a smaller representation that has the same integrity as the original data.

- **Data modeling/discovery:** For identifying patterns, algorithms can be used e.g., Apriori algorithm, that helps discover relationships between subjectively unrelated agriculture data.

- **Data Solution analysis:** Analysis of the results made by the data modeling/discovery step.

Therefore, knowledge from science will need to be effectively integrated within efforts to accomplish the goals of predictive and prescriptive analytics. Even with this additional complication, the potential of tools based upon emerging data science capabilities offers significant promise to more effectively optimize operations and create value within the agricultural sector.

2.10 Conclusion

The process of moving toward a better and more efficient productivity in the agricultural sector required the implication of modern technologies. Agriculture field can be understood, controlled and improved with these technologies such as IoT and Big data analysis.

Chapter 3

Related work in smart farming

3.1 Problem Statement: Land suitability

Usually new farmers to the agriculture field lack enough knowledge about the characteristics of soil for crop cultivation Somov et al. [2018]. They are not aware of the fact that agriculture land needs to be assessed before cultivation. Therefore, analyzing the land suitability becomes a mandatory prerequisite for crop cultivation, which leads to maximizing production. Farmers used to rely on manual data collection and soil testing labs, in order to acquire the properties of the soil, such method may not be sufficient enough to help them, and sometimes the data lacks accuracy.

Land suitability for agriculture and plantation can be determined by its elements. However, land elements are overused and exploited. Many lands are facing different problems like soil erosion, water logging, groundwater depletion, heavy run-off, productivity losses, etc. These lands degrades are threatening the food and energy securities, water availability and quality, biodiversity, and human life.

3.2 Weather factor in land suitability

Weather condition can significantly affect agricultural lands. The effect is because a change in the weather condition affects agricultural practices to thrive and yield increase. Some of these parameters include temperature, nutrient levels, soil moisture, and water availability to crops and livestock. Studies of crop yields have indicated that high temperature extremes, which lead to high vapor pressure deficits, can decrease rainfed crop yields in a variety of crops and regions. Also, the concentration of CO2 increases light intensity Bradford et al. [2017]. The climate has effects on them, though in different ways. These effects should be appropriately understood. As to know the extent of their influence on agricultural output and possible ways to curtail these effects. Climate change is expected to have a negative impact and decrease agricultural suitability at the global scale, although many studies considered the most negative outcomes are likely to occur in tropical and sub-tropical systems. However, in temperate regions, climate change impacts on agricultural suitability are uncertain, as in temperate regions that support a majority of global agricultural land tend to have higher yields than tropical areas due in large part to higher soil fertility, as shown in the figure below.

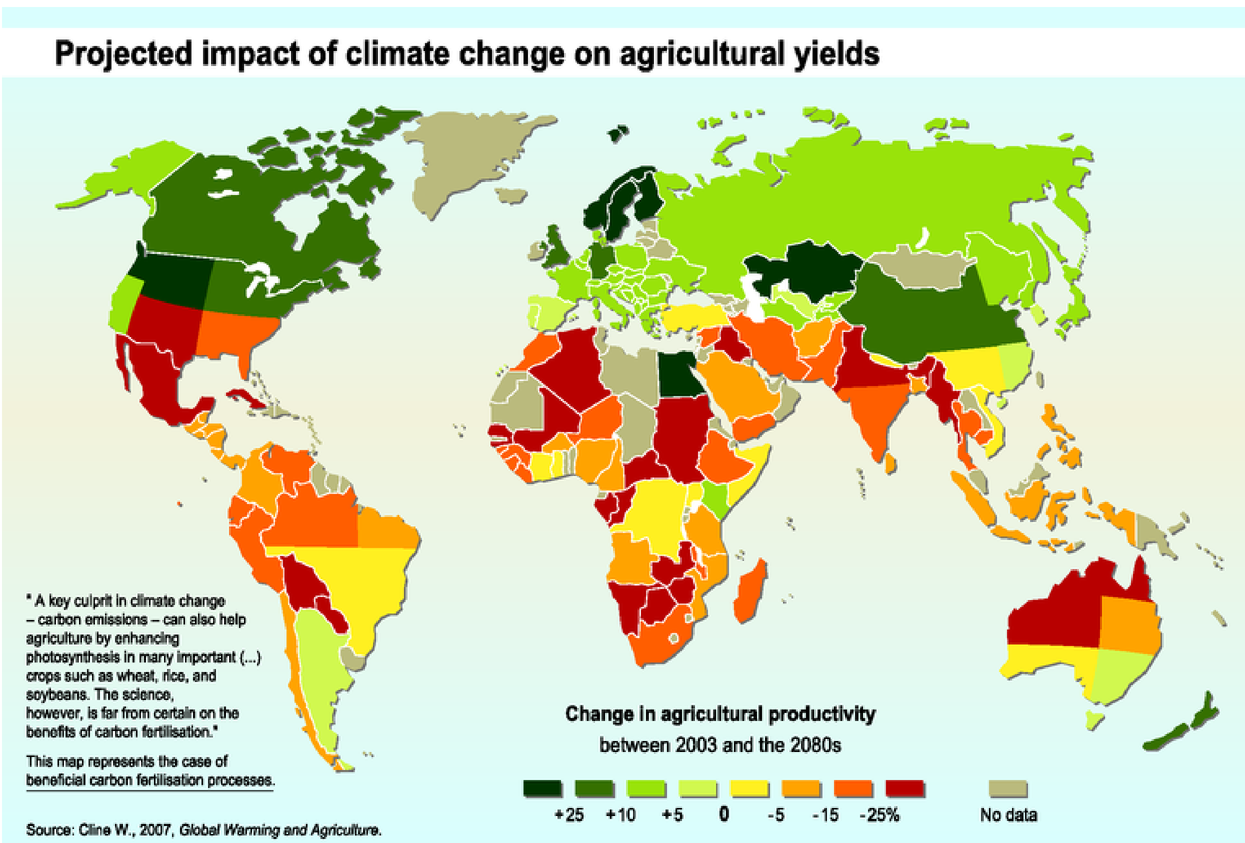

Figure 3.1: Climate change impact on crop yields

The map shows projected national productivity increase/decrease by 2080 (compared to 2003 levels). It is likely that some divergence will occur among regions concerning agricultural production. Under bird's eye view, countries falling within the tropics and sub tropical zones, with developing economies (e.g. Algeria), shall be losing in terms of agricultural production, whereas countries in temperate zones, sharing developed economies, are considered to gain. Most of the developing countries are highly dependent on developed economies for the production and exports of agricultural goods, and climate change is anticipated to cause significant losses in terms of growth as well as export opportunities Wheeler and Von Braun [2013].

Another factor that can affect land suitability and agriculture productivity is Humidity. In order to keep the stomata open, it is important to reduce the evaporation of the plant when there is more irradiation. By keeping the humidity in the environment high, evaporation will be reduced. In addition, the temperature can be lowered by introducing humidity, as a result of which the plant has to cool less through evaporation. Finally, the crop can be slightly moistened so that the evaporating water can cool the crop or the farm temperature De Goeij and Soeters [2016].

Figure 3.2: Stomata opening from Bad to Good

3.3 Related work

Some of the related works, and which our project is based on, are the following:

3.3.1 Sensors Driven AI-Based Agriculture Recommendation Model for Assessing Land Suitability

In Vincent et al. [2019], the author introduced a solution for integrating AI in the agriculture domain specifically in land suitability analysis, to improve productivity. In order to achieve this goal, he proposed a model (Neural network, MLP) combined with IoT. The output would be the affiliation of the land to one of the classes (Most suitable, Suitable, Moderately suitable, Unsuitable). The proposed solution consists of three layers:

- **Sensors**: they are responsible for collecting data (input data) from the farm. The used sensors are: pH, soil moisture, salinity, and electromagnetic sensor.
- **Raspberry Pi 3**: handles the collected data from the sensors installed on the field, and moves it to a cloud storage through the internet.
- **AWS Cloud network**: the proposed model performs the classification of the land suitability based on the collected data, after the training phase.

The proposed model was accurate at 99%, and the MLP (Multi Layer Perceptron with 4 hidden layers) offered better results and more accuracy for classification of land suitability when compared to NN (Neural Network).

3.3.2 IoT and agriculture data analysis for smart farm

Another work has been done by Muangprathub in this field. In Muangprathub et al. [2019], the author applied optimization on agriculture in order to enhance productivity, improve quality and reduce cost, by monitoring temperature, pressure and soil moisture. In order to achieve this goal, the author used the Internet of Things (IoT) technology to construct a Wireless Sensor Network (WSN). The system is mainly composed of three parts:

- **Control Box (Hardware)**: controller of IoT devices implemented on the crop fields, such as sensors (temperature, humidity, soil moisture and MCUnode. . . etc.).
- **Web-based application**: this application is used to visualize real-time data captured by “IoT” devices (sensors) and to manage watering for the agriculture fields.

- **Mobile application**: this is an end user application for the farmer to allow Automatic/-Manual control of watering on the fields. It can be automated based on the analyzed data, or manually by switching on/off by the farmer.

In order to extract knowledge from the data captured by the system, the author applied data analysis methods. The process of data analysis goes through four major steps:

- **Data pre-processing**: The captured data may be incomplete or inconsistent, which may affect results. Thus, this step is required to enhance the quality of data and to improve accuracy and efficiency.

- **Data reduction**: Reduce the data to a smaller representation that has the same integrity as the original data using numerosity reduction.

- **Data modeling/discovery**: For data analysis and identifying patterns, the author used Apriori algorithm to extract Association rules (if/then statements) that help discover relationships between subjectively unrelated agriculture data.

- **Solution analysis**: Analysis of the results made by the data modeling/discovery step.

The author contribution was applying data mining using Association rules technique to extract knowledge and information about the effects of environment and climate on the crop yields.

3.4 Synthesis

In this section, we present a comparison between the mentioned work presented in this chapter. This comparison stands on five different parameters which are : domain, objective of the work, criterion which covers the additional setting to the approach such as sensors and so on. In addition, we add the used method and data type for each work which are the main point in our work due to their importance especially in the process of decision making.

Table 3.1: Related work comparison

	Muangprathub et al. [2019]	Vincent et al. [2019]	**Our work**
Domain	Agriculture	Agriculture	Agriculture
Objective	Water automation	Land suitability classification	Land suitability classification and field monitoring
Criterion	Temperature and humidity	pH, soil moisture, salinity	Temperature, humidity, pressure, and precipitation
Data type	Real time and stored data	Stored data	Real time and stored data
Method	Association rules (if/then)	Neural Network and Multi-layer perceptron	Long-Short Term Memory (LSTM RNN)

3.5 Conclusion

We deduce that land suitability is a major field of studies that affects agriculture productivity. Finding the best land ensures better crops quality. Therefor, our work focuses on studying the properties of the land and defining its suitability for the plant.

Chapter 4

Design and Contribution

4.1 Introduction

Our aim is to construct a robust system that delivers real time analytics, historical analysis and data visualization in an intelligent way by combining artificial intelligence with IoT and Data analytics. In this chapter, we are going to discuss our system architecture and the role of each layer in details, also the algorithm used to achieve land suitability prediction.

4.2 Proposed architecture

In order to achieve our goal, we proposed an architecture composed of three layers, interacting with each other and exchanging data in real time. Each layer is responsible for a specific task that outputs data for the next layer starting from the physical field of the plant (the farm). The first layer acquires the data as numerical values from the farming field, and then it passes it to the second layer for analytics and data visualization. This last, stores the historical data to the cloud, while doing the real time analytics locally. The third layer is considered as a storage facility that also provides the historical data analysis/analytics service. The global system architecture is illustrated in the figure below.

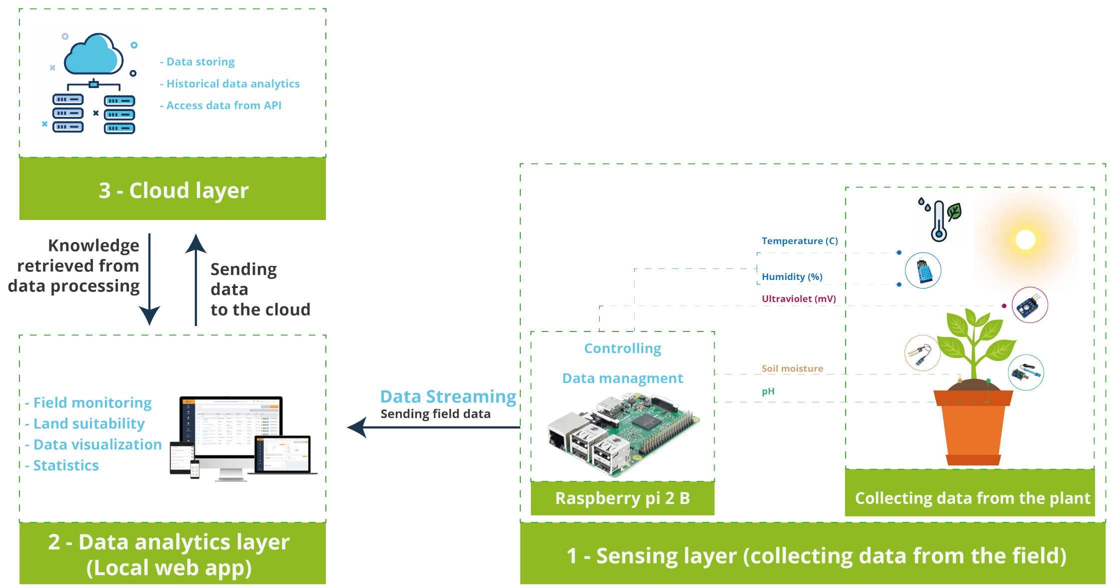

Figure 4.1: The proposed architecture

This approach prioritizes real-time analytics over the historical one. It delivers information based on the received streams of data. The process and the role of each layer will be discussed in the next section.

4.2.1 Architecture description

1. **Sensing layer:** This layer represents the physical part of the system. It is responsible for sensing the environment of the farm and reporting back to the server. The sensing process is achieved, as mentioned earlier, through the deployed sensors on the field (Temperature sensor, Humidity sensor, Soil moisture sensor, UV sensor). All the sensors are controlled by the *Raspberry Pi*, which gathers all the data and redirects it to the server. It can be scheduled to collect data weekly, daily, hourly...etc. The plant will be monitored all the time, which results in generating huge amounts of data that will be passed to the next layer for analytics and processing.

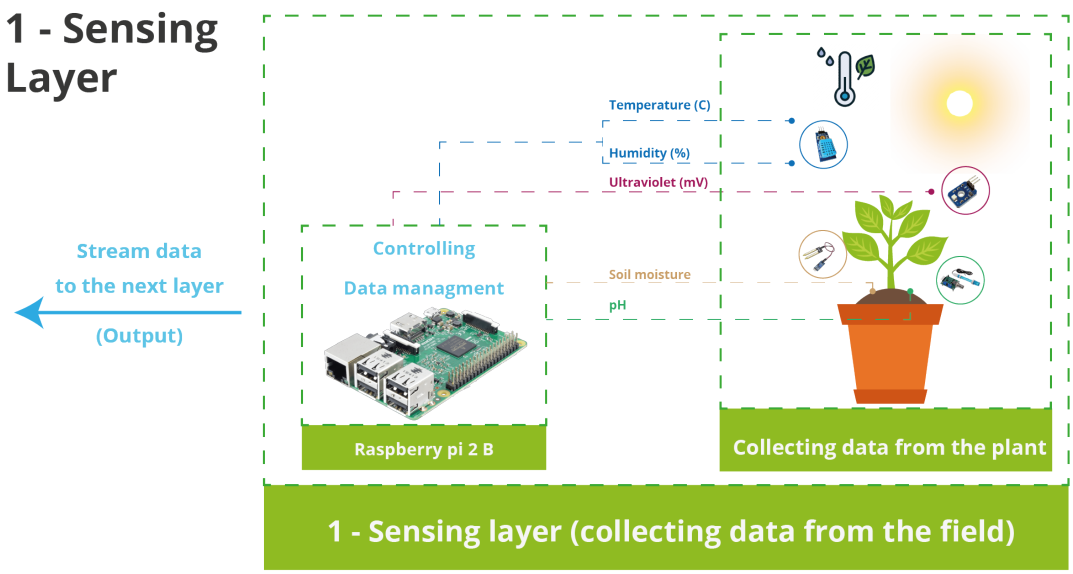

Figure 4.2: 1 - Sensing layer (collecting data from the field)

The sensing layer produces streams of data that are generated in real time, which describe the current status of the environment of the plant. These streams are outputted directly to the server using streaming methods, such as *KAFKA* to be handled by the Data Analytics layer.

2. **Data analytics layer:** After the data acquisition process, all information will be passed to this layer which is responsible for data analytics and field monitoring. The collected data must be stored locally at this layer, data structures are required such as Data bases. The data base contains tables that defines different types of the stored data, the main historical part is structured in Fig 4.3.

	Unnamed: 0	Year	Month	Day	Precipitation	Relative Humidity	Pressure	Temp_avg	Temp_min	Temp_max	Suitability
0	0	2009	January	1	0.15	72.07	99.43	9.86	4.90	14.92	Moderately Suitable
1	1	2009	January	2	0.11	63.27	98.94	11.45	8.61	15.95	Moderately Suitable
2	2	2009	January	3	0.12	72.29	98.53	11.41	6.22	16.77	Moderately Suitable
3	3	2009	January	4	3.98	78.63	98.14	11.20	8.31	16.33	Moderately Suitable
4	4	2009	January	5	0.08	65.45	98.23	9.34	5.82	15.31	Moderately Suitable
...	...	...	...	...	...	...	...	...	...	...	...
4103	4103	2020	March	27	1.58	55.19	98.00	13.30	5.96	19.61	Suitable
4104	4104	2020	March	28	8.38	70.06	98.21	13.47	10.23	17.83	Suitable
4105	4105	2020	March	29	0.91	54.82	98.26	13.35	7.95	19.35	Suitable
4106	4106	2020	March	30	0.00	54.72	98.46	13.30	6.24	19.79	Suitable
4107	4107	2020	March	31	0.07	47.62	98.38	15.07	9.20	21.90	Suitable

Figure 4.3: Historical data of the field

At this level, the farmer has the ability to monitor the farming field through the web platform. The web platform provides different services, such as a dashboard for visualizing and monitoring the field, giving real time information about the weather, the land, and crops production. The data analytics layer also provides the land suitability service, which decides whether the land is suitable for a specific plant or not using artificial intelligence (Deep Learning), based on the collected data earlier (the process will be discussed in details in the next section).

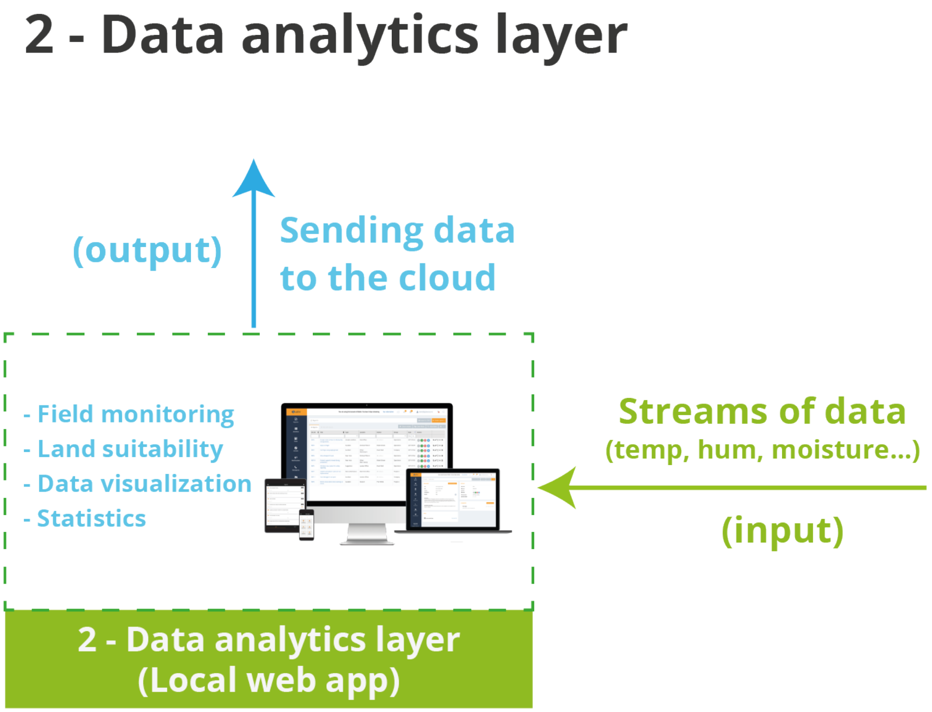

Figure 4.4: 2 - Data Analytics layer

3. **Cloud layer:** The third layer, is cloud based. It is a cloud server, which extends the local server abilities, providing the platform with more functionalities and features that can be used through the internet, such as:

   - **Cloud storage:** The sensing layer produces huge amounts of data that cannot be stored on a local server only. Thus, having a cloud server, *Google Drive* in our case, allows the farmer to further extend the storage capacity of the platform.

   - **Historical data analytics:** The stored data on the cloud will be used for historical analytics/analysis by extracting meaningful patterns from the collected data over the years. The larger the dataset is, the longer the process takes.

   - **Access data through API:** Having the data stored on the cloud allows the platform to access it from anywhere through the internet with API requests. The data will then be sent in structured data format for later use (JSON, XML...).

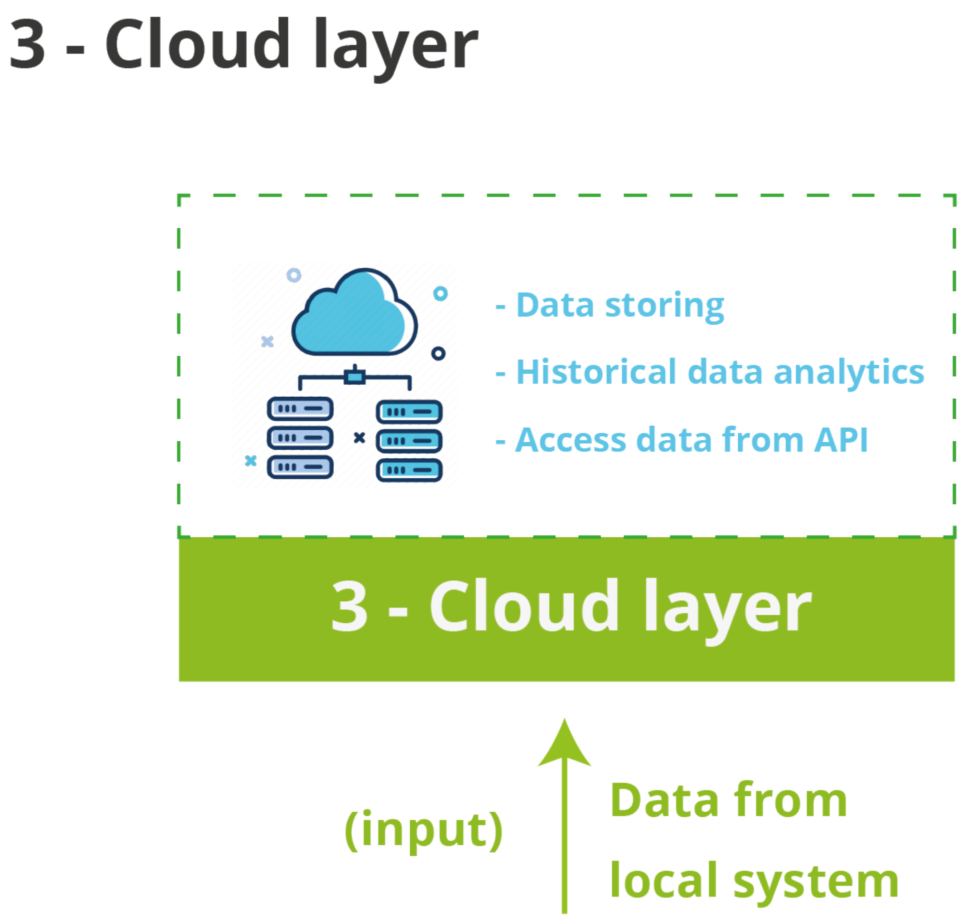

Figure 4.5: 3 - Cloud layer

4.2.2 Land suitability process

In order to make a decision about the land suitability, the process must run through 4 steps.

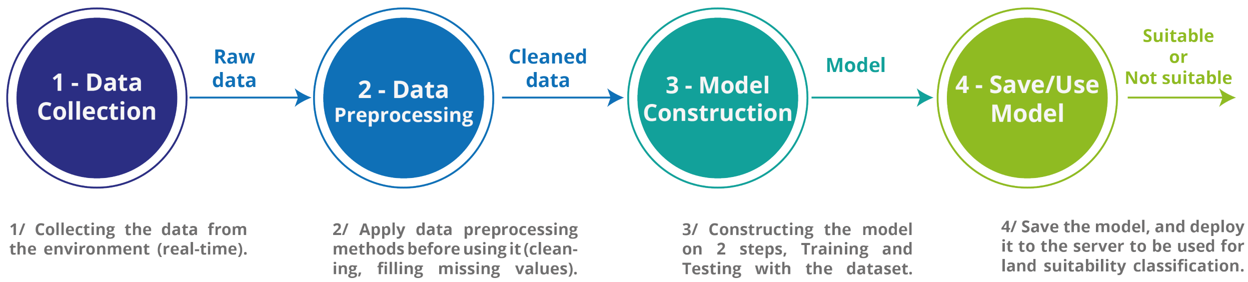

Figure 4.6: Land suitability process

1. **Data collection:** The decision system relies on the collected data. In our case the data is the numerical values which describe our farming environment. The output of this process is the collected data (raw data), it is stored as JSON format and then sent to the local server via streaming framework (KAFKA).

2. **Data pre-processing:** Collecting huge amounts of small numerical values will surely produce incomplete datasets. The raw data can have missing values which will result in error if directly fed to the model. Therefor, a pre-processing phase is necessary for cleaning the data. For increasing accuracy and efficiency, the data must be normalized into the range (0, 1). The output will be clean consistent dataset ready to be used with the model.

3. **Model construction:** Choosing the best model and architecture has a major impact on prediction results. Among the various Machine/Deep learning methods and algorithms, we chose an algorithm with time series consideration. This phase goes on 2 steps:

   (a) Constructing the model and architecture (number of layers, number of neurones, activation functions...etc.).

   (b) Training the model and testing it (making predictions to test the model).

   The algorithm will be discussed in details in "Used Algorithm" section.

4. **Save/Use model:** After creating the model, it must be saved (exported) and deployed to the web platform so that users can have access to the Land suitability service. After the

training and testing phases with the datasets, the model should now be able to make the decision about the suitability of the land for a specific plant, the output will be one of the 4 classes: *"Best suitability", "Suitable", "Moderately suitable" or "Unsuitable".*

4.2.3 UML Diagram

In this section, we illustrate our platform functionalities which the user can benefit from. Services and features are shown in the use case diagram below.

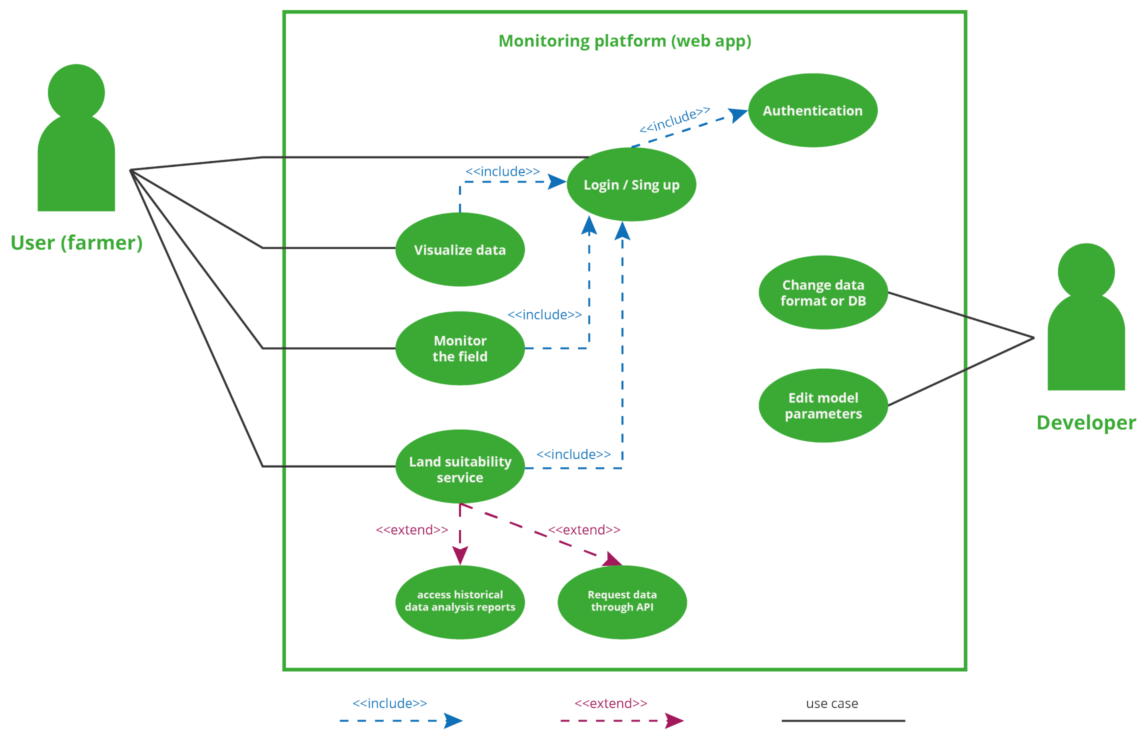

Figure 4.7: Use case diagram of the platform

4.3 Used Algorithm

4.3.1 Deep learning

Modern Artificial Intelligence systems rely on Neural Networks for decision making and predicitons. Deep learning or deep structured learning can be defined as special kind of neural networks composed of multiple layers, these layers perform better than the traditional Neural Networks (used in machine learning) when it comes to information persisting. Kumar et al. [2018]. One of the major advantages and differences of DL over ML, is the feature extraction step, which is achieved automatically as information propagates through the deep layers. However, DL still requires a lot more data than ML, which can be hard to provide in some situations.

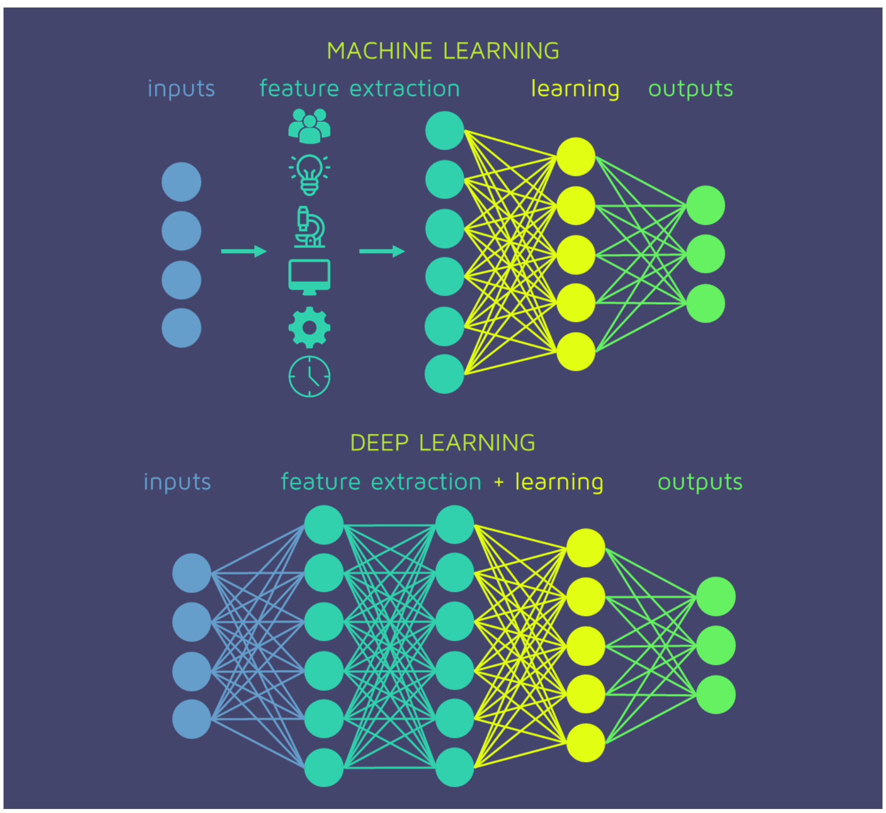

Figure 4.8: Machine Learning vs. Deep Learning

4.3.2 Reccurent Neural Network (RNN)

Reccurent Neural Network, in contrast to ordinary Neural Network, employs feedback loops where the output from each step (e.g. step n-1) is fed back to the network to affect the outcome of the current step (e.g.step n) based on parameters that define the impact of the previous outputs, this process is repeated for each subsequent step Al-Smadi et al. [2018]. RNNs have proven to achieve good results on sentence level sentiment analysis, speech recognition, time series data...etc.

4.3.3 Long-Short Term Memory network

Long-Short Term Memory (LSTM) Networks are a special kind of RNN. They were designed to address the issue of long term dependency in RNN. LSTMs are good in remembering information for long time. Since more previous information may affect the accuracy of model, LSTMs become a natural choice of use Al-Smadi et al. [2018]. LSTM is composed of a module called *"Repeating Module"*, it has four neural network layers interacting in a unique way. Repeating module also has three gate activation functions: σ_1, σ_2, σ_3 and two output activation functions ϕ_1 and ϕ_2 as shown in Figure 4.9.

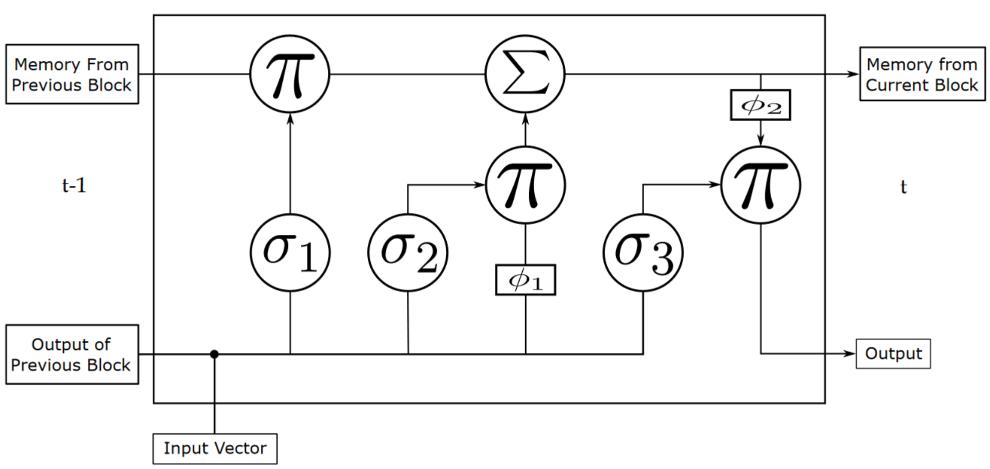

Figure 4.9: LSTM Repeating Module

Basically, choosing LSTM architecture was based on the type of data we have. Since we are dealing with time series, which means a data related to time and was recorded in a sequential way. LSTM proved to be the best in such scenarios, where it can remember some information from past records. This can be beneficial in our situation, because land suitability cannot be predicted based on scattered or unsorted data.

4.3.4 LSTM pseudocode and architecture

Process

Our algorithm works as follows:

1. ***Gathering data:*** this process is accomplished as explained earlier, through IoT devices (Sensors and Raspberry Pi).

2. ***Preparing data:*** before feeding the data to the model, it must run through preparation methods, filling missing values, data cleaning, normalization and finally reshape it according to the LSTM input (a 3-dimensional array).

3. ***Creating the model:*** choosing the right model is important for creating an efficient learning model. The creation process is accomplished with some parameters: number of neurons, number of layers, activation functions...etc.

4. ***Training:*** The dataset will be splitted in 2 parts. The first part will be used for training the model, in other words adjusting the weight in order for the predictions to match the expected results.

5. ***Evaluation:*** The second part of the dataset, will be used for testing and evaluating the model. The testing data is smaller and different from the training data.

6. ***Prediction:*** After the training and testing phases, the model can be used for making predictions, in our case about land suitability.

PseudoCode

Algorithm 1 LSTM prediction model pseudocode

```
normalize_data(0,1)                                   ▷ Normalizing data values in range 0 and 1
x_train, y_train, x_test, y_test ← split_data(data,25)    ▷ Split data to training and testing sets (25% testing)
reshape_data(data)                                    ▷ Reshape data according to LSTM input
model ← CreateSequentialModel()                       ▷ Creating and configuring the model
model.add_LSTM_layer(nbr_lstm, sigmoid)
model.add_NN_layers(nbr_nn)
model.compile()
for nbr_epochs do                                     ▷ Training LSTM model
    for batch_size do
        model.fit()
    end for
end for
results ← model.predict(x_test)                       ▷ Testing the model and making predictions
```

Architecture

Our model was created with an input layer, 3 hidden layers and 1 output layer. The input layer receives a 3-dimensional array of 4 features (4 columns), each feature is a 1-dimensional array.

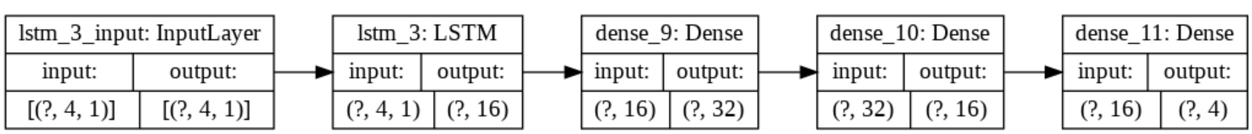

Figure 4.10: LSTM model architecture and layers

4.4 Conclusion

In this chapter we explained our architecture and the used algorithm. Our system is composed of Hardware part (Internet of Things) and a Software part (Web Platform with Deep learning, LSTM-RNN specifically), in order to bring a new approach for enhancing the farming field. In the next chapter we discuss the implementation and results.

Chapter 5

Implementation and results

5.1 Introduction

Application development process comes after a set of steps which their main goal is to develop a product capable of responding to the farmers and expert needs.

After presenting and discussing the theory part of our project and the details about the used approaches. We shall now go through our developed idea and the results obtained from it.

In this chapter, we present the used tools including the platforms used to implement our system. Next, we give some system's interfaces which give the obtained results. In the end, we give some discussions and analysis of the results given by the constructed models.

5.2 Development tools and used platforms

Our system is composed, as mentioned earlier, of two main parts: **Hardware** and **Software**. The hardware represents the physical part of the system deployed on the field, and the Software represents the core system and the server which resides on a computer or data centers.

In order to make a robust and an efficient system, we combined some of the best advanced technologies and programming languages in both parts, Hardware and Software. The used technologies are listed below with a brief definition and explanation.

5.2.1 Hardware

The hardware part is mostly composed of three entities:

Raspberry Pi 2 model B

The Raspberry Pi is a low cost, credit-card sized computer that plugs into a computer monitor or TV, and uses a standard keyboard and mouse ras [2015]. In our system, the Raspberry Pi is considered as the controller of all sensors, and the data collector.

- 100 Base Ethernet
- 4 USB ports
- 40 GPIO pins
- Full HDMI port
- Combined 3.5mm audio jack
- Micro SD card slot
- VideoCore IV 3D graphics core

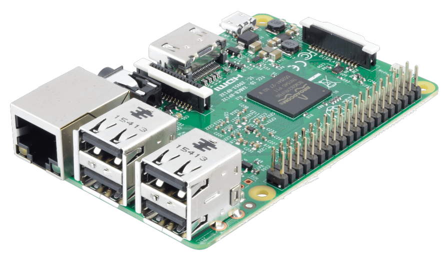

Figure 5.1: Raspberry Pi 2 model B

The connection between the Raspberry Pi and sensors is established with electricity wires using the **Breadboard** as an intermediate linking tool. An example of the device connected to the DHT11 sensor (temperature and humidity sensor) is illustrated below:

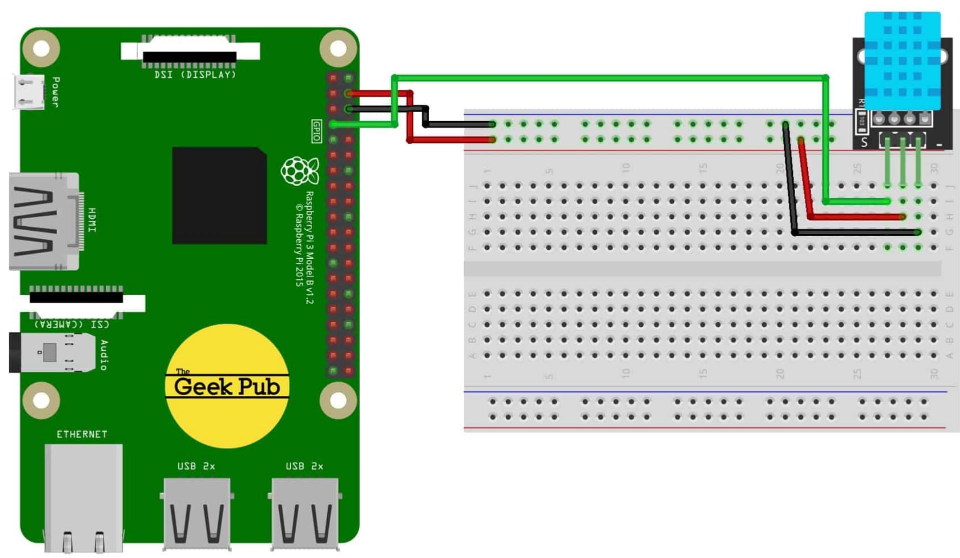

Figure 5.2: Raspberry Pi 2 connected to DHT11 sensor via the Breadboard

Arduino Uno R3

Arduino is an open-source electronics platform based on easy-to-use hardware and software. Arduino boards are able to read inputs - light on a sensor, a finger on a button and turn it into an output - activating a motor, turning on an LED, publishing something online. The board can be told what to do by sending a set of instructions to the microcontroller. To do so we use the Arduino programming language (based on Wiring), and the Arduino Software (IDE), based on Processing.

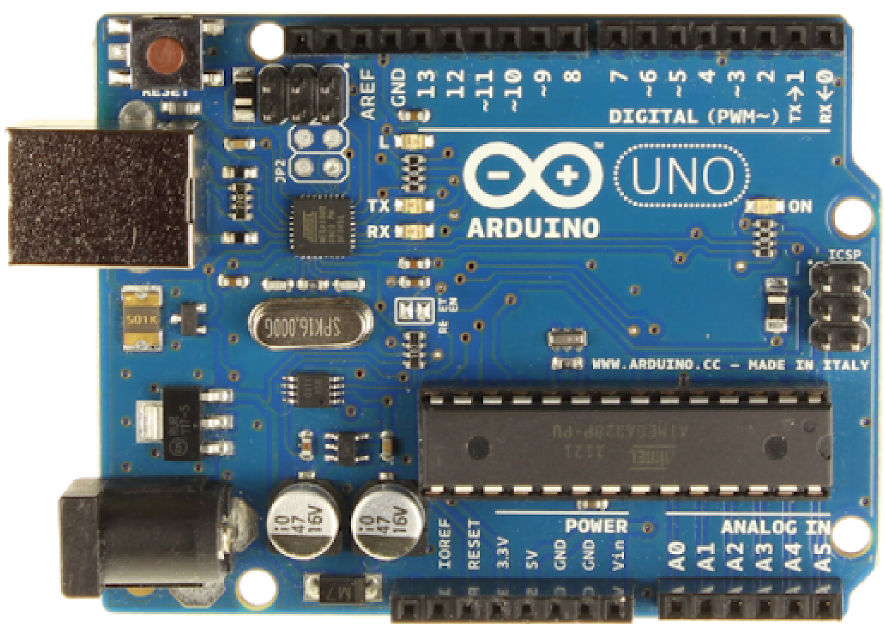

Figure 5.3: Arduino Uno R3

We used the Arduino along with the Raspberry Pi in order to handle the Analog outputs of some sensors with the Arduino and forward it back to the Rasoberry Pi.

Sensors

For the system to work on real-time, sensors are required. As mentioned in the previous chapters, sensors are small electronic devices deployed (or planted) on the field of the plant. We used these devices to capture different weather and land properties, to be used in the process of land suitability classification. The used sensors are listed in the table below:

Sensor	Place	Type of value
Temperature	Air	$\circ$ C
Humidity	Air	%
Pressure	Air	%
Soil moisture	Planted in soil	%
pH	Planted in soil	0-14

Table 5.1: Used sensors

We should notice that the data provided by the sensors is not identical to the historical data, and some properties cannot be collected in real-time, that's due to the lack of availability of sensors in our country (Biskra).

Server (personal computer)

For running the web platform (the server), we used a personal laptop with the following specs:

- **Ram:** 16 GB
- **CPU:** Intel Core i5-8350U
- **Storage:** 256 SSD

The laptop is able to handle all the coming streams of data from the Raspberry Pi, and process it. However, scaling the app for a large number of devices and sensors to cover a whole field may need a bigger server.

5.2.2 Software

Google Colab

Training a deep learning model can require extensive CPU/GPU workload, that's why we used Google Colab cloud platform for this task. Colaboratory is a Google research project created to help disseminate machine learning education and research. It's a Jupyter notebook environment that requires no setup to uses and runs entirely on cloud.

Figure 5.4: Google Colab

Visual Studio Code (Code Editor):

VS Code is a code editor made by Microsoft, it's a light-weight tool that allows the editing of source code files. It comes with a huge library of extensions that ease the work for developers. Our choice landed on VS Code, since our project contains different programming languages (python, html, js, css).

Figure 5.5: VS Code

Python & Django:

For developing the model (Deep learning model) and the web application, we used Python programming language. Python is an interpreted, object-oriented, high-level programming language with dynamic semantics pyt [2020].
It's high-level data structures, dynamic typing and the huge set of libraries makes it the best choice for AI and IoT programming.
Some of the most important libraries we used:

Figure 5.6: Python

- **Django**: is a web framework for creating dynamic and scalable web apps, we used it to create the server web app *Smart Green*.
- **Tensorflow & Keras**: are used for machine/deep learning to create and train the model.
- **Pandas**: used for all kind of data pre-processing, data analysis, data analytics, etc.

- **Kafka**: provides an easy API to interact with the kafka server from both parties (producer and consumer), for creating topics, sending and receiving messages.
- **Websockets**: used for exchanging data between the back-end server and the front-end dashboard, updating charts, accessing historical data api page, etc.

Charts.JS:

For the data visualization, we used a JavaScript library called "Chart.js". It is a free open-source JavaScript library for creating plots and charts, which supports 8 chart types: bar, line, area, pie (doughnut), bubble, radar, polar, and scatter. Created by London-based web developer Nick Downie in 2013.

Figure 5.7: Chrat.js

5.3 System interfaces

As mentioned in previous chapters, our app is Web based, farmers interact with different app interfaces. Each services has its own interface, which is designed specifically to be easy to use and to interact with.

5.3.1 Home page

The Home page is the first page to land on. It presents features and services of the system, the home page also provides access to other interfaces of the system. The navigation bar allows the user to Log in or Sign up, or to navigate to the dashboard...etc.

Figure 5.8: Smart Green Home page

5.3.2 Log in/Sign up

Users are able to log in to access the system services, if the user does not have an account, he can then Sin up.

Sign in/Log in

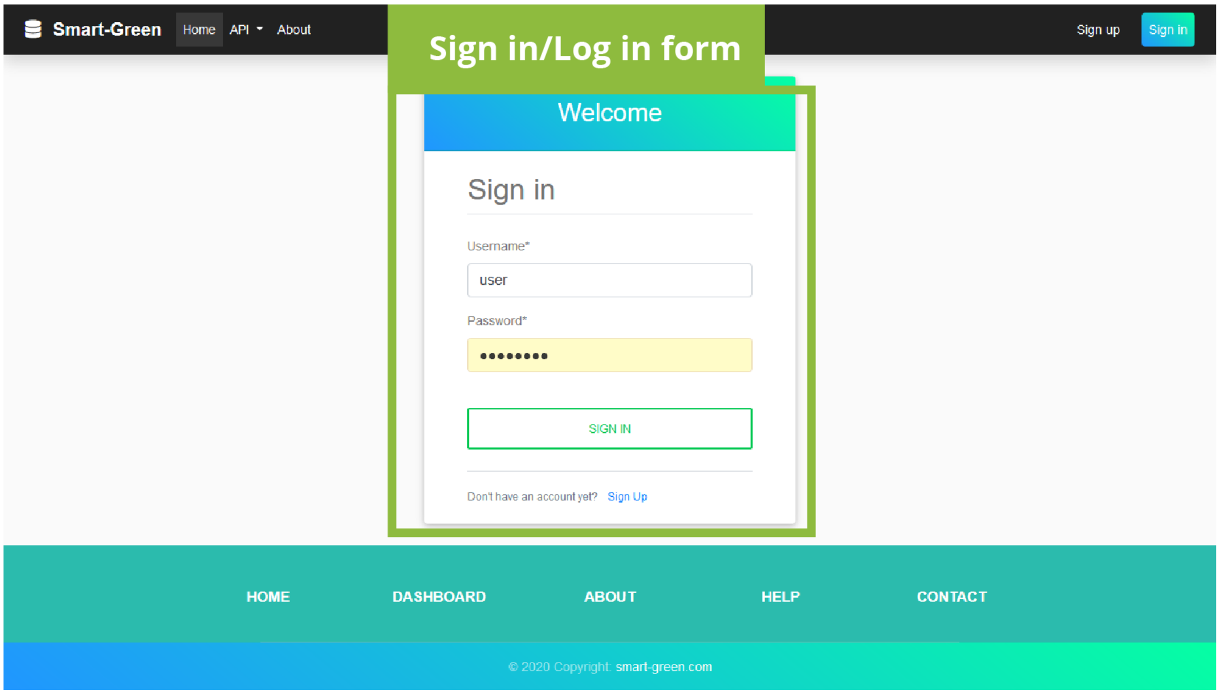

Figure 5.9: Sign in / Log in form

The login method allows the user to access the dashboard for the different services. All user data is stored on the local storage database. Information may contain: *Full name, username, password, email, profile image.*

Sign up

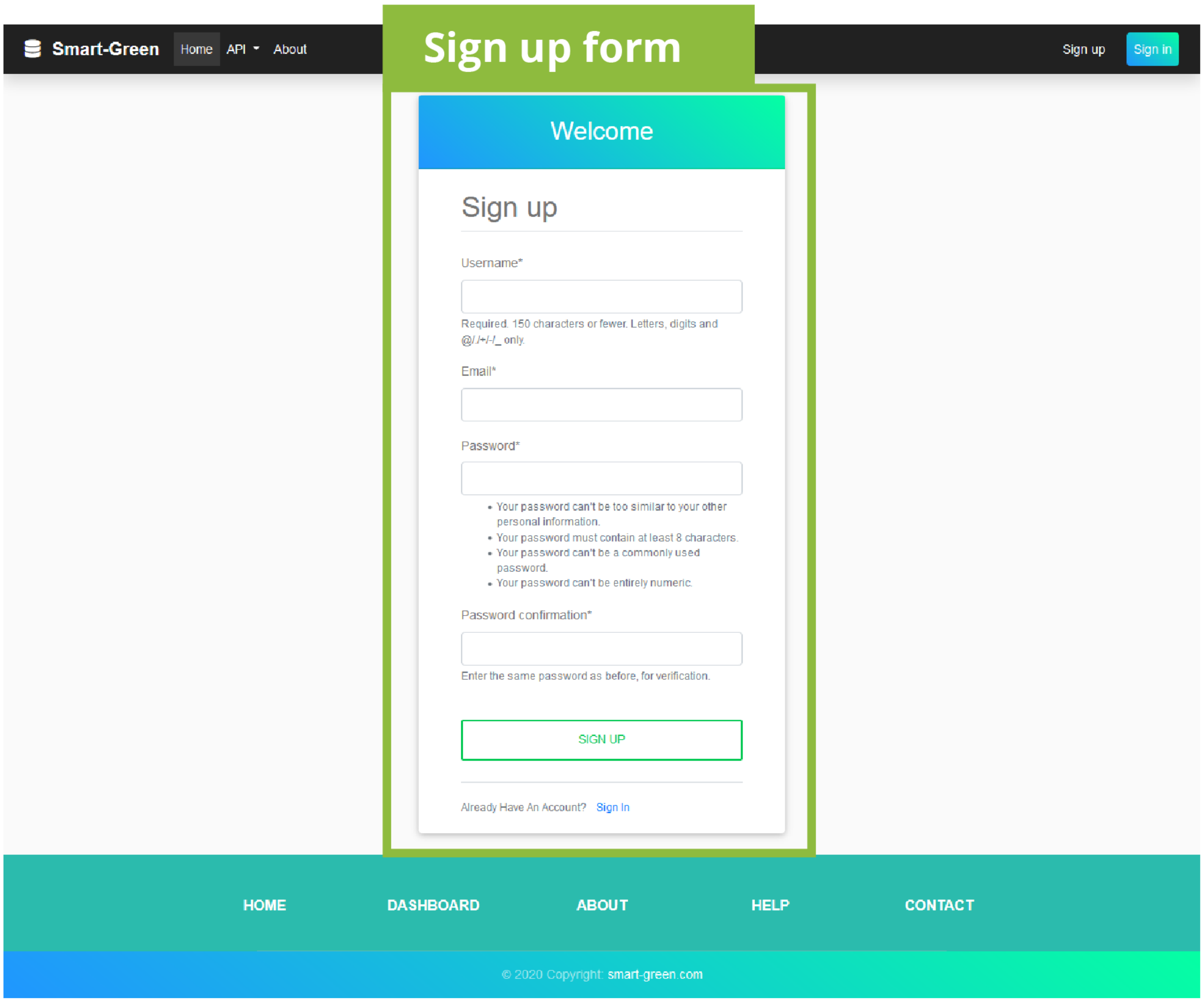

Figure 5.10: Sign up form

5.3.3 Dashboard

Real-time data visualization

Real time visualization allows the farmer to monitor the crop fields, receiving live streaming data from the deployed sensors. The Raspberry Pi sends sensors collected information through the Kafka framework, acting as a Producer of data. The code snippet below elaborates how Raspberry Pi is programmed to send the data to the server:

```
"""
    This code resides on the Raspberry Pi
"""
from kafka import KafkaProducer
from json import dumps
# start the producer
producer = KafkaProducer(bootstrap_servers='localhost:9092',
value_serializer=lambda msg: dumps(msg).encode('utf-8'))
# collect data from sensors
data = {
    'Temperature': getTemperature(),
    'Humidity': getHumidity(),
    'Moisture': getMoisture(),
    'pH': getPH()
}
# send data to consumer (server app)
producer.send('WeatherData', data)
producer.flush()
```

Listing 5.1: Raspberry Pi producer code

The collected data is then displayed as charts on the dashboard. This last, receives data from the field by listening to the kafka broker, whenever a message (data) arrives, the server automatically displays and updates charts on real-time, this task is illustrated on the code below:

```
"""
    This code resides on the Server app.
"""
from kafka import KafkaConsumer
from json import loads
from .websocket.websocket import start_websocket, updateCharts
#
def consumer_kafka():
    #   start kafka Consumer to receive data from the raspberry pi
    #   Raspberry pi (producer) ====> Server app (consumer)
    topic = 'WeatherData'
    consumer = KafkaConsumer(topic, bootstrap_servers=['localhost:9092
'])
    #   start websocket
    #   Server app ====> farmers dashboard
    start_websocket(updateCharts)
    print('Kafka Consumer started...')
```

```
        # receiving messages from producer
        for message in consumer:
            data = loads(message.value)
            # send data to the dashboard to update charts
            send_charts_data(data)

```

Listing 5.2: Server consumer code

The received streams of data are also stored on the local database of the server. Database management is done through Django ORM (Object Relational Mapping), which uses Object Oriented paradigm for manipulating databases, tables are structured as classes called Models, records as objects, etc. When the server receives data it creates a new object and stores it on the databases table.

```
    """
        Database model for daily values recorded from the raspberry pi
    """
    class Daily_real_time(models.Model):
        date = models.DateTimeField(db_column='Date', auto_now=True)
        temperature = models.IntegerField(db_column='Temperature')
        pressure = models.FloatField(db_column='Pressure')
        humidity = models.FloatField(db_column='Humidity')
        pH = models.FloatField(db_column='pH')
        moisture = models.FloatField(db_column='Moisture')

```

Listing 5.3: Real time data model

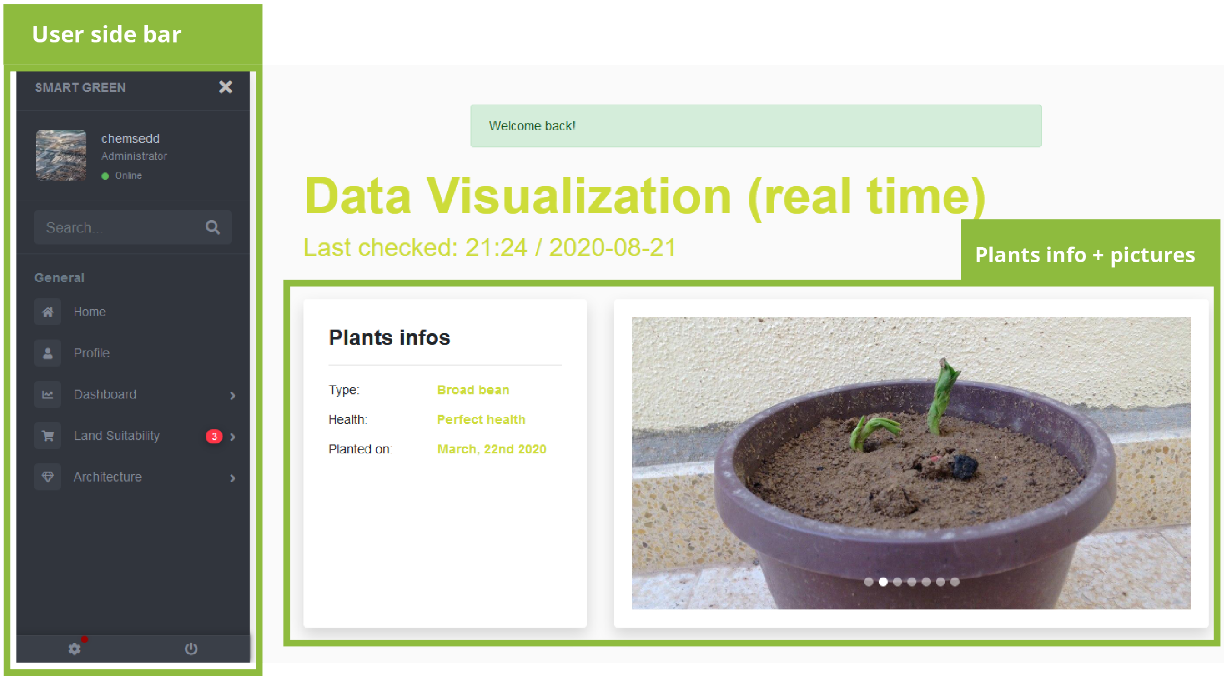

Figure 5.11: Smart Green Dashboard 1

Historical data

Data is collected and stored on the database, thus users will have access to it. In order to display historical data about the field, the user selects Month and Year corresponding to the date and the server will fetch the data from the database stored on the local server, and send back all information to the dashboard to be displayed in charts.

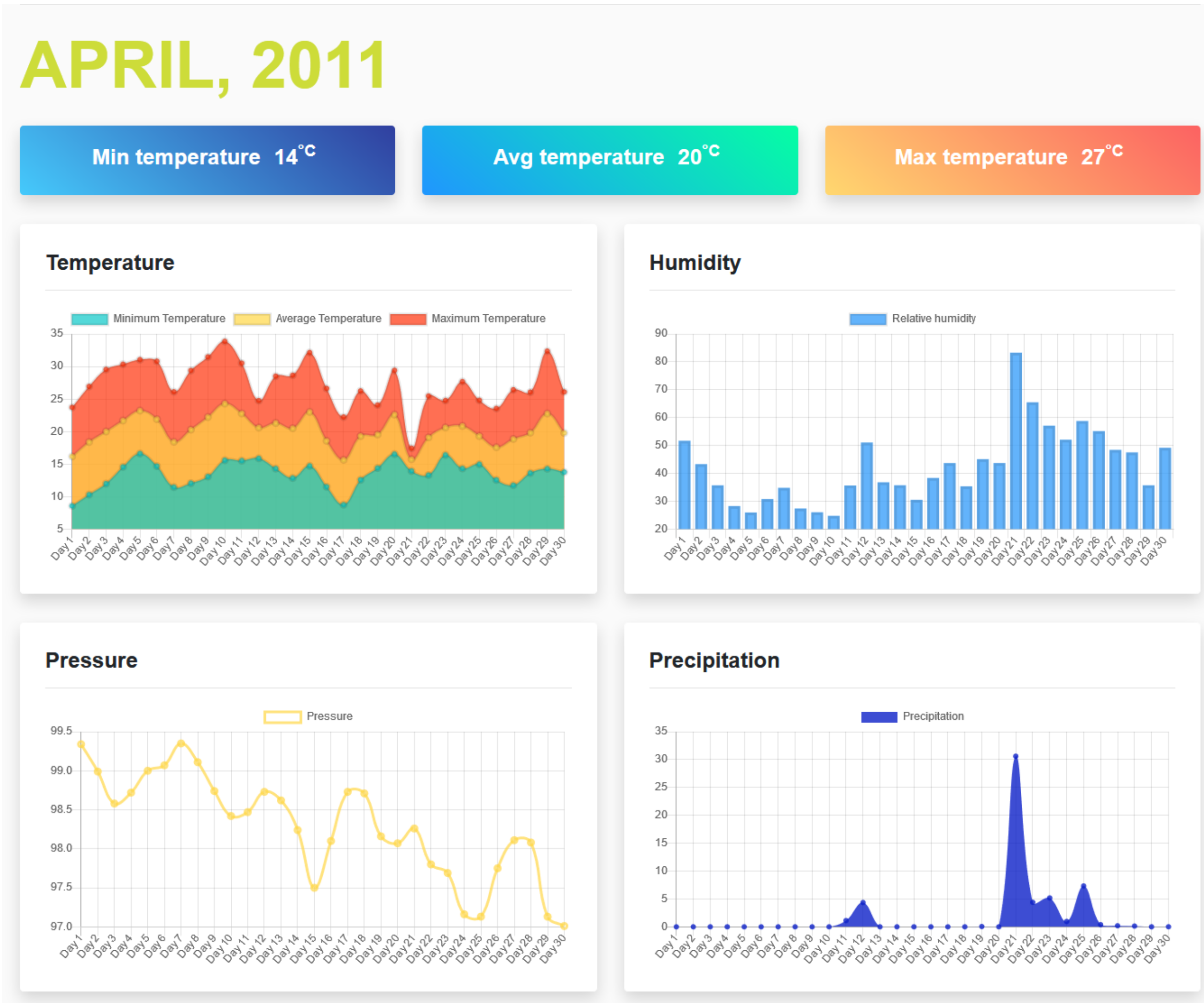

Figure 5.12: Historical Data

All previous records and values are stored in a table described in the following code snippet:

```
"""
    Database model for historical recorded from the raspberry pi
"""
class Monthly_records(models.Model):
    date = models.DateTimeField(db_column='Date', auto_now=False)
    # temperature
    min_temp = models.FloatField(db_column='min_temp')
    avg_temp = models.FloatField(db_column='avg_temp')
    max_temp = models.FloatField(db_column='max_temp')
    # precipitation
    min_prec = models.FloatField(db_column='min_prec')
```

```
        avg_prec = models.FloatField(db_column='avg_prec')
        max_prec = models.FloatField(db_column='max_prec')
        # humidity
        min_rel_humid = models.FloatField(db_column='min_rel_humid')
        avg_rel_humid = models.FloatField(db_column='avg_rel_humid')
        max_rel_humid = models.FloatField(db_column='max_rel_humid')
        # pressure
        min_pressure = models.FloatField(db_column='min_Pressure')
        avg_pressure = models.FloatField(db_column='avg_Pressure')
        max_pressure = models.FloatField(db_column='max_Pressure')

```

Listing 5.4: Historical data model

Land suitability prediction

The most important service is Land Suitability Prediction. Farmers can have a prediction about their land, whether it is suitable, needs more work (fertilization...etc) or completely unsuitable for growing the plants. The farmer needs to specify weather condition in order for the model to predict. The input data in sent to the server with POST method. As shown in the figure below.

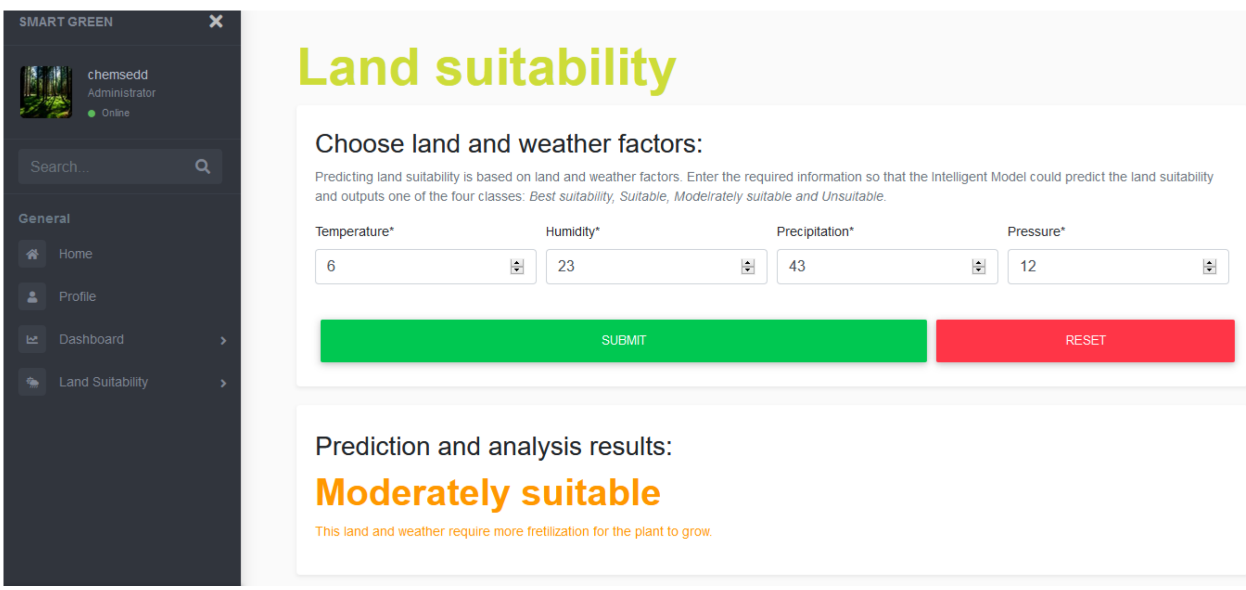

Figure 5.13: Land Suitability service

The following code is responsible for adapting the data to the model input shape, and make the prediction.

```
"""
    Prepare data before feeding it to the model
    - convert to numpy array
    - Reshape data to a 3D arrray
"""
def prepare_data(data):
    _data = np.array(data)
    return np.reshape(_data, (1, 4, 1))


"""
   Make a prediction about land suitability
   using the given data
"""
def make_prediction(model, data):
    _input = prepare_data([data])
    results = model.predict(_input)
    # select the class with highest probability
    results = np.argmax(results, axis=1)
    return results

```

Listing 5.5: Prepare input data and make predictionl

5.4 Obtained results and discussion

Input data is normalized in the range of 0 and 1 which is useful in Deep Learning for identifying patterns between features. Land suitability label (output) is coded in a 1-dimensional array of 4 columns, representing the affiliation to one of the 4 classes as shown in the table below.

Class	**Encoding**	**Output**
Unsuitable	[1, 0, 0, 0]	0
Moderately suitable	[0, 1, 0, 0]	1
Suitable	[0, 0, 1, 0]	2
Best suitability	[0, 0, 0, 1]	3

Table 5.2: Classes and output encoding

The training process of the model, was carried in google colab. Training went for 200 epochs with a batch size of 5. Our LSTM model was trying to learn the pattern between weather data and the land suitability label, in order to define and predict correctly whether the land will be suitable or not. We can see at the beginning of the training, accuracy was very low with a high loss value, as show in the figure below.

```
history = model.fit(x_train, y_train, epochs=200, batch_size=5, validation_data=(x_test, y_test))

Epoch 1/200
617/617 [==============================] - 2s 3ms/step - loss: 1.2145 - accuracy: 0.4473 - val_loss: 1.1429 - val_accuracy: 0.5755
Epoch 2/200
617/617 [==============================] - 1s 2ms/step - loss: 1.0398 - accuracy: 0.5774 - val_loss: 0.9719 - val_accuracy: 0.5881
Epoch 3/200
617/617 [==============================] - 1s 2ms/step - loss: 0.9127 - accuracy: 0.6089 - val_loss: 0.8084 - val_accuracy: 0.6115
Epoch 4/200
617/617 [==============================] - 1s 2ms/step - loss: 0.7055 - accuracy: 0.6933 - val_loss: 0.6198 - val_accuracy: 0.7702
Epoch 5/200
617/617 [==============================] - 1s 2ms/step - loss: 0.5666 - accuracy: 0.7725 - val_loss: 0.5301 - val_accuracy: 0.7790
```

Figure 5.14: Beginning of the training with LSTM model

After 200 epochs, at the end of the training, accuracy has a higher value (**0.97**), and a lower loss value (**0.06**) comparing to the beginning.

```
Epoch 195/200
617/617 [==============================] - 1s 2ms/step - loss: 0.0728 - accuracy: 0.9646 - val_loss: 0.0771 - val_accuracy: 0.9630
Epoch 196/200
617/617 [==============================] - 1s 2ms/step - loss: 0.0704 - accuracy: 0.9718 - val_loss: 0.0931 - val_accuracy: 0.9552
Epoch 197/200
617/617 [==============================] - 1s 2ms/step - loss: 0.0720 - accuracy: 0.9688 - val_loss: 0.0727 - val_accuracy: 0.9727
Epoch 198/200
617/617 [==============================] - 1s 2ms/step - loss: 0.0738 - accuracy: 0.9688 - val_loss: 0.1129 - val_accuracy: 0.9464
Epoch 199/200
617/617 [==============================] - 1s 2ms/step - loss: 0.0728 - accuracy: 0.9721 - val_loss: 0.0577 - val_accuracy: 0.9796
Epoch 200/200
617/617 [==============================] - 2s 2ms/step - loss: 0.0667 - accuracy: 0.9721 - val_loss: 0.1124 - val_accuracy: 0.9503
```

Figure 5.15: Ending of the training with LSTM model

The improved accuracy over time, from **0.44** to **0.97** means that the LSTM model was able to fit the data. Predicting correctly most of the time the land suitability, the validation accuracy is also considered very high, it ranges between **0.95** and **0.97**. The trained model uses a different set of data for validation that was not included in the training set, thus it was able to predict correctly for completely different set of data.

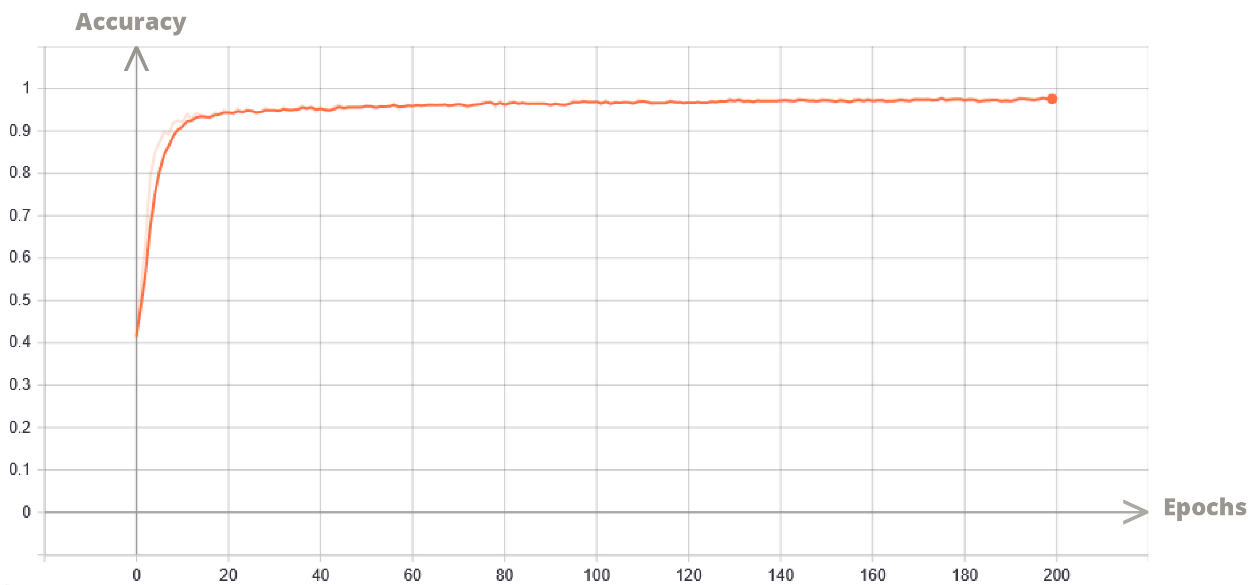

Figure 5.16: LSTM accuracy

For calculating the loss, we used *Categorical Cross Entropy* function. It's the most common function when it comes to classification problems. Categorical Cross Entropy increases as the predicted probability diverges from the actual label. The model works on minimizing the loss function, which distills all aspects of our algorithm into a single numerical value that describes how efficient our model is.

At the earlier stages of the training, loss function had a high value (**1.2**), LSTM tries to learn form the data and thus minimizing the loss by using the back propagation as training goes. After few epochs the value drops significantly to **0.06**, which indicates that the model succeeded at extracting features from the data set and identifying the pattern to correctly predict the land suitability.

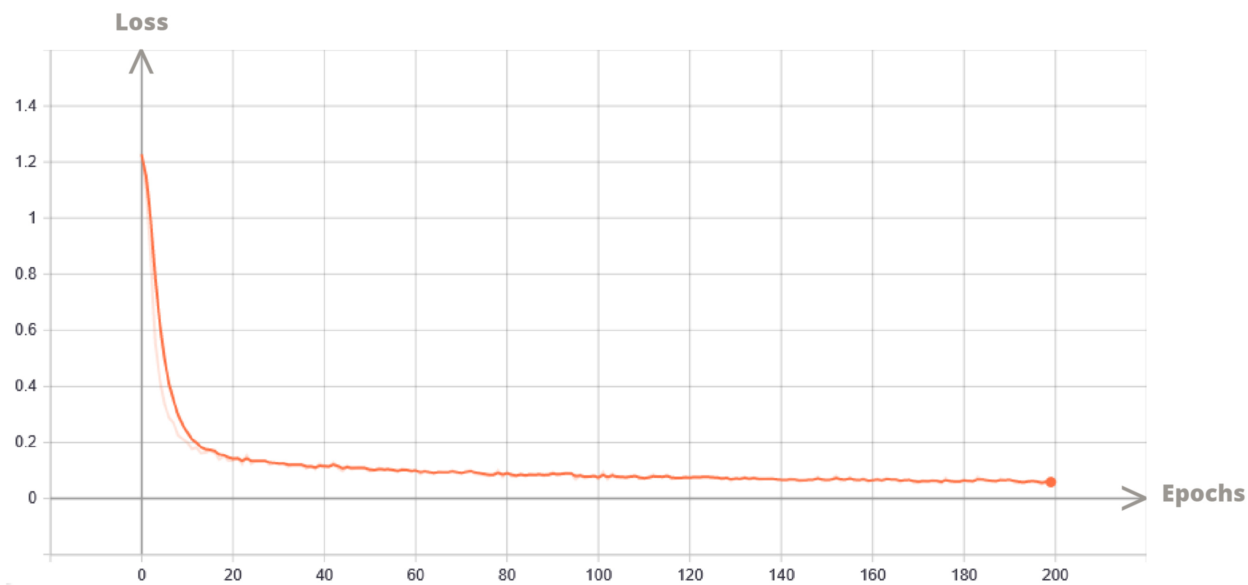

Figure 5.17: LSTM loss

Prediction results are stored on an array and plotted with the training labels as scattered dots on the same figure to demonstrate the accuracy of our model (Figure 5.18).

The X axis represents the testing set, which is composed of more than 1000 row (25% of all the data). Whereas the Y axis represents Land suitability classes, and the output is mapped to one of these classes (0, 1, 2 or 3) corresponding the to highest probability of affiliation.

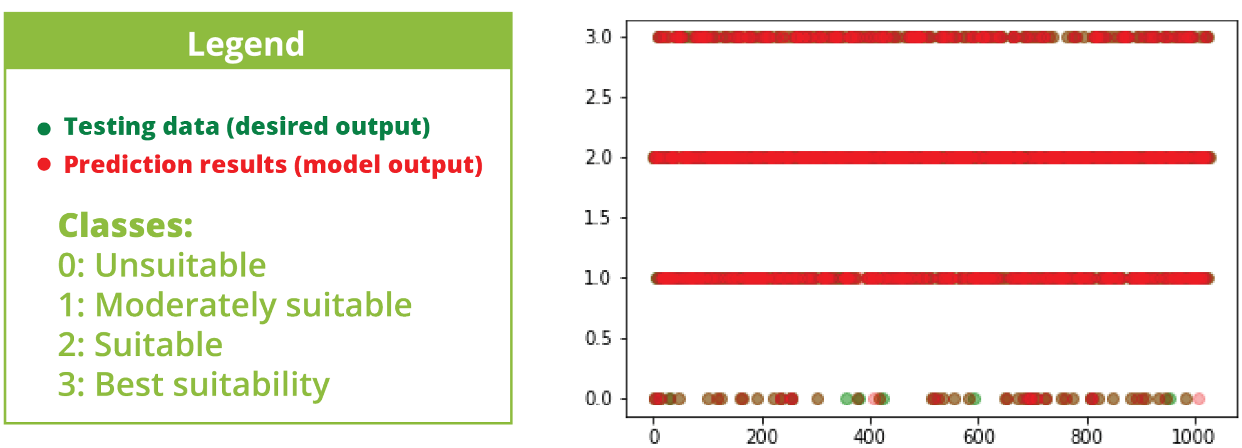

Figure 5.18: LSTM prediction results

We can see that the predicted results and the desired output are almost stacked upon each other, which indicates that our model was able to predict the output nearly identical to the testing data. At some points the model predicted wrong results and that is illustrated on the dots plotted in different places, and that can be beneficial for avoiding over fitting.

5.5 Conclusion

In this chapter we presented the implementation of our system as well as the obtained results. We carefully chose the best combination between hardware and software tools in order to achieve our goal. The results were promising, our model had a high accuracy in prediction, which reflects the well structured system and architecture.

Chapter 6

Conclusion and Perspectives

6.1 Conclusion

Land suitability presents a very important side of agriculture and farming, our work mainly focuses on providing the best prediction on that matter. The proposed architecture which is a combination of Deep Learning and Internet of Things showed efficiency in predicting land suitability based on the collected data about weather and soil. Compared to the related works on smart farming and land suitability, our solution also focused on data streaming method. Collecting data the best way will surely improve results. Our solution also uses Long-Short Term Memory (LSTM) model for making the prediction with time series data instead of classic Artificial Neural Networks, as patterns emerge in time series rather than considering the data just scattered values. However, our solution suffered from shortage in data. The historical data was only for the past 9 years, which is considered relatively low at the count of 1 value per day.

6.2 Perspectives

Our work can be extensible and enhanced for better improving the farming field. As a future work, we are willing to introduce more features to our system, such as cameras for monitoring plants health. Images will be captured on a daily basis and stored on the server databases. All the stored images will then be fed to a Convolutional Neural Networks (CNN) for image processing and extracting useful information about plants health. Enriching the dataset will surely result in higher accuracy in prediction by adding more sensors on the field (UltraViolet...).

From another perspective, we wish to embed GPS trackers in all farming vehicles in order to be able to locate each one on the field from the app on the server. Providing farmers with full control over their fields will surely enhance crop yields and ease the task even more.

References

What is a raspberry pi? https://www.raspberrypi.org/help/what-is-a-raspberry-pi/, Aug 2015.

Python executive summary. https://www.python.org/doc/essays/blurb/, 2020.

Mohammad Al-Smadi, Omar Qawasmeh, Mahmoud Al-Ayyoub, Yaser Jararweh, and Brij Gupta. Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of arabic hotels' reviews. *Journal of Computational Science*, 27:386 – 393, 2018. ISSN 1877-7503. doi: https://doi.org/10.1016/j.jocs.2017.11.006.

John B Bradford, Daniel R Schlaepfer, William K Lauenroth, Charles B Yackulic, Michael Duniway, Sonia Hall, Gensuo Jia, Khishigbayar Jamiyansharav, Seth M Munson, Scott D Wilson, et al. Future soil moisture and temperature extremes imply expanding suitability for rainfed agriculture in temperate drylands. *Scientific reports*, 7(1):1–11, 2017.

Min Chen, Shiwen Mao, and Yunhao Liu. Big data: A survey. *Mobile networks and applications*, 19(2):171–209, 2014.

M Chi, J Shen, Z Sun, F Chen, and JA Benediktsson. Oil spill detection based on web crawling images. In *Proc. IEEE Int. Geosci. Remote Sensing Symp.*, pages 1–4, 2014.

Mingmin Chi, Antonio Plaza, Jon Atli Benediktsson, Zhongyi Sun, Jinsheng Shen, and Yangyong Zhu. Big data for remote sensing: Challenges and opportunities. *Proceedings of the IEEE*, 104 (11):2207–2219, 2016.

Cam Cullen (VP of Global Marketing). Sandvine releases 2019 global internet phenomena report. https://www.sandvine.com/press-releases/sandvine-releases-2019-global-internet-phenomena-report, 2020.

Laetitia De Goeij and Sandra Soeters. Importance of humidity for plant growth. http://www.anthura.nl, 2016.

International Fund for Agricultural Deve, World Food Programme, and F.A.O. of. *The State of Food Insecurity in the World 2015: Meeting the 2015 International Hunger Targets: Taking Stock of Uneven Progress*. FAO, 2015. ISBN 9789251088005.

Ravi Gorli. Future of smart farming with internet of things. *Journal of Agriculture and Water Works Engineering*, 1(1), 2017.

Ibrahim Abaker Targio Hashem, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, and Samee Ullah Khan. The rise of "big data" on cloud computing: Review and open research issues. *Information systems*, 47:98–115, 2015.

T. Hoeren and B. Kolany-Raiser. *Big Data in Context: Legal, Social and Technological Insights*. SpringerBriefs in Law. Springer International Publishing, 2017. ISBN 9783319624617.

Mohammad Jabraeil Jamali, Bahareh Bahrami, Arash Heidari, Parisa Allahverdizadeh, and Farhad Norouzi. *Towards the Internet of Things: Architectures, Security, and Applications*. 01 2020. ISBN 978-3-030-18468-1.

Andreas Kamilaris, Feng Gao, Francesc X Prenafeta-Boldu, and Muhammad Intizar Ali. Agri-iot: A semantic framework for internet of things-enabled smart farming applications. In *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*, pages 442–447. IEEE, 2016.

Andreas Kamilaris, Andreas Kartakoullis, and Francesc X Prenafeta-Boldú. A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture*, 143:23–37, 2017.

Ravi Kishore Kodali, Vishal Jain, and Sumit Karagwal. Iot based smart greenhouse. In *2016 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, pages 1–6. IEEE, 2016.

Chandra Krintz, R Wolski, N Golubovic, B Lampel, V Kulkarni, B Sethuramasamyraja, B Roberts, and B Liu. Smartfarm: Improving agriculture sustainability using modern information technology. In *KDD Workshop on Data Science for Food, Energy, and Water*, 2016.

Jitendra Kumar, Rimsha Goomer, and Ashutosh Kumar Singh. Long short term memory recurrent neural network (lstm-rnn) based workload forecasting model for cloud datacenters. *Procedia Computer Science*, 125:676–682, 2018.

James Manyika, Michael Chui, Jacques Bughin, Richard Dobbs, Peter Bisson, and Alex Marrs. *Disruptive technologies: Advances that will transform life, business, and the global economy*, volume 180. McKinsey Global Institute San Francisco, CA, 2013.

Abdelhak Merizig, Hamza Saouli, Meftah Zouai, and Okba Kazar. An intelligent approach for enhancing the agricultural production in arid areas using iot technology. In *International Conference on Advanced Intelligent Systems for Sustainable Development*, pages 22–36. Springer, 2018.

Mehdi Mohammadi, Ala Al-Fuqaha, Sameh Sorour, and Mohsen Guizani. Deep learning for iot big data and streaming analytics: A survey. *IEEE Communications Surveys & Tutorials*, 20(4): 2923–2960, 2018.

Jirapond Muangprathub, Nathaphon Boonnam, Siriwan Kajornkasirat, Narongsak Lekbangpong, Apirat Wanichsombat, and Pichetwut Nillaor. Iot and agriculture data analysis for smart farm. *Computers and electronics in agriculture*, 156:467–474, 2019.

Keyur K Patel, Sunil M Patel, et al. Internet of things-iot: definition, characteristics, architecture, enabling technologies, application & future challenges. *International journal of engineering science and computing*, 6(5), 2016.

Kantamneni Raviteja and M Supriya. Iot-based agriculture monitoring system. In *Data Engineering and Communication Technology*, pages 473–483. Springer, 2020.

Ammar Rayes and Samer Salam. Internet of things from hype to reality. *The Road to Digitization; River Publisher Series in Communications; Springer: Basel, Switzerland*, 49, 2017.

Satish Shankar, Francois van Raemdonck, and Dalton Maine. Can agribusiness reinvent itself to capture the future? https://www.bain.com/insights/can-agribusiness-reinvent-itself-to-capture-the-future, 2016.

Andrey Somov, Dmitry Shadrin, Ilia Fastovets, Artyom Nikitin, Sergey Matveev, Oleksii Hrinchuk, et al. Pervasive agriculture: Iot-enabled greenhouse for plant growth control. *IEEE Pervasive Computing*, 17(4):65–75, 2018.

Jesús Martín Talavera, Luis Eduardo Tobón, Jairo Alejandro Gómez, María Alejandra Culman, Juan Manuel Aranda, Diana Teresa Parra, Luis Alfredo Quiroz, Adolfo Hoyos, and Luis Ernesto Garreta. Review of iot applications in agro-industrial and environmental fields. *Computers and Electronics in Agriculture*, 142:283–297, 2017.

Durai Raj Vincent, N Deepa, Dhivya Elavarasan, Kathiravan Srinivasan, Sajjad Hussain Chauhdary, and Celestine Iwendi. Sensors driven ai-based agriculture recommendation model for assessing land suitability. *Sensors*, 19(17):3667, 2019.

Tim Wheeler and Joachim Von Braun. Climate change impacts on global food security. *Science*, 341(6145):508–513, 2013.

Sjaak Wolfert, Lan Ge, Cor Verdouw, and Marc-Jeroen Bogaardt. Big data in smart farming–a review. *Agricultural Systems*, 153:69–80, 2017.

Lewis Ziska, Allison Crimmins, Allan Auclair, Stacy DeGrasse, JF Garofalo, AS Khan, Irakli Loladze, AA Pérez de León, A Showler, J Thurston, et al. Ch. 7: Food safety, nutrition, and distribution. *The impacts of climate change on human health in the United States: A scientific assessment*, pages 189–216, 2016.

Meftah Zouai, Okba Kazar, Guadalupe Ortiz Bellot, Belgacem Haba, Nadia Kabachi, and M Krishnamurhty. Ambiance intelligence approach using iot and multi-agent system. *International Journal of Distributed Systems and Technologies (IJDST)*, 10(1):37–55, 2019.