International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. , No. (Mon-20..)

http://dx.doi.org/10.12785/ijcds/XXXXXX

IoT and Deep learning-based approach for an efficient land suitability prediction in smart farms

Toufik Bendahmane[1], Abdelhak Merizig[1], Khaled Rezeg[1], Okba Kazar[1,2] and Saad Harous[3]

[1]*LINFI Laboratory, Computer Sciences Departement, Mohamed Khider University of Biskra, Biskra, ALGERIA*
[2]*College of Computing and Informatics, Department of Computer Science, University of Sharjah, College of Arts, Sciences Information Technology, University of Kalba, Sharjah, United Arab Emirates*
[3]*College of Computing and Informatics, Department of Computer Science University of Sharjah, UAE*

E-mail address : toufik.bendahmane@univ-biskra.dz, a.merizig@univ-biskra.dz, k.rezeg@univ-biskra.dz, o.kazar@uaeu.ac.ae., harous@sharjah.ac.ae.

Received Mon. 20, Revised Mon. 20, Accepted Mon. 20, Published Mon. 20

Abstract: The earliest known way for humans to make a living is through farming. Smart farming is a new, improved vision of agriculture that incorporates the use of new technologies. In recent times, farmers have increasingly depended on technology to efficiently carry out their daily responsibilities and enhance the quality of their crops. In agriculture, land suitability is an important aspect, which describes how well area is conducive for plant growth. Experts in the field of land suitability can determine it or use mathematical tools to make accurate predictions. Artificial techniques have been proven to be efficient prediction tools for this purpose. Empowered by the Internet of Things and Big Data, Artificial Intelligence (AI) is capable of handling these kinds of tasks and easing the burden on farmers and experts. The devices, used to improve farming, generate data in several formats, which might lead to ambiguous data. This paper proposes an ontology-based solution to deal with the heterogeneity problem. Moreover, this paper uses a deep learning-based solution that uses streamed weather data generated from sensors. In fact, our system uses the long-short-term memory model to predict land suitability. The model exhibited encouraging outcomes that could potentially influence the field of agriculture.

Keywords: Agriculture 4.0, Land Suitability, Internet of Things, Machine learning, Deep Learning, Long-Short Term Memory, Ontology.

1. INTRODUCTION

Agriculture has served as the primary sustenance for both people and animals throughout the past 12,000 years. Due of its significance, authority and private entities are always trying to improve the return from farming the land, which leads to increasing the incoming resources and meet the population requirements. The latest technological development, if used properly, can enhance the farming field in a substantial way. In fact, the application of new technological developments in traditional agriculture resulted in a new term called smart agriculture [1], [2]. Smart agriculture refers to the integration of Internet of Things (IoT) technologies with conventional agricultural practices [3]. This combination will greatly facilitate field monitoring, particularly in the production chain process. In the conventional paradigm of agriculture, farmers are responsible for nearly all aspects of the process, including aggregation and crop collection. In addition, experts face limitations when they want to study different phenomena and/or diseases that affect crops. The emerging concept of agriculture revolves around the implementation of sensors on the field to collect data about several characteristics of the environment, such as humidity, temperature, wind speed and so on [4]. The collected data can be used by experts and data analysts for decision-making process purposes [5], [6]. In addition, it could be to study different phenomena related to the seeds or the fertility of the land. Agricultural land suitability prediction is a significant concern that producers prioritize greatly [7]. To optimize crop quality, farmers must possess precise knowledge regarding the optimal timing and location for seed planting, taking into account several elements and criteria, including water availability, weather conditions, soil composition, and fertilizer. Experts need a huge amount of historical data to make the prediction as accurate as possible. The new developments in technology can be very helpful in this domain. Using sensors and actuators [8] [9], a huge amount of historical data can be collected, and then, using AI techniques to study this data, accurate information about the land and the crops is deduced. In this work, we aim to design and implement a system that uses AI techniques and IoT technology to collect and analyze data that helps farmers make appropriate predictions about land suitability. Usually, new farmers do not have sufficient understanding of the soil properties required for agricultural

E-mail address: toufik.bendahmane@univ-biskra.dz

production [10]. They lack awareness of the necessity of doing an assessment of agricultural land prior to cultivation. Consequently, it is essential to assess the suitability of the area before cultivating crops in order to optimize production. [11], [12]. Traditionally, farmers rely on manual data collection and soil testing labs to know the soil's properties. These methods may not lead to the aimed goal because the data may not be accurate [13], [14]. The elements of land can determine whether it is suitable for agriculture and plantations. Nevertheless, terrestrial components are excessively utilized and exploited. Various regions are encountering diverse issues, including soil erosion, water logging, groundwater depletion, excessive run-off, and productivity losses. The deterioration of land poses a significant danger to food and energy security, water availability and quality, biodiversity, and human life. Weather conditions can significantly affect agricultural lands. This phenomenon is the result of a modification in the atmospheric conditions that impacts the ability of agricultural operations to flourish and enhance crop production. Several factors that influence the conditions for crops and cattle include temperature, nutrient concentrations, soil moisture, and water accessibility. Research on crop yields has shown that very high temperatures, resulting in elevated vapour pressure deficits, can reduce the yields of rain-fed crops across different crops and areas. Interconnected devices or physical objects via the internet, known as IoT technology, represent the new evolution in artificial Intelligence [15]. This approach uses different kind of communications. Data collected, this way, may be of different types. This problem is due to the nature of data gathered by devices that are generally heterogeneous [16]. In fact, these data are collected in real time which will generate a sort of Big Data [17]. This kind of data is usually handled using data streaming process [18], [19]. Smart agriculture brings a new vision which injects the use of AI technology in the traditional agriculture. This combination increases the production and economic profits [20]. The emergence of sensors, which are based on different systems, leads to some problems about how is this data exchanged among these systems. Usually, to avoid such this issue which is related to understanding heterogeneous data or exchanges signals between sensors known as the interoperability issue. In fact, land suitability represents an issue to the farmers in two manners the first consists of the quality of food or plant the second is related to the nature of the field area. The primary contributions of this study are:

- Develop a distributed architecture to ensure interoperability inter among devices.
- Propose an ontology domain to ensure data understanding between different generated data;
- Propose a deep learning model for land suitability prediction;
- Develop a distributed architecture to ensure interoperability among devices;
- Propose a framework to visualize data in real time;
- Perform several comparisons of our strategy with alternative machine learning classifiers.

The paper is organized as follows: section 2 describes related works about smart agriculture including land suitability. In section 3 we present our proposed framework. Section 4 explains the prediction model for land suitability. The obtained results are discussed in section 5. Finally, section 6 concludes this paper and gives perspectives for future works.

2. Literature Review

Since the emergence of sensors, actuators and the idea of the IoT many developers/researchers have incorporated these technologies in many domains such as industry or agriculture. In the literature, many works have used the internet of things as solution in different field. In this section, we present some works that have used the IoT as solution to problems in agriculture field. Agriculture field has gained many advantages from this technology such as collecting data, irrigation systems, monitoring the field and detecting the disease on the corps. In addition, land suitability is a major challenge due to its importance to the countries especially the ones that have large areas. In the work of Bu and Wang [21] presented an IoT based system for smart agriculture on edge-cloud computing to enhance food production. The authors used deep reinforcement learning that is based on four principal layers: the data collection layer, the edge computing layer, the data transmission layer, and the cloud computing layer. The implemented model is used to predict the needed amount of water in order to enhance crop growth. However, the author did not take into account the land suitability, which could affect the corps growth. When Hsu et al. in [22] have designed a platform that uses cloud fog computing for agriculture purposes. The main idea is to present a new creative IoT system layer composed of a set of layers instead of the traditional IoT layers. In fact, the authors integrated the fog computing layer into this system. According to the authors, this architecture ensures data gathering and analysis from different types of devices. However, they did not mention how to deal with the heterogeneity of collecting data from these devices. Pathak et al. [23] proposed a field monitoring system that is based on IoT technology to deal with the problem of crop's irrigation. The main objective of the developed system is to optimise the quantity of water used for the corps' irrigation. The main idea is to use the collected climatic parameters to predict the level of water in the field and thus decide the amount of water that should be poured to irrigate the crops. Also, Dos Santos et al. in [24] have proposed a model to measure crop productivity and anticipate problems. The proposed model, named AgriPrediction, is based on both the ARIMA model and LoRa IoT technology. The authors combined a set of technologies, such as wireless network range systems, with a prediction method to notify the farmers for some possible recommendation. The system uses different climate factors (soil humidity and temperature) to decide what actions should be taken. However, the authors did not take into consideration all climate factors that could affect crop productivity, which is a shortcoming of this work. Chen et al. [25] proposed a platform called agriTalk, which is an IoT-based platform for the precision farming of soil cultivation. The authors ensure connections between the sensors and actuators in order to preserve farming precision. The objective of their solution is to increase the number of crop cultivation resources through turmeric cultivation. They used network time protocol in their platform. However, interoperability between devices is not mentioned in this work. In [26] Senapaty et al. have applied IoT solution-based machine learning to increase crop productivity. The solution consists of analysing soil nutrients to enhance precision in agriculture. In addition, the proposal includes a crop recommendation model. The machine learning model is represented as a combination of a support vector machine with a directed acyclic graph and fruit fly optimisation technique. Where Shevchenko et al. in [27] have proposed a solution to deal with the impact of climate changes on land suitability. The solution aimed to deal with risks that impact food security. The machine model is used to deal with the problem under different carbon emission scenarios. However, interoperability between devices is not mentioned these works.

3. Proposed methodology

Agriculture field is evolving with the evolution of new technologies supported by Artificial Intelligence. There are many problems related to traditional agriculture that limit the farmers' profits. One of these significant problems is land suitability, which is one of the most known issues in traditional agriculture. The study of suitability of a given land helps the farmer to decide which kind of crops should be implanted and what should be added to this land to support other crops. Through data analysis and studying historical data related to weather condition, including soil conditions, this issue should be covered. Similar to other fields, agriculture field may benefit from different ideas using new development in technology, such as sensors that collect data about the environment. The main use of the new development in technologies is to reduce humans' effort in different fields, and this is especially true in agriculture domain. Farmers nowadays use technology to complete their daily tasks in the field to improve crops quality. To tackle these kinds of problems, solutions based on multi-layer architecture have been proposed. These solutions analyze collected data, visualize these data in an easy-to-understand to for farmers, and predict which crops are more suitable to be implanted in this land. These solutions combine the use of IoT and data analytics to study a new phenomenon. In this section, we propose an architecture that consists of a set of layers. Every layer has a distinct responsibility. The first layer installed in the field, collects data from the physical area using sensors and actuators. The second layer consists of data analysis which covers the pre-processing operations. The last layer is a Cloud IoT approach, this layer represent the process of model construction and stores the data on cloud servers. Besides, our proposal it ensures data analytics in real-time over historical one through data visualization displayed as data insights for the users.

A. *Architecture description*

In this subsection, we present the constituents employed in our suggested approach and their respective functions. Our architecture is designed to address the issue of land suitability problem through an interoperability protocol. It is a multi-layer architecture based on IoT to predict the land suitability and visualize the data in real-time. In addition, since we are using different farm settings, including number of sensors and actuators which it might produce various kinds of data. To deal with issue, our solution is to use an ontology to ensure the interoperability protocol exchanged inter-systems. Figure 1 shows the proposed architecture composed of the following set of components:

1) **End User:** This component is the interface to our system. The user can access all operations available in the system by clicking on the desired operation. The system supports all types of devices such as pads, phones, etc... In fact, this component aims to interact the users with the Cloud services, which in our case data analysis, data visualization and more importantly, the land suitability prediction. To perform a suitability detection for a given land, the user should introduce some parameters related to a blessed land. These parameters are weather condition and data of the land such as PH and nature of it to check if the given area suitable for a specific crop.
2) **Sensing layer:** called Physical field layer, its primary role is gathering data from the field and sending it to the sink. This layer is made up of : Temperature sensor, Humidity sensor, Soil moisture sensor, UV sensor and PH sensor. All sensors are controlled by the sink which is modeled by Raspberry Pi in our case. The primary role of the sink is to transmit the collected data to the IoT Edge layer. The gathered data can be scheduled weekly, daily, or hourly according to the necessity of data and the studied area. We this configuration, the farm will be monitored in each lap of time, which will generate vast amounts of data that will be treated as big data. The resulted data will be passed to the next level to the analysis purposes.
3) **IoT Edge layer:** This component is an IoT edge layer. We can call this layer a middleware, in our architecture, which connects the processing layer (cloud IoT layer) and the physical layer (sensing layer). This layer also ensures the interoperability between different types of data collected from the field. In this layer; we use an ontology domain to unify the data collected from the field. A MQTT

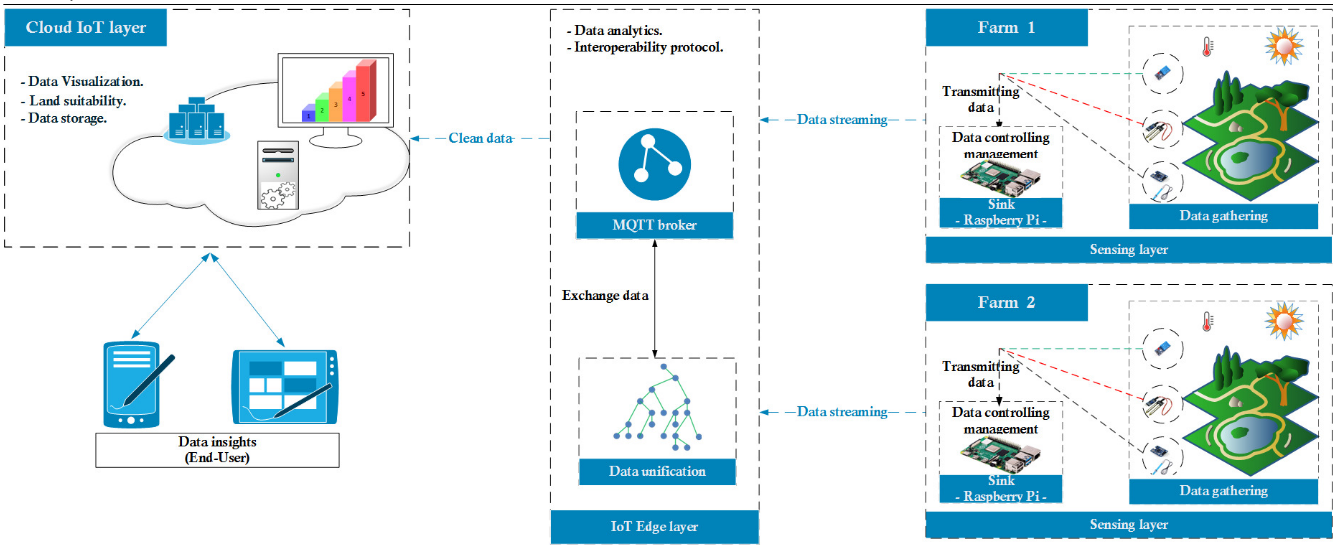


Figure 1. Land suitability monitoring and prediction system.

broker is used to facilitate the interaction between the physical layer and this layer through the sink. Actually, the MQTT broker is used to ensure the exchanged messages between devices as publish and subscribe form [28]. It contains the following modules:

- **Interoperability protocol:** Before we start the data analysis operation, we need to unify the different types of data. After receiving the collected data from each farm sites as data streaming transmitted into the MQTT broker. Then we perform the interoperability, we use an ontology domain attributed in this layer which will be presented in the next section.
- **Collected data analysis (preprocessing):** After collecting data from different areas, this module occupies the pre-treatment process which covers data cleaning. This operation it is data streaming based method which will collect data in real-time. After this operation, the extracted meaningful patterns from the collected data will be transformed into the Cloud IoT layer for the model creation.

4) **Cloud IoT layer:** This component is a cloud-based layer. It extends the local server abilities by using the Cloud server's resources. Furthermore, this layer gains from cloud advantages such as resources pooling and service on demand [29], [30]. This layer also offers a platform with the following multi functionalities accessible via Internet.

- **Cloud storage:** The different sensors collected huge amount of data (Big Data) [31], which requires a lot of storage space. The cloud provides the user with unlimited data storage space compared to the local server.
- **Data visualization:** It allows the farmers to visualize data in real-time as it is collected from the field. The data visualization makes data more understandable. The data will then be sent in an unstructured data format such as JSON and XML for later use.
- **Model construction or Machine learning module:** The cloud offers high computing resources, which can run very complex algorithms to predict the land suitability. In our work, we use a deep learning method which is a neural network-based method. In our proposal, we use a machine learning algorithm that handles the land suitability prediction.

B. *Dataset description*

To construct our model to predict the suitability of the land based on remote sensing data, we utilized meteorological data collected from different regions over twenty years. This dataset consists of various features such as soil moisture, precipitation, temperature, relative humidity, and Pressure. We focused on using deep learning techniques to predict land suitability across a heterogeneous set of land covers, including natural vegetation, croplands, and human infrastructure.

C. *Land Suitability prediction process*

To explain the role of each element in the architecture depicted in previous figure (see figure 1), we use the following flowchart (see figure 2). As described in figure 2, our system consists of settings, especially at the physical layer (sensing layer) installed in each farm. To ensure that a given area could produce healthy plants, our proposal is composed of two different main sub-systems. Our system provides the land suitability prediction in a specific area. To deal with the interoperability problem between other data generated from sub-systems (different farms), we present

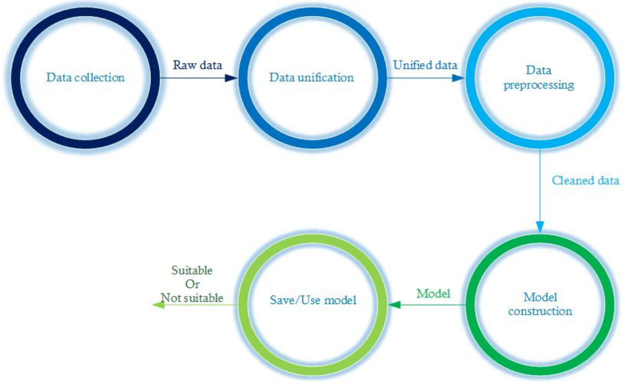


Figure 2. Land suitability process.

the previous flowchart's various operations. The issue related to interoperability due to the nature of generated data could be presented in other formats (JSON and CSV in our case or maybe include XML format).

1) **Data collection:** this step is the important to construct the data especially data streams. The system receives a numerical data from our farm where the output (raw data) is stored as unstructured data (JSON file). At the end the local server saves these data using streaming framework such as KAFKA [32], [33].
2) **Data unification:** since we have many types of data generated from each farm field, this is present big challenge related to understanding each type. To ensure an excellent manner to process these data, this step introduces a protocol to unify them. The primary objective of this operation is to prepare data for pre-processing stage. In fact, MQTT broker handles the unification process, what we also called the interoperability process. The main idea is to use the collected data from the field. In IoT, interoperability can be defined as two systems that can communicate and share information or data services via devices [34]. The devices generate different kinds of data with several data formats, including XML, JSON, or even CSV (Noura et al., 2019). As we have seen in our proposed architecture, our system may generate data from devices as the previous data format. According to [15] semantic interoperability in IoT can be ensured through different ideas which given in three ways; information model, data model, or using ontologies. The existence of different kind of sensors can generate different data types. To deal with the mention problem, semantic interoperability is needed and gathered data from devices needs the use of sensor ontologies [35]. Since the ontology well known in the semantic field through many works in literature our interest is to use them [36], [37]. Our idea is to use domain ontology to unify the collected data to start the data analytics task.

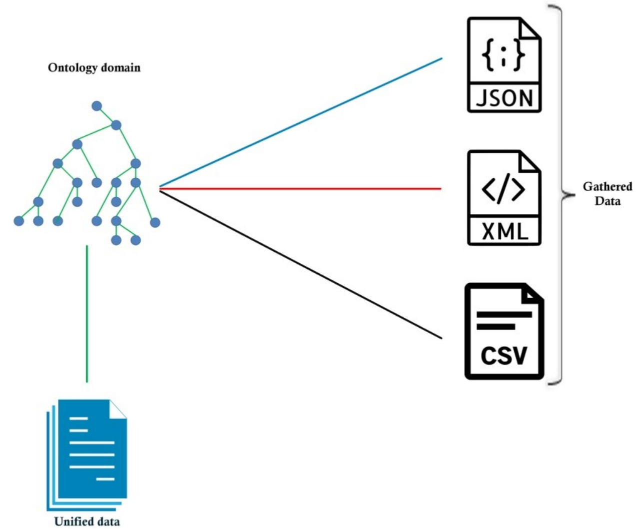


Figure 3. Data unification model.

To extract information from each file using the ontology and to find the right concept from one data format to another, we execute the equation presented in equation 1 [38]. The primary role to this operation is to calculate the similarity degree from concept to concept, and then we create a unified file at the end. In this work, we use XML file type as unified data file. The idea behind this decision is due to the tools and library that make file generation easier than other types.

$$W(term) = tf{-}idf(term) = tf(term){-}log_2\left(\frac{N}{df(term)}\right) \tag{1}$$

W represents the term weight, tf is the term frequency, where N is the number of documents in the collection, and df is the document frequency, which is the number of times the word appears in the other documents. After unified data construction, the MQTT brokers send the result to start the data preprocessing. The following figure represents the ontology used to unify concepts between different data format. The presented ontology inspired from [39].
To fetch a data from the used ontology given above (see figure 4), we propose the following algorithm to extract the pertinent data. After extracting the classname we compare them (data format: csv, xml, and JSON files) to unify the data. After that we send the data for preprocessing step in order to create our prediction model. Our algorithm consists of several steps, given as follows:

- Step 1. Given the data format received from the sink, which could be one of the three different formats as discussed earlier,. This data format is presented as a query to our algorithm, and

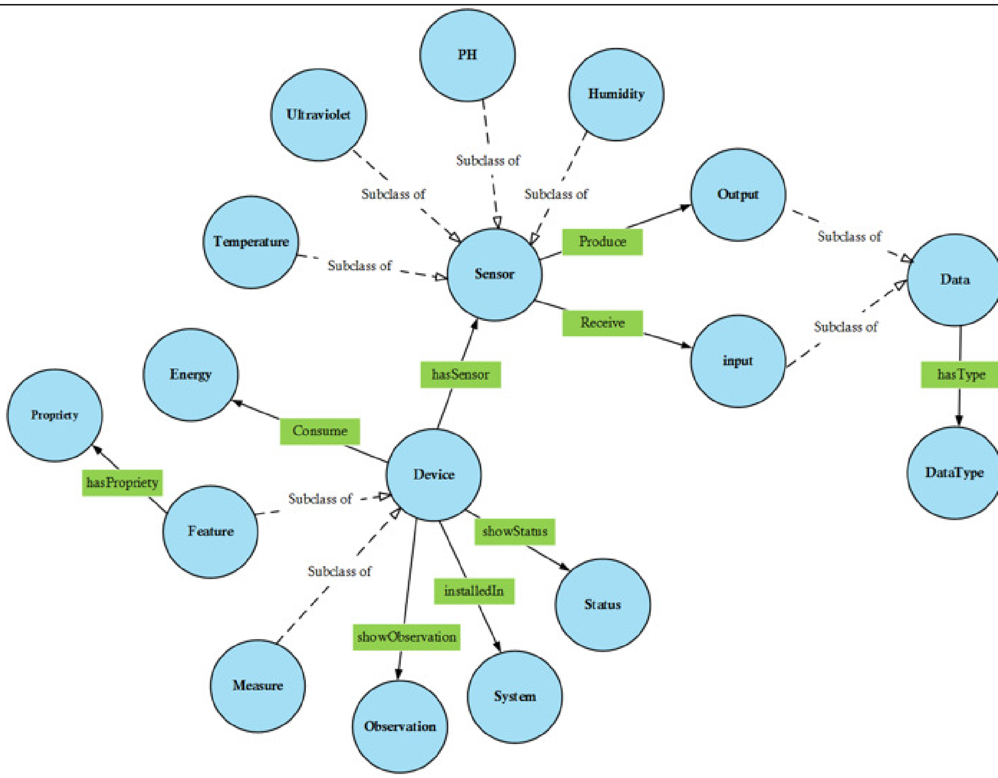


Figure 4. Proposed Domain ontology.

Algorithm 1 Fetch in ontology.

Input: Q: query : data from source file, O: ontology domain.
Output: C: a list of concepts and its signification.
Begin
Let o *class_node* of data from O;
Let q set of *name_class* of Q;
for each name n in q **do** ▷ extract class name from ontology.
 $i \leftarrow 0$;
 Let S an array of real values
 for each class_node o in O **do** ▷ calculate degree of similarity between concepts. Tf-idf
  $S[i] \leftarrow similarity(o, n)$;
  $i \leftarrow i + 1$;
 end for
 $classes \leftarrow ExtractClassName(S)$; ▷ Extract class name with the highest value.
 $C \leftarrow addtolist(classes, n)$;
end for
Return C;
End.

we also give the used ontology.

- Step 2. For each name 'n' from class name 'q' in query Q we fetch for the appropriate concept that is near to 'n'.
- Step 3. We calculate the degree of similarity of each name 'n' and compare it with each class node of ontology O and we store it in an array S.
- Step 4. Last step we extract the class node that has the highest degree of similarity and save in list named 'C' with the current class name 'n'. when we finish this algorithm, we return each concept with its meaning using the ontology resulted in list 'C'.

3) **Data pre-processing:** Accumulating vast quantities of minuscule numerical values will always result in incomplete datasets. If the raw data contains missing values and is immediately inputted into the model, it will lead to an error. Hence, a preliminary stage of data pre-processing is required to ensure data cleanliness. To improve accuracy and efficiency, it is essential to standardise the data to a range of 0 to 1. The result will be a tidy and reliable dataset that is prepared for usage with the model.
4) **Model construction:** The selection of the optimal model and architecture significantly influences the accuracy of predictions. From the several Machine/Deep learning methods and algorithms available, we have selected an approach that takes into account time series data. This phase consists of two steps:
   - Designing the model and its architecture, in-

cluding determining the number of layers, number of neurons, and activation functions.
- Performing model training and testing by producing predictions to evaluate the model's performance. The algorithm that was utilized will be examined in the "Prediction algorithm" section.

5) **Save/Use model:** Once the model is created, it needs to be saved and uploaded to the web platform in order to provide users with access to the Land suitability service. Following the completion of the training and testing processes using the datasets, the model should now possess the capability to determine the appropriateness of the land for a particular plant. The resulting output will fall into one of four categories: "Best suitability," "Suitable," "Moderately suitable," or "Unsuitable". In the use step, the system generates data visualization the inspected area using choropleth to display the results.

4. Prediction model for land suitability

LSTM Networks are a specific type of Recurrent Neural Network (RNN). Their purpose was to tackle the problem of prolonged reliance in RNN. LSTMs have exceptional ability in preserving information over prolonged durations. Due to the potential impact of prior information on model correctness, LSTMs are often used for this purpose [40]. The LSTM architecture consists of a module known as the "Repeating Module" which comprises four neural network layers that interact in a distinctive manner. The repeating module is equipped with three gate activation functions: σ_1, σ_2, σ_3 and two output activation functions ϕ_1 and ϕ_2 as shown in Figure 5. The selection of the LSTM architecture

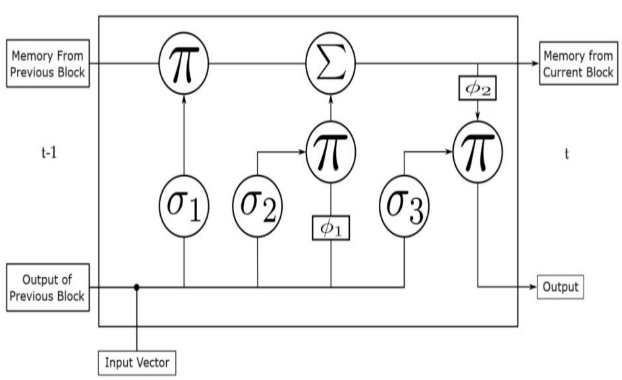


Figure 5. LSTM Repeating Module.

was primarily determined by the characteristics of the data at hand. Given that we are working with time series, which refers to data that is associated with time and has been recorded in a sequential manner. LSTM has demonstrated superior performance in situations when it is capable of retaining and recalling information from past data points. Our circumstance can benefit from this approach as the prediction of land suitability is not possible using dispersed or unsorted data. To create our prediction model as we mentioned before we use the LSTM method. Our algorithm follows the next steps:

1) *Data collection:* This procedure is carried out, as previously described, using Internet of Things (IoT) devices such as sensors and Raspberry Pi.
2) *Data preparation:* involves several steps before inputting it into the model. These steps include filling in missing values, cleaning the data, normalizing it, and reshaping it into a 3-dimensional array suitable for LSTM input.
3) *Model creation:* selecting the appropriate model is crucial for developing an effective learning model. The creation process is executed by specifying certain parameters, such as the quantity of neurons, the number of layers, and the activation functions used.
4) *Training:* The dataset will be partitioned into two segments. The initial segment will be employed for model training, specifically for fine-tuning the weights to align the predictions with the anticipated outcomes.
5) *Evaluation:* The second portion of the dataset will be utilized for the purpose of testing and evaluating the model. The testing data is both smaller in size and distinct from the training data.
6) *Prediction:* Once the training and testing steps are complete, the model can be utilized to make predictions, specifically about land suitability.

Algorithm 2 LSTM prediction model pseudo code.

Input: Datasets of weather condition.
Output: LSTM_model
Begin
normalize data(0,1); ▹ Normalizing data values in range 0 and 1.
$x\_train, y\_train, x\_test, y\_test \leftarrow splitdata(data, 25)$; ▹ Split data to training and testing sets (25% testing)
reshape data(data) ▹ Reshape data according to LSTM input
$model \leftarrow CreateSequentialModel()$ ▹ 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
$results \leftarrow model.predict(x\_test)$; ▹ Testing the model and making predictions
End.

5. Results and discussion

This section outlines the configurations that enable us to implement our suggested design, hence our prediction model. Moreover, this section summarises the attained outcomes and provides a discussion related to the subject matter.

A. Simulation's settings and configurations

Our proposed approach was implemented in Python on a machine with an Intel i7 processor and consisting of 16 GB RAM. The architecture of our model consists of an input layer, three hidden layers, and one output layer. The input layer is provided with a three-dimensional array consisting of four features, where each feature is represented by a one-dimensional array with four columns.

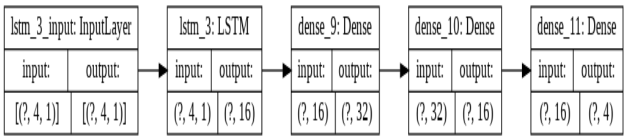


Figure 6. LSTM model architecture and layers.

B. Obtained results and Discussions

In this subsection we present some results about our proposed system, in fact, it shows some plots for our model resulted from training and constructing model. Also, it gives the data visualization about the degree of suitability for a given area.

1) *Interfaces:* As we mentioned earlier, our system is web-based application which give the possibility to the user to use anywhere. Plus, this application could be used by any means whether mobile or other device. Our system also gives the possibility to help the experts of farmers to give a real-time data streaming from the field with two types of data weather data and soil data including PH values and quantity of water from the humidity of the sol.
To check if the given area is suitable or not the next figures (see figure 7 and 8) help the experts to see the degree of suitability of chosen area according to its climatic factors. The next figure (see figure 8) helps

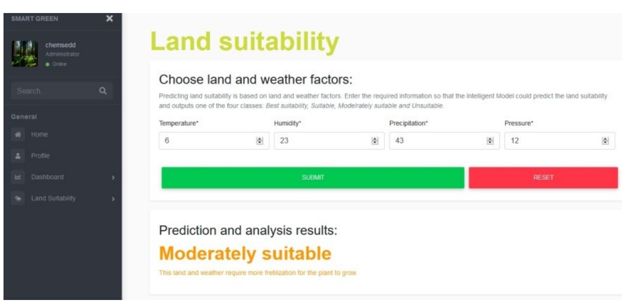


Figure 7. Land suitability from a chosen preference.

the experts to visualize the degree of suitability of chosen area as choropleth map results from Cloud IoT layer. Figure 8 shows suitability degree from 1 to 4 (unsuitable, moderately suitable, suitable and best suitability respectively).

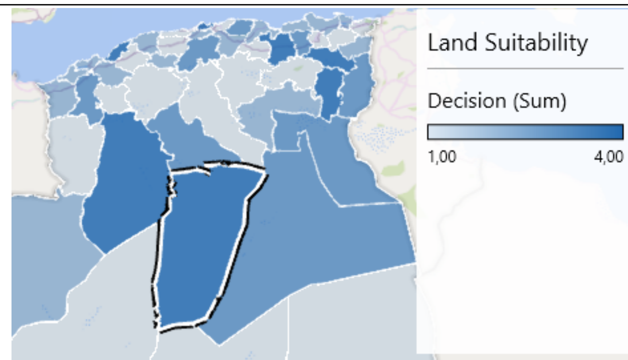


Figure 8. Land suitability using visualization.

2) *Model creation:* the next figure shows the results after training step. As we can see, we have four classes, which correspond Unsuitable, moderately suitable, suitable and Best suitability, respectively. From fig-

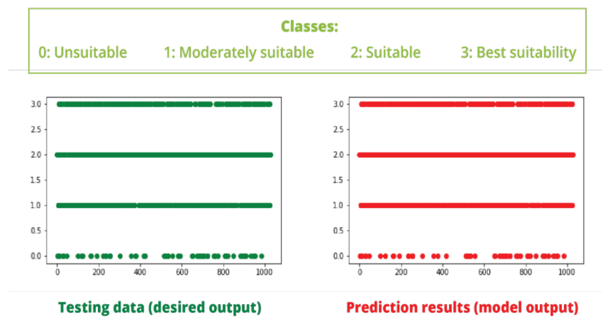


Figure 9. LSTM model -a- Results.

ure 9 we can see that the testing data presented in green and the prediction results presented in red are almost identical, which shows the effectiveness of our model. Figure 10 shows the difference between the desired output and the predicted output, which is given by our model. Google Colab was used for the

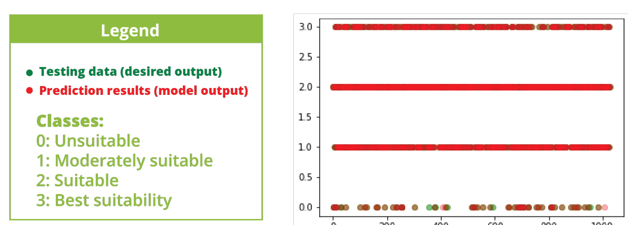


Figure 10. LSTM model -b- Results.

model's training phase. 200 epochs of training with a five-person batch size were conducted. In order to accurately identify and predict whether the land will be appropriate or not, our LSTM model was trying to understand the relationship between weather data and the land suitability label. As seen in the image

below, accuracy was very poor at the start of the training, with a large loss value.
After 200 epochs, the accuracy at the end of the training has increased to 0.98, while the loss value has decreased to 0.06 compared to the initial values. The LSTM model was able to match the data, as seen by its increasing accuracy over time, which went from 0.44 to 0.98. Predicting land suitability accurately the majority of the time, the validation accuracy, which varies from 0.95 to 0.98, is likewise thought to be very high. The trained model was able to predict accurately for an entirely separate collection of data since it used a different validation set of data that wasn't part of the training set.

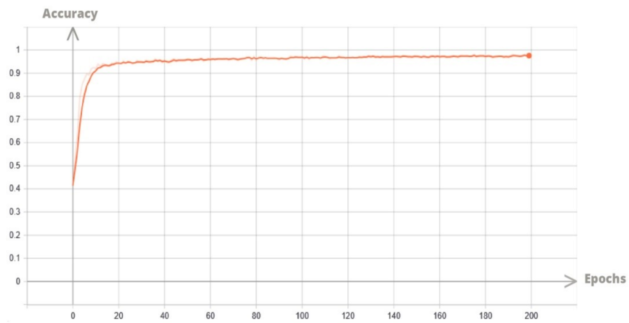


Figure 11. LSTM accuracy.

We employed the Categorical Cross Entropy function to compute the loss. It is the predominant function in the context of categorization difficulties. The Categorical Cross Entropy metric rises as the anticipated probability deviates from the true label. The model operates by minimizing the loss function, which condenses all components of our algorithm into a single numerical measure that quantifies the effectiveness of our model. During the initial phases of the training, the loss function had a high value of 1.2. The LSTM model endeavors to learn from the data and subsequently reduce the loss by employing backpropagation as the training progresses. After a few epochs, the value decreases dramatically to 0.06, indicating that the model successfully extracted information from the dataset and identified the pattern to accurately predict land suitability.

C. Evaluation metrics

Performance evaluation is an essential stage in validating the efficiency of a machine learning system. Various metrics are employed for this objective, including Recall, Precision, F1 score, Accuracy, and Confusion Matrix [41] [42]. These equations are calculated using the number of correct and incorrect class as refereed True positive/negative and False positive/negative. The used metrics are measured as follows.

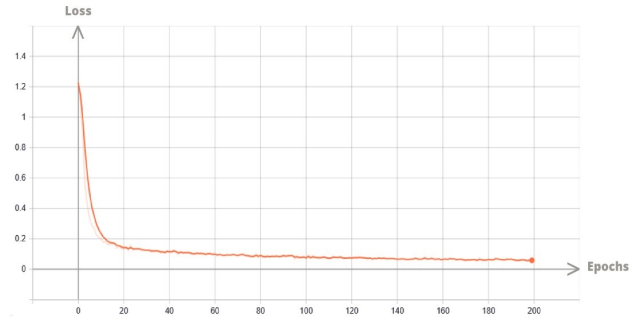


Figure 12. LSTM loss.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 - score = \frac{2 \times (precision \times recall)}{precision + recall} \quad (5)$$

TABLE I. Comparative performance of ML models.

Model	Accuracy	Precision	Recall	F1-score
SVM	0.6533	0.6468	0.6533	0.6338
Linear Regression	0.5012	0.4504	0.5012	0.3949
Logistic Regression	0.4663	0.3683	0.4663	0.4038
LSTM	**0.9805**	**0.9809**	**0.9805**	**0.9805**

As we can see from obtained results dipected in Table I which gives a comparative study between LSTM and other ML techniques. Particularly, the LSTM model demonstrated superior performance in capturing the temporal dynamics of the remote sensing data compared to more traditional machine learning approaches. Our deep learning model was able to achieve an overall accuracy of 98% in predicting land suitability, which is a significant improvement over the baseline pixel-based classification accuracy of 85%. Further, the model was able to provide detailed insights into the important drivers of land suitability, with the Sentinel-derived vegetation indices and soil moisture data being the most predictive features.

The bellow figure (see Figure 13) explains the differences between standard deviation changes over evaluation metrics the prediction accuracy. Where the next figure 14 explains different confusion matrix between ML techniques and LSTM. As we can see the LSTM has the highest accuracy and lowest errors for different land use classes compared to the other methods.

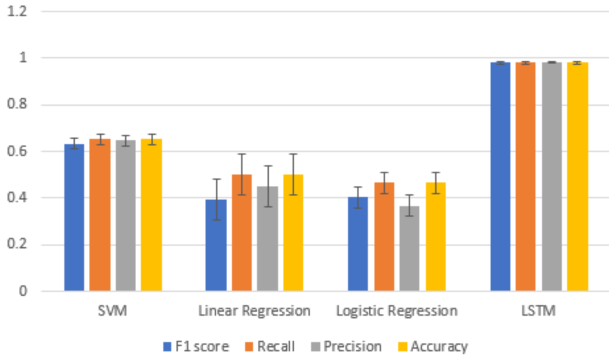
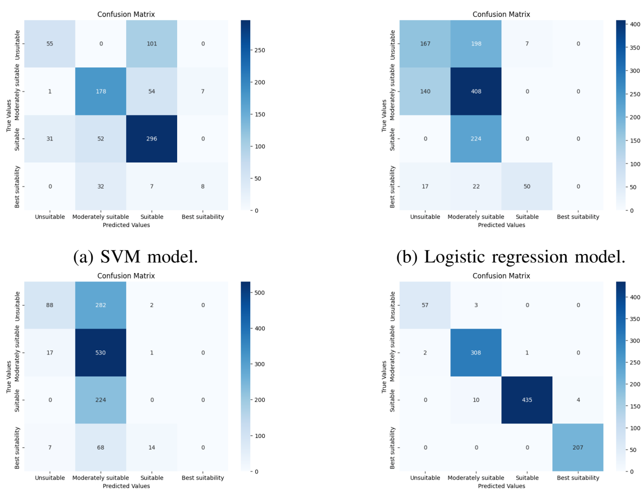


Figure 13. Comparison between different ML models.



(a) SVM model.

(b) Logistic regression model.

(c) Linear regression model.

(d) LSTM model.

Figure 14. Confusion matrix for different models

6. Conclusion and Future Work

Land suitability is a crucial factor in agriculture and farming, and our primary objective is to offer accurate predictions in this regard. The suggested architecture, which integrates DL and IoT, demonstrated efficacy in predicting land suitability by leveraging data on weather and soil conditions. In contrast to the existing literature on smart farming and land appropriateness, our strategy additionally prioritized the utilization of a data streaming mechanism. Gathering data using the most effective method will undoubtedly enhance outcomes. Our method utilizes the Long-Short Term Memory (LSTM) model to make predictions with time-series data, instead of using traditional Artificial Neural Networks and other ML methods. This is because time series data exhibits patterns over time, rather than being composed of isolated data points. As have seen in the previous sections, farms generate different types of data which will create a problem of data heterogeneity. Our solution consists of proposing ontology to deal with issue. In addition, we have proposed an algorithm to fetch information through data formats. Unfortunately, our solution was hindered by a lack of data. The available historical data spans a period of only nine years, which is relatively limited, with only one data point recorded per day.

In the future, we are open to incorporating additional functionalities into our system, like cameras to keep an eye on the health of the plants. Every day, pictures will be taken and kept on the server databases. Convolutional Neural Networks (CNN) will thereafter receive all of the saved photos in order to process them and extract pertinent data regarding the health of the plants. The addition of more sensors to the field will undoubtedly improve forecast accuracy by enriching the dataset. From an alternative angle, we would like to equip every agricultural truck with a GPS tracker so that the server software can locate each one on the field. Giving farmers complete management authority over their farms will undoubtedly increase agricultural output and make the work even easier.

References

[1] S. Prathibha, A. Hongal, and M. Jyothi, "Iot based monitoring system in smart agriculture," in *2017 international conference on recent advances in electronics and communication technology (ICRAECT)*. IEEE, 2017, pp. 81–84.

[2] I. Mat, M. R. M. Kassim, A. N. Harun, and I. M. Yusoff, "Smart agriculture using internet of things," in *2018 IEEE conference on open systems (ICOS)*. IEEE, 2018, pp. 54–59.

[3] K. Lakhwani, H. Gianey, N. Agarwal, and S. Gupta, "Development of iot for smart agriculture a review," in *Emerging Trends in Expert Applications and Security: Proceedings of ICETEAS 2018*. Springer, 2019, pp. 425–432.

[4] D. D. K. Rathinam, D. Surendran, A. Shilpa, A. S. Grace, and J. Sherin, "Modern agriculture using wireless sensor network (wsn)," in *2019 5th international conference on advanced computing & communication Systems (ICACCS)*. IEEE, 2019, pp. 515–519.

[5] M. Ghasemaghaei, S. Ebrahimi, and K. Hassanein, "Data analytics competency for improving firm decision making performance," *The Journal of Strategic Information Systems*, vol. 27, no. 1, pp. 101–113, 2018.

[6] J. Ye, "Generalized ordered weighted simplified neutrosophic cosine similarity measure for multiple attribute group decision making," *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, vol. 14, no. 1, pp. 51–62, 2020.

[7] F. Harahap, R. Sitompul, A. Rauf, D. Harahap, H. Walida *et al.*, "Land suitability evaluation for oil palm plantations (elaeisguenensis jacq) on sitellu tali urang julu, pakpak bharat district," in *IOP Conference Series: Earth and Environmental Science*, vol. 260, no. 1. IOP Publishing, 2019, p. 012116.

[8] S. Kim, M. Lee, and C. Shin, "Iot-based strawberry disease prediction system for smart farming," *Sensors*, vol. 18, no. 11, p. 4051, 2018.

[9] R. Khebbache, A. Merizig, K. Rezeg, and J. Lloret, "The recent technological trends of smart irrigation systems in smart farming: a review," *International Journal of Computing and Digital Systems*, vol. 14, no. 1, pp. 10 317–10 335, 2023.

[10] A. Somov, D. Shadrin, I. Fastovets, A. Nikitin, S. Matveev, O. Hrinchuk *et al.*, "Pervasive agriculture: Iot-enabled greenhouse for plant growth control," *IEEE Pervasive Computing*, vol. 17, no. 4, pp. 65–75, 2018.

[11] S. Amini, A. Rohani, M. H. Aghkhani, M. H. Abbaspour-Fard, and M. R. Asgharipour, "Assessment of land suitability and agricultural production sustainability using a combined approach (fuzzy-ahp-gis): A case study of mazandaran province, iran," *Information Processing in Agriculture*, vol. 7, no. 3, pp. 384–402, 2020.

[12] A. A. Sulaiman, Y. Sulaeman, N. Mustikasari, D. Nursyamsi, and A. M. Syakir, "Increasing sugar production in indonesia through land suitability analysis and sugar mill restructuring," *Land*, vol. 8, no. 4, p. 61, 2019.

[13] M. M. Subashini, S. Das, S. Heble, U. Raj, and R. Karthik, "Internet of things based wireless plant sensor for smart farming," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 10, no. 2, pp. 456–468, 2018.

[14] F. Wang, Z. Zhang, C. Liu, Y. Yu, S. Pang, N. Duić, M. Shafie-Khah, and J. P. Catalão, "Generative adversarial networks and convolutional neural networks based weather classification model for day ahead short-term photovoltaic power forecasting," *Energy conversion and management*, vol. 181, pp. 443–462, 2019.

[15] M. Noura, M. Atiquzzaman, and M. Gaedke, "Interoperability in internet of things: Taxonomies and open challenges," *Mobile networks and applications*, vol. 24, pp. 796–809, 2019.

[16] M. Al-Osta, A. Bali, and A. Gherbi, "Event driven and semantic based approach for data processing on iot gateway devices," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 12, pp. 4663–4678, 2019.

[17] S. V. Limkar and R. K. Jha, "A novel method for parallel indexing of real time geospatial big data generated by iot devices," *Future generation computer systems*, vol. 97, pp. 433–452, 2019.

[18] M. Munir, S. A. Siddiqui, M. A. Chattha, A. Dengel, and S. Ahmed, "Fusead: Unsupervised anomaly detection in streaming sensors data by fusing statistical and deep learning models," *Sensors*, vol. 19, no. 11, p. 2451, 2019.

[19] Y. Liu, K. Tountas, D. A. Pados, S. N. Batalama, and M. J. Medley, "L1-subspace tracking for streaming data," *Pattern Recognition*, vol. 97, p. 106992, 2020.

[20] Ö. Köksal and B. Tekinerdogan, "Architecture design approach for iot-based farm management information systems," *Precision Agriculture*, vol. 20, pp. 926–958, 2019.

[21] F. Bu and X. Wang, "A smart agriculture iot system based on deep reinforcement learning," *Future Generation Computer Systems*, vol. 99, pp. 500–507, 2019.

[22] T.-C. Hsu, H. Yang, Y.-C. Chung, and C.-H. Hsu, "A creative iot agriculture platform for cloud fog computing," *Sustainable Computing: Informatics and Systems*, vol. 28, p. 100285, 2020.

[23] A. Pathak, M. AmazUddin, M. J. Abedin, K. Andersson, R. Mustafa, and M. S. Hossain, "Iot based smart system to support agricultural parameters: A case study," *Procedia Computer Science*, vol. 155, pp. 648–653, 2019.

[24] U. J. L. dos Santos, G. Pessin, C. A. da Costa, and R. da Rosa Righi, "Agriprediction: A proactive internet of things model to anticipate problems and improve production in agricultural crops," *Computers and electronics in agriculture*, vol. 161, pp. 202–213, 2019.

[25] W.-L. Chen, Y.-B. Lin, Y.-W. Lin, R. Chen, J.-K. Liao, F.-L. Ng, Y.-Y. Chan, Y.-C. Liu, C.-C. Wang, C.-H. Chiu *et al.*, "Agritalk: Iot for precision soil farming of turmeric cultivation," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5209–5223, 2019.

[26] M. K. Senapaty, A. Ray, and N. Padhy, "Iot-enabled soil nutrient analysis and crop recommendation model for precision agriculture," *Computers*, vol. 12, no. 3, p. 61, 2023.

[27] V. Shevchenko, A. Lukashevich, D. Taniushkina, A. Bulkin, R. Grinis, K. Kovalev, V. Narozhnaia, N. Sotiriadi, A. Krenke, and Y. Maximov, "Climate change impact on agricultural land suitability: An interpretable machine learning-based eurasia case study," *IEEE Access*, 2024.

[28] P. Jutadhamakorn, T. Pillavas, V. Visoottiviseth, R. Takano, J. Haga, and D. Kobayashi, "A scalable and low-cost mqtt broker clustering system," in *2017 2nd International Conference on Information Technology (INCIT)*. IEEE, 2017, pp. 1–5.

[29] A. Merizig, O. Kazar, and M. L. Sanchez, "A multi-agent system approach for service deployment in the cloud," *International Journal of Communication Networks and Distributed Systems*, vol. 23, no. 1, pp. 69–92, 2019.

[30] S. A. Butt, M. I. Tariq, T. Jamal, A. Ali, J. L. D. Martinez, and E. De-La-Hoz-Franco, "Predictive variables for agile development merging cloud computing services," *IEEE Access*, vol. 7, pp. 99 273–99 282, 2019.

[31] Y. Drohobytskiy, V. Brevus, and Y. Skorenkyy, "Spark structured streaming: Customizing kafka stream processing," in *2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP)*. IEEE, 2020, pp. 296–299.

[32] M. T. Tun, D. E. Nyaung, and M. P. Phyu, "Performance evaluation of intrusion detection streaming transactions using apache kafka and spark streaming," in *2019 international conference on advanced information technologies (ICAIT)*. IEEE, 2019, pp. 25–30.

[33] B. Blamey, A. Hellander, and S. Toor, "Apache spark streaming, kafka and harmonicio: a performance benchmark and architecture comparison for enterprise and scientific computing," in *International Symposium on Benchmarking, Measuring and Optimization*. Springer, 2019, pp. 335–347.

[34] J. Kiljander, A. D'elia, F. Morandi, P. Hyttinen, J. Takalo-Mattila, A. Ylisaukko-Oja, J.-P. Soininen, and T. S. Cinotti, "Semantic interoperability architecture for pervasive computing and internet of things," *IEEE access*, vol. 2, pp. 856–873, 2014.

[35] S. Aydin and M. N. Aydin, "Semantic and syntactic interoperability for agricultural open-data platforms in the context of iot using crop-specific trait ontologies," *Applied Sciences*, vol. 10, no. 13, p. 4460, 2020.

[36] L. Feng, G. Chen, and J. Peng, "An ontology-based cognitive model for faults diagnosis of hazardous chemical storage devices," *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, vol. 12, no. 4, pp. 101–114, 2018.

[37] M. N. Omri, "Fuzzy ontology-based querying user'requests under uncertain environment," *International Journal of Cognitive Infor-*

matics and Natural Intelligence (IJCINI), vol. 14, no. 3, pp. 41–59, 2020.

[38] A. Merizig, H. Saouli, M. L. Sanchez, O. Kazar, and A.-N. Benharkat, "An extended data as a service description model for ensuring cloud platform portability," in *Advanced Intelligent Systems for Sustainable Development (AI2SD'2018) Volume 5: Advanced Intelligent Systems for Computing Sciences*. Springer, 2019, pp. 736–744.

[39] D. Hooda and R. Rani, "Ontology driven human activity recognition in heterogeneous sensor measurements," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 12, pp. 5947–5960, 2020.

[40] M. Al-Smadi, O. Qawasmeh, M. Al-Ayyoub, Y. Jararweh, and B. Gupta, "Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of arabic hotels' reviews," *Journal of computational science*, vol. 27, pp. 386–393, 2018.

[41] A. Merizig, H. Belouaar, M. M. Bakhouche, and O. Kazar, "Empowering customer satisfaction chatbot using deep learning and sentiment analysis," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 3, pp. 1752–1761, 2024.

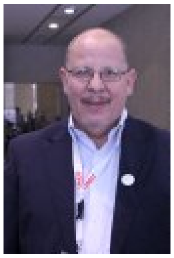
[42] D. Houfani, S. Slatnia, O. Kazar, I. Remadna, H. Saouli, G. Ortiz, and A. Merizig, "An improved model for breast cancer diagnosis by combining pca and logistic regression techniques," *International Journal of Computing and Digital Systems*, vol. 13, no. 1, pp. 701–716, 2023.

Toufik Bendahmane Obtained his Msc degree by 2015 from Mohamed Khider University, Biskra, Algeria, He is working on an artificial intelligence field. Toufik Bendahmane is now an Assistant Professor at the computer science department of Biskra University. Also, he is a member of LINFI Laboratory at the same University. His research interests include Agriculture 4.0, Ontology, Interoperability, Cloud Computing and Internet of Things. He can be contacted at toufik.bendahmane@univ-biskra.dz.

Abdelhak Merizig Obtained his PhD degree by 2018 from Mohamed Khider University, Biskra, Algeria, He is working on an artificial intelligence field. Abdelhak Merizig is now an Assistant Professor at the computer science department of Biskra University. Also, he is a member of LINFI Laboratory at the same University. Abdelhak Merizig is a member of the scientific committee in several International conferences and act reviewer in different journals. His research interests include Agriculture 4.0, multi-agent systems, Human machine Interaction, Natural Language Processing, Cloud Computing and Internet of Things. He can be contacted at a.merizig@univ-biskra.dz.

Khaled Rezeg was born in Biskra (Algeria) in 1968. He received the engineer degree in computer science from Annaba University, Annaba, Algeria, in 1992, his magister degree in computer science in artificial intelligence and imagery (2003), his Doctoral degree (2011), and his university habilitation degree (2015) from the University of Biskra (Algeria). He is a full professor in the computer science department of the University of Biskra since 2021. He is a researcher with the LINFI laboratory since 2013. He has been the head of a team in the same laboratory since 2013. Pr. Rezeg is a member of many TPCs at conferences and acts as a reviewer in several journals. In fact, Pr Rezeg is interested in: Geographic Information Systems, Human-Machine interface engineering, ontologies, and automatic natural language processing. He can be contacted at k.rezeg@univ-biskra.dz.

Okba Kazar obtained a state engineer's diploma and master's and doctoral degrees from the University of Constantine (field of computer science and artificial intelligence). He has published more than 400 papers in international journals and communications at international conferences. He participated as a chair and session chair in international conferences. He supervised more than 40 PhD and 100 master's projects. He has published three books: "The Artificial Intelligence Handbook," "Big Data Security," and "AI with the Practice." Moreover, he has published more than 14 book chapters. His main research area is artificial intelligence, he is interested in multi-agent systems and their applications, PHM in medical and industrial fields, enterprise resource planning (ERP), advanced information systems, robotics, web services, semantic web, big data, IoT and cloud computing. He held the rank of visiting professor at many universities in Europe, especially in France and Spain. Pr. Okba Kazar was a professor in the Department of Computer Science at the University of Biskra, where he contributed to its founding. He also founded and was director of the Smart Computer Science Laboratory at the University of Biskra. He was a visiting professor at the United Arab Emirates University (UAE, Abu Dhabi). He currently works at the University of Sharjah, Kalba branch, in the Emirate of Sharjah. He can be contacted at okazar@sharjah.ac.ae



Saad Harous received the Ph.D. degree in computer science from Case Western Reserve University, Cleveland, OH, USA, in 1991. He has more than 30 years of experience in teaching and research in three different countries, including USA, Oman, and United Arab Emirates. He is currently a Professor with the College of Computing and Informatics, University of Sharjah, United Arab Emirates. His teaching interests include programming, data structures, design and analysis of algorithms, operating systems, and networks. He has published more than 200 journal articles and conference papers. His research interests include parallel and distributed computing, P2P delivery architectures, wireless networks, and the use of computers in education and processing Arabic language. He can be contacted at harous@sharjah.ac.ae.