IoT and AI Based Smart Soil Quality Assessment for Data-Driven Irrigation and Fertilization

Nyakuri Jean Pierre, Bizimana Judith, Bigirabagabo Aaron, Kalisa Jean Bosco, Gafirita James, Munyaneza Midas Adolphe, Nzemerimana Jean Pierre
IoT and AI Based Smart Soil Quality Assessment for Data-Driven Irrigation and Fertilization

1 Nyakuri Jean Pierre, 2 Bizimana Judith, 3 Bigirabagabo Aaron, 4 Kalisa Jean Bosco, 5 Gafirita James, 6 Munyaneza Midas Adolphe, 7 Nzemerimana Jean Pierre.

1,2 Post graduate students, University of Rwanda-African Center of Excellence in Internet of Things (ACEIoT)

3,4,5,7 Lecturer, Rwanda Polytechnic- Integrated Polytechnic Region College of Gishari (IPRC-Gishari)

6 Lecturer, Rwanda Polytechnic- Integrated Polytechnic Region College of Kigali (IPRC-Kigali)

*Corresponding Author’s Email: njpindian@yahoo.fr

ABSTRACT

Purpose: The rapidly growing demand for food due to rapid population growth in East Africa is one of the challenging issues and the sustainable way of tackling it, is to enhance the agriculture activities to satisfy the need of increasing farm productivity. However, the climate change, limited water resources and poor soil fertility reduces crops yields. In attempt to solve these challenges, Internet of thing (IoT) in conjunction with artificial intelligence (AI) techniques is increasingly being used in agriculture sector. This study investigates an integration of IoT and a deep learning (DL) driven solution for smart irrigation and fertigation by assessing soil nutrients and soil water content dynamics in Eastern province of Rwanda for optimization of these scare resources while increasing yields productivity.

Methodology: The research data for analysis was collected from KABOKU-KAGITUMBA irrigation scheme, and data on soil moisture and soil nutrients was gathered over a six-month period from 36 sensor nodes that were installed in approximately 70 hectares with 6 pivots for irrigation. The collected data in real time by sensors was sent to an IoT platform and incorporated with the forecasted weather information there after a deep learning based model used to predict when to irrigate and when to fertigate and the notification sent to farmer with recommendations. The irrigation valves were automatically actuated based on the predictions. The study's main software tools for gathering, displaying, and analyzing real-time data streams were Things Speak, Tensor Flow Lite, and the Arduino Software (IDE). A prototype was finally implemented effectively.

Findings: The resulting model showed that can perform well with an accuracy of 91.7% and it can work well when deployed in the remote area with minimum internet connection.

Unique Contribution to Practice: since the currently technologies used in irrigation and fertilization are manual or based on threshold values for automatic irrigation, we recommend the implementation of this solution since it will guarantee data-driven farming, which will help to protect the environment and ensure the optimization use of water resources. Additionally, this will result in lower operating cost, which will raise earnings.

Keywords: IoT, Smart Farming, Smart Irrigation, Deep Learning, Smart Fertilization
INTRODUCTION

In Rwanda about 96% of rural households depend mainly on agriculture for their income[1]. Low productivity and resource waste are, nevertheless, becoming more prevalent. The effectiveness in managing water utilization is an issue shared by farmers of various crops and animals in arid and semi-arid regions.

One way of ensuring the efficient management of this scarce resource is the use of an automated irrigation system. In addition to the issue of water waste, fertilizer waste results from careless application. The Government of Rwanda has implemented numerous policies to enhance the growth of this industry in an effort to address these and other challenges. One such policy is the ICT4Ag (ICT for Agriculture) strategy, which aims to create an environment that is favorable for the creation, adoption, and use of ICT[2]. Reaching a large audience and ensuring accelerated productivity and greater efficiency in the agricultural industry are both made possible by the use of ICT in agriculture. Through the creation of websites and other interfaces through which such information can be accessed, ICT enables farmers and other stakeholders to receive the required and pertinent information on time. Through ICT-driven learning initiatives, farmers will also be able to advance their knowledge and abilities. Additionally, this will result in better access to information and employment possibilities.

Internet of Things (IoT) and Machine Learning provide capabilities that can be exploited in enhancing data driven farming, which is a concept referred to as smart farming[3]. The use of automated management systems and technology replaces conventional agricultural practices in smart farming [4]. Smart agriculture integrates the Internet of Things (IoT) and Information and Communication Technology (ICT) developing technologies to boost the optimization of various management models and the technologies used in agricultural production. Long-term, this may result in less consumption of agricultural resources and more production in this industry. IoT allows for the deployment of a variety of devices and sensors in farms to track a variety of soil and environmental characteristics, with the collected data being communicated over the internet in almost real-time [5]. Farmers may observe the state of the field from any location. Local farms can make better decisions by using this information to swiftly assess the data and combining it with outside sources of information, like weather services. Artificial intelligence (AI) for forecasts and function automation can help to further improve this [6]. IoT has been used to improve irrigation systems and to ensure data driven fertilization as stated in [7]. For example, an automated irrigation system in which the soil moisture condition is monitored and maintained at desired levels by the use of an automatic watering system, a Raspberry Pi is used in an IoT based smart irrigation solution. In this system a webcam is used to monitor the growth of plants and the control of the watering system done [8].

The presence of nutrients in the soil is monitored and analysed based on the principle of colorimetric.. Data from the field is collected using specially created NPK sensors, and the data is then transferred to the Google cloud for archiving and display. A fuzzy logic inference system is utilized to identify errors in the sensed data[9][10]. These solutions demonstrate the IoT's possibilities, which can be further utilized. However, in order to increase efficiency, it is also necessary to apply the most recent AI approaches, which is why this study presents the application of deep learning. Due to DL algorithms' capacity to extract high-level features from data, DL was chosen. This DL capacity outperforms other conventional Machine Learning techniques [11].

This study, therefore, presents a prototype for an IoT and a deep learning driven solution for smart
fertigation. A deep learning model is used to predict when to irrigate using data that is analyzed on an ARM Cortex 4 based Arduino Nano 33 BLE sensing that collects real-time information on soil moisture and soil nutrients. Farmers receive text message alerts with recommendations for fertilizers to add if soil nutrient deficits are found. According to the forecasts and occasionally, the irrigation valve is automatically opened. This system's successful deployment will decrease water and fertilizer waste while boosting productivity.

The data used for training the machine learning model was collected by STES group, which is a commercial IoT company in Rwanda that offers smart irrigation for six months. This presented solution is seen as an improvement to the existing system that is based on threshold values which does not take into account the changes in soil quality over time.

The rest of the paper is organized as follows; the next section highlights a review of related systems, section III describes the research methodology, section IV outlines the system level design for the solution, Section V presents results, analysis and discussions followed by a conclusion and future works.

RELATED WORKS

This section outlines a state-of-the-art analysis of existing related solutions. Firstly, smart irrigation systems are reviewed followed by a review of smart fertilization systems and lastly an analysis of how artificial intelligence is applied to improve such systems.

Smart Irrigation Systems

To begin with, an automated irrigation and monitoring system consisting of temperature and moisture sensors, a raspberry pi and a water pump is proposed. Smart phones module is used for communication just by irrigating based on the water requirements for the plants alone may not mean that optimal levels of required moisture is reached[12]. Some water will be lost in the soil by environmental factors. The use of smart phone for communication is also not appropriate for remote locations where ownership and connectivity are a challenge hence the need for improvements. Secondly, a smart irrigation system has been developed in [13]. In the solution, sensors like soil moisture sensor, water level sensor and the temperature sensor are connected to the microcontroller. Upon receiving the signal from those sensors, the microcontroller gives the appropriate output that turns on the relay and operates the water pump[14]. This solution uses threshold values to actuate for irrigation. This does not take into account the trends and gradual change of soil properties. Thirdly, [15] suggests a wireless sensor network-based smart farm irrigation system that uses an Android phone to remotely monitor and control a drip irrigation system. The use of mobile phones is a challenge for remote location in Africa, necessitating the need for improvements of this solution. Last but not least, an automated irrigation system has been proposed in [16]. The solution measures the soil's properties, including soil moisture, pH, and humidity. The LCD is used to display the sensed values and provided recommendations. This solution is based on threshold values that are not recommended for time series data hence the need to use AI. Even though these solutions are an improvement from the manual irrigation processes, there is also a need to integrate machine learning to enable predictions for a more efficient data driven process thus the need for this study.

Smart Fertilization Systems

These systems are mainly focused on ensuring fertilizers are applied only when needed and in right
amounts. First, a smart system is suggested so that farmers can receive all pertinent information about how to improve soil fertilization and agriculture by receiving climate change information from IoT devices [17]. This solution can help in guiding the farmers in the usage of fertilizers, however does not integrate this to an irrigation process to ensure maximum utilization of resources. Secondly, in another related solution, an IoT device is used to sense the agricultural data and it is stored into the Cloud database. Cloud based big data analysis is used to analyze the data visualization, fertilizer requirements, analysis of the crops, as well as market and stock requirements for the crop. The use of a mobile app may be challenging to rural farmers hence the need for alternative notification methods. Training on the cloud also means additional connectivity and energy costs. Moreover, a generic IoT framework for improving agriculture yield by effectively scheduling irrigation and fertilization based on the crops’ current requirements, environmental conditions and weather forecasts is proposed in [18].

The use of a mobile application in this solution makes it challenging to implement in a rural setting. In addition, a novel method that uses IoT and Wireless Sensor Network in Farming is presented. Soil Moisture sensor and humidity sensor monitor the soil properties. These results are updated to the IoT Server. Based on the values for humidity of the soil, drip irrigation to the plant is on/off through solenoid valve [19]. This solution applies the use of thresholds that does not take into account a progressive change in soil parameters. Moreover, the solution does not integrate soil fertility sensors which are important in determining the soil properties. Finally, a novel Nitrogen-Phosphorus-Potassium (NPK) sensor is designed using Light Dependent Resistor (LDR) and Light Emitting Diodes (LED) [20]. The principle of colorimetric is used to monitor and analyze the nutrients present in the soil. This solution does not integrate to an irrigation system and with the customized sensor it is not easy to tell the exact amounts of NPK in the soil hence the need for improvement. Such solutions show the potential of IoT and machine learning in fertigation. However, there is a need to not only notify the farmers via a cheap solution but also use more advanced ML techniques so as to predict when to apply the fertilizers and integrate the process to the irrigation function so as to reduce wastages.

Use of Artificial Intelligence in Irrigation and Fertilization

Different Artificial Intelligence (AI) techniques have been used in improving irrigation and fertilization efficiency. First, a model for Decision Support System (DSS) which is used in the Internet of Things (IoT) based agricultural application is proposed [21]. In another solution, the system uses Soil Moisture, DHT11 and also pH Meter sensors to obtain soil moisture, humidity and acidity (pH) soil data. Similarly, a Fuzzy logic controller is used to compute input parameters (e.g. soil moisture, temperature and humidity) and to produce outputs of motor status. Additionally, the technology stops the motor to conserve energy when rain is forecast and protects the crop grown with solar panels from unrestricted rain [22].

Also, an Internet of Things based irrigation system that works at reducing the frequency of irrigation while increasing the rate of production through the use of fuzzy logic is proposed. The Naive Bayes algorithm and machine learning have been used to construct an AI expert system that uses sensor data from agriculture. This work was used to analyze the quality of fertilizer, the effectiveness of insecticides, and the quantity of water needed to irrigate the crops. In another solution, an IoT device is used to sense the agricultural data and it is stored into the Cloud database. Big data analysis powered by the cloud is utilized to examine crop requirements, market and stock needs, and fertilizer requirements. Then, a prediction is made using data mining techniques, and
the farmer receives the information via a mobile app [23].

Various calculations and channels are also developed to acquire and manage the colored images of the dirt examples. Due to the increasing availability of labeled datasets, there is a move to deep learning methods for more accurate prediction systems that can lead to an improvement of the existing solutions. This study therefore proposes the use of deep learning-based models. The study also explores pruning and quantization techniques so as to make the models lighter to run on embedded devices.

**RESEARCH METHODOLOGY**

The methods, tools and processes and steps undertaken in the research are outlined in this section. The research steps are first presented followed by the AI model training steps and lastly the selected system development method used in developing the prototype.

**Research Steps**

The research process began in the month of January 2021 with a review of related literature in an endeavour to identify gaps for further study. Then, an idea of possible topic was proposed and the research proposal was developed. After the research proposal’s approval, the research process proceeded with the selection of tools and methods, data collection, AI model design and training, system design, prototyping and documentation of results.

**Data Collection**

Due to the limited time, the data used for the study was collected in collaboration with STES group, a commercial company in Rwanda that offers smart farming solution. The data was collected for a period of 6 months from March 2021 to September 2021, of interest was soil moisture data and when the crops were irrigated. Microsoft excel was used to format the data and remove fields that were not needed for our solution. This data was collected from KABOKU-KAGITUMBA Irrigation scheme, where 35 devices are installed in about 70ha with 6 pivots for irrigation.

**Machine Learning Steps**

The machine learning steps followed the normal steps undertaken in any machine learning algorithm. Figure 1 shows the machine learning steps. It began with data collection. The quality of data determined how accurate the model would be. Data was collected for a period of 6 months to ensure there was enough data for training. The sensors were also calibrated from time to time to ensure the quality was maintained through the process. The collected data was cleaned in preparation for training. Fields that were not needed were deleted, dealt with missing values, corrected errors and removed duplicates. The data was then converted into JSON format and uploaded into the cloud based machine learning platform. The data was also separated into validation set and training sets. The data was then pre-processed so as to generate features for training. After pre-processing the machine learning model was designed and the inputs to the model training were the soil moisture level and soil temperature.

The raw features of the data were selected and a Keras Neural Network was selected for training the model with the expected output being a classification as to whether there is a stress condition in the soil or ideal condition. The model was then trained. At this stage different number of epochs, dense layers and learning rates were experimented till the optimum training parameters were reached. The on-device resources needed by the model were also tested. This step also included
hyper parameter tuning, used to tune model parameters for improved performance. The test data was used to evaluate the performance of the model on unseen data. This unseen data is meant to be somewhat representative of model performance in the real world, but still helps tune the model (as opposed to test data, which does not). The model was also optimized for performance in an embedded device. After evaluation, the model was converted into a tiny machine learning model that can run on an embedded device in readiness for deployment. Using further (test set) data which had, until this point, been withheld from the model (and for which class labels are known), were used to test the model in both the cloud environment and on the embedded device; a better approximation of how the model will perform in the real world.

![Machine learning steps](image)

Fig. 1: machine learning steps

**Software Tools**

The main software tool used in the study include: The open-source Arduino Software (IDE), TensorFlow Lite which is an open-source, product ready, cross-platform deep learning framework for converting the pre-trained model in TensorFlow to a special format that can be optimized for speed or storage. In addition, ThingsSpeak which is an IoT analytics platform service was used to aggregate, visualize, and analyze live data streams.

**System Development Model**

The Prototype model was selected as the prototype development model. The basic idea in Prototype model is that instead of freezing the requirements before a design or coding proceeds, a throwaway prototype is built to understand the requirements. This prototype is developed based on the currently known requirements.

**SYSTEM LEVEL DESIGN**

In this section the embedded system level design for the proposed system is presented. The high-level system architecture is first presented, followed by the embedded system block diagram, system hardware components, system software tools and thereafter the system PDL and flowcharts was presented. Lastly the Machine learning process and tools are explained.
**System Block Diagram**

The sensors collect real-time soil moisture and temperature readings, predictions are done on whether irrigation or fertilization is needed. In case irrigation is needed, irrigation valves are actuated, and the parameters observed till the required levels have been achieved. In case there is a need for fertilization, a notification is sent to the farmer via GSM on the amount of fertilizer to be added for the next irrigation cycle. The collected data is also sent to things speak cloud platform for future analytics and research. Figure 2 gives the embedded system block diagram showing the embedded components of the system.

![System Block Diagram](image)

**Fig. 2: System block diagram**

**System PDL**

A Program Description Language [PDL] was used to describe the algorithm for the working of the system as shown below;
System Hardware Components

The embedded system is made up of the following components:

- The Arduino Nano 33 BLE Sense which combines a tiny form factor, different environment sensors and the possibility to run AI using TinyML and TensorFlow Lite.
- The soil NPK sensor which is suitable for detecting the content of nitrogen, phosphorus, and potassium in the soil.
- The soil moisture sensor which consists of two probes that are used to measure the volumetric content of water.
- SIM900 GSM/GPRS shield for communication.
- The Micro DC 3-6V Micro Submersible Mini water pump.
- A 4-Channel Relay interface board that allows one to control and actuate multiple outputs.

AI Model Design

This section outlines the smart irrigation AI model. The first section describes the dataset used for training of the model followed by a description of the training process and thereafter a validation and testing of the model.

Datasets

Data used for training of the model was collected with the help of STES Group Rwanda for a
period of six months. The data was collected using soil moisture and temperature sensors distributed across a farm in Rwanda. With the existing dataset, thresholds are applied to irrigate the farm when stress conditions are detected. The dataset included 208,890 observations. The data had to first be cleaned to remove values that had errors also columns that were not needed for the training of the model were removed. Any duplicates were also checked and omitted.

Data Formatting

The dataset was first separated into two different sets, with the labels, irrigate and no irrigation. Each class had a total of 1290 observations. 20% of the data from each class was randomly separated as test data. Data was then converted from CSV to JSON for a convenient upload into the digital signal processing pipeline powered by edge impulse, an embedded ML training platform. In addition, 20% of the data is separated as validations set, leaving the remaining 60% as the training set. Figure 3 gives a plot of the data training data.

![Fig. 3: A plot of the no irrigation dataset.](image)

Model Training

A Neural Network classifier based on Keras and tensor flow light was used to train the model. A learning rate of 0.0005 with 300 training cycles was applied with 2 dense layers that were used with 20 and 10 neurons respectively. The code below describes the model architecture.

<table>
<thead>
<tr>
<th>Response</th>
<th>Irrigation</th>
<th>No irrigation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irrigation</td>
<td>96%</td>
<td>4%</td>
</tr>
<tr>
<td>No irrigation</td>
<td>18.2%</td>
<td>18.8%</td>
</tr>
<tr>
<td>F1 Score</td>
<td><strong>0.94</strong></td>
<td><strong>0.86</strong></td>
</tr>
</tbody>
</table>

Training Output

From the validation set, a training performance accuracy of 91.7% was achieved with a loss of 0.22. Table 1 shows the confusion matrix for the validation set.
The resulting model was also small enough to run on an embedded device with a projected RAM usage of 1.7K and a ROM requirement of 17.3K. The model is packaged into a tiny library that can run on an Arduino ARM cortex boards.

RESULTS, ANALYSIS AND DISCUSSIONS

In this section different results from the study are presented, analyzed and discussed. In the first section data collection sensor calibration is presented, the deep learning model evaluation results are outlined and discussed and lastly the results from the prototype presented.

Data Collection Sensor Calibration

The sensors deployed in the field by STES Group Rwanda for data collection were calibrated to verify the precision and reproducibility of sensor measurements. Sensors that are calibrated are the prerequisite for precise, reliable and reproducible measurement results. Calibration was done as it is one of the key prerequisites for effective quality assurance. Figure 4 shows the results during calibration for a section of the farm. The conditions in the soil were classified as shown in figure 4. The classified conditions include; Extreme stress condition, Stress condition, Field capacity, Excess condition. For the stress and extreme stress conditions, irrigation was needed.

Deep Learning Model Evaluation

So as to evaluate the performance of the deep learning model. The test data was used for inference on the cloud and on the embedded device and the results were compared so as to determine if optimization affects the model performance. The results were also compared with those best on open source benchmarking datasets and other machine learning algorithms.

Inference Results

From the inference results the model was able to predict the need or no need for irrigation with the same accuracies in both the cloud and the embedded device. This shows that the training of a deep learning model for inference in the resource constrained devices, do not affect the performance of the model.

Optimization Effects

From the evaluation on test data, the model performed with an accuracy of 93% slightly above that from the validation set which was 91%. However, when the model was quantized by conversion
from a float 32 model to an int 8 model. The resources needed on the device reduced while at the same time reducing the accuracy. This shows that in case of a big model one can opt to optimize so as to reduce the required resources. However, this will affect the accuracy and a trade-off must be considered.

**Data Set Performance Evaluation**

So as to evaluate the performance of the dataset, a comparison was done with an online open dataset from googled sets. The data was collected from a system with a wireless sensor network that sensed real-time soil moisture and temperature and ML applied to automatically control of an irrigation system. The results show that our datasets gave better accuracy, hence support the fact that the dataset used was better than existing model datasets in the areas of study. Figure 5 shows the accuracies achieved with different algorithms.

![Algorithm Accuracies](image)

**Fig. 5:** Accuracies achieved by using different algorithms

**Effect of Training Cycles on Accuracy**

An analysis was also done to find out how the number of training cycles affect the accuracy for the deep learning model.

![Accuracy as no of Epochs Increase](image)

**Fig. 6:** Results analysis for training cycles affects the accuracy.
The accuracy for the model increases as more training cycles are used. The best training accuracy was achieved at with 300 epochs. After this peak the accuracy did not increase further showing this is the optimum accuracy value.

**System Prototype Results**

All the components were connected and the functionality of the system tested in a lab environment. The soil moisture and temperature were captured in real-time and the collected values fed into the DL model that correctly predicted when to irrigate or not. Based on the prediction the water pump was actuated to irrigate the farm for a specified duration.

From the analysis our solution outperforms the existing solutions in the following ways. It has the capability of smart irrigation while at the same time monitoring soil parameters, Machine learning is used to predict when to irrigate, it is the only solution among those reviewed that uses deep learning, in other solutions either fuzzy logic or naïve Bayes and support vector machine are used. For all reviewed solutions that use machine learning only our solution applies it at the edge device. This makes our prototype appropriate for the African market where connectivity and energy are challenges. This also helps reduce on operation costs.

**CONCLUSION, RECOMMENDATIONS, AND FUTURE WORKS**

This study presents a prototype for a smart irrigation and fertilization system. The solution uses a deep learning model that has been packaged for deployment on an edge device. The edge based architecture for our solution makes it appropriate for the African market where existing cloud based solutions are difficult to deploy due to connectivity challenges. The deep learning model was tested in different environments and using different data sources, it was also compared against other algorithms applied on similar datasets. The results show that our model performed best with an accuracy of 91.7%. The results also show that the number of training epochs affects the model accuracy and optimization reduces the resources needed on the device to run a machine learning model but affects the accuracy of the model.

We recommend the implementation of this solution as it will indeed ensure the practice of data driven farming and thus help conserve the environment and ensure maximum utilization of resources. This will also lead to reduced operation costs and thus increase profits.

Future works will involve collection of more data so as to tune the performance of the model, development of a model for prediction of the needed fertilizers and the implementation and further testing of the solution.
REFERENCES


