Smart IoT Monitoring System for Agriculture with Predictive Analysis

Alaa Adel Araby[1], Mai Mohamed Abd Elhameed[1], Nada Mohamed Magdy[1], Loa'a Ahmed Said[1], Nada Abdelaal[1], Yomna Tarek Abd Allah[1], M. Saeed Darweesh[2,3], Mohamed Ali Fahim[4], Hassan Mostafa[1,3]

[1]*Faculty of Engineering, Cairo University, Giza, Egypt*

[2]*Electronics and Communications Department, Institute of Aviation Engineering and Technology, Giza, Egypt*

[3]*University of Science and Technology, Nanotechnology Program, Zewail City of Science and Technology, October Gardens, 6th of October, Giza 12578, Egypt*

[4]*Plant Protection, Central Laboratory for Agricultural Climate (CLAC), Giza, Egypt*

Abstract— **The Internet of Things (IoT) technology has the means to shape the future of many industries. Data is the language of the communication between different nodes through the network; the networks are the communication channel. The cloud is the home and destination of the data which adds intelligence through data analytics software, Precision agriculture uses the IoT features to help in managing crops production, by optimizing the quality of the crops through applying required nutrients and reduce the harmful impacts on the environment due to the application of excess pesticides. In this paper, we deployed a sensing network to gather the field data of some crops (Potatoes, Tomatoes, etc.), then fed these data to a machine learning algorithm to get a warning message finally displaying both the data and the warning message through a Graphical User Interface (GUI).**

Key words—Internet of Things (IoT); Precision Agriculture (PA); Machine Learning (ML); Message Queuing Telemetry Transport (MQTT.

I. INTRODUCTION

Precision agriculture is a new concept in agriculture, it is defined as the farm management system using information technology to identify, analyze and manage the variability of fields to ensure profitability, sustainability, and protection of the environment. In precision agriculture, new information technologies can be used to make better decisions about many aspects of crop production. It is obvious that precision agriculture increases the efficiency that can be realized by understanding and dealing with the natural variability found within a field [1].

Late Blight is a dangerous disease in agriculture field that affects many crops such as potatoes and tomatoes, it is identified by black/brown lesions on leaves and stems that may be small at first but as many lesions accumulate, the entire plant can be destroyed in only a few days after the first lesions are observed. The late blight caused the Irish potato famine of the 1840s which resulted in the death or emigration of over two million people from Ireland and it has been a continuous threat to potato cultivations in Egypt. The potato cultivation in Egypt occupying about 15% of the total vegetable cultivation area [2].

Machine learning is using precision agriculture benefits to prevent the late blight disease spreading, by using data collected about the climatological and soil conditions of the plants, the algorithm gives advice to the farmer regarding what should be done to prevent the infection as soon as possible.

IoT is also used to collect the sensors data from the field so that the data and the advice of ML algorithm can be available on a GUI platform, which makes it easier to have continuous monitoring of the field.

In [3] precision agriculture is used to provide agriculture solution using Artificial Neural Network machine learning algorithm which is used for performing data prediction on data collected by sensors. The use of IoT devices system provides an automated solution for data prediction. The produced result will be helpful for the farmer to take an accurate decision.

The system will give all prior knowledge in advance to the farmer for taking proper decision. The proposed system will be used to improve the detection of diseases and predict how the disease will spread in the crop field.

In this paper, we will focus on a certain disease which is a late blight for a certain crop which is potatoes using regression machine learning algorithm then give advice to the farmer to avoid it and prevent its spread as much as it possible.

The paper is organized as follows: In Section II, the IoT architecture of the system prototype is presented by illustrating how the data is collected then sent to the GUI application. Section III illustrates the machine learning algorithm which used and how it's applied to the sensors data. Section IV discusses the whole system including the results. The conclusion of this work is discussed in Section V.

II. CLOUD-BASED IOT ARCHITECTURE

The proposed cloud-based IoT platform architecture consists of several layers [4]. Fig. 1 describes the 3 layers of the system: the perception layer, network, and gateway layer, and application layer. The implementation of each layer will be discussed in detail through these subsections:

A. Perception Layer

The perception layer is the physical layer that contains the hardware of the sensing nodes on which the data is collected to be transmitted to the gateway layer. This layer is the front-end layer of the IoT system [5] and it consists of several nodes, each node consists of 3 main components: sensors, microcontroller, and a communication module.

Sensors: Which are used in precision agriculture to measure the different environmental attributes needed for the targeted application. Air temperature sensor, air humidity sensor, and soil moisture sensor are used to measure environmental and yield conditions. The prototype node collects data by interfacing these sensors with the microcontroller.

Microcontroller: it is responsible for collecting data measured by sensors. The microcontroller used is the Node Microcontroller Unit (NodeMCU) [6]. In this presented work, the platform used NodeMCU that is equipped with ESP8266 Wi-Fi modules. The NodeMCU can connect to the next layer, the gateway using a protocol called MQTT [7], [8].

Communication module and protocol: Wi-Fi module and MQTT protocol [9] are used to send the data collected to the gateway layer. This protocol which runs on TCP/IP connection uses publish/subscribe communication pattern, sending data from sensors attached to NodeMCU continuously to gateway layer is defined as the publisher so that this node is defined as publisher MQTT client. The MQTT client publishes the different data in a message-oriented where every message is published to a specific address called topic. To distinguish data, each sensor readings are published on a specific topic. The main distributor of the messages in the topics is a node that called an MQTT broker which is responsible for forwarding the messages between the sender and multiple receivers so that MQTT broker forward the topic message to subscriber MQTT client which is the next layer, the gateway layer.

B. Gateway Layer

The gateway layer is the bridge between the perception layer and the application layer. The different perception layer nodes deployed in the agricultural field collect the sensors data and send it to the gateway layer. The gateway is implemented using R-Pi 3 microcontroller [10]. It provides the needed processing power and storage that ensure that all the capture sensor data is relayed to the cloud server for analysis [7]. In Fig. 2, the architecture of the system shows that the R-Pi 3 is both the subscriber MQTT client and the broker. By installing mosquito software on R-Pi 3 which is a message broker that implements the MQTT protocol [11].

The R-Pi 3 is able to subscribe to the data from the same topics that the publisher MQTT client publishes in, the data is then analyzed and stored on R-Pi 3 to perform some machine learning algorithms which will be discussed in the next section to predict the disease or take an action depending on field condition, the predicted action, and the data are then being sent from R-Pi 3 to the application layer.

C. Application Layer

The Application Layer is responsible for connecting between the end user and the sensed data by dynamic application with data visualization. Data visualization is a technique in which the visual features can be used to code different attributes of data [12]. When data is visualized, it can be viewed by the end user. The end user has privileges to view and analysis the data without any modification to the sensed data. To design the back-end layer with the previous specification, the architecture is built via online database server as shown in Fig. 3.

The implemented back-end are based on an online MySQL

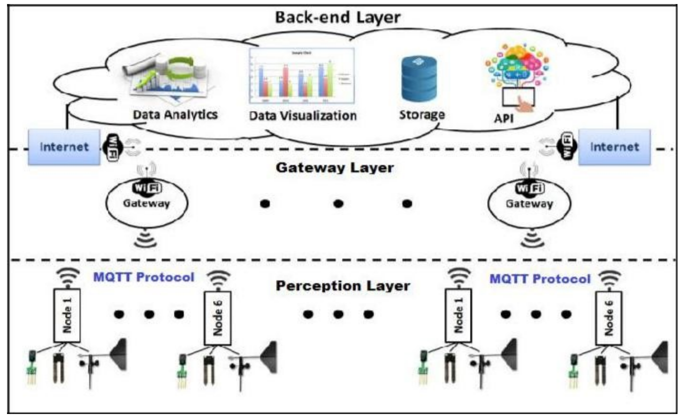

Fig. 1. Cloud-based IoT architecture

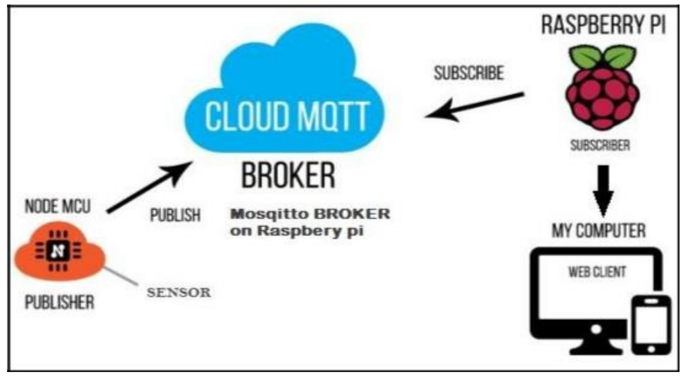

Fig. 2. System architecture communicating by MQTT

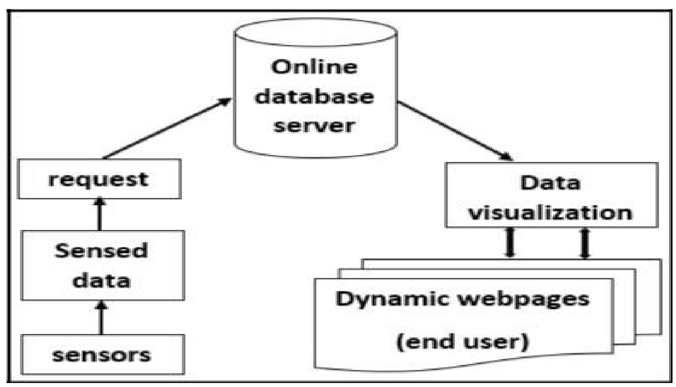

Fig. 3. Cloud server architecture

server through a free web hosting server (000webhost) which is receiving the data from the gateway by post method which receive the desired data with specific key which has an agreement from transmitter and receiver to extract the sensed data from the request and storing it to be accessed by the end user with a minimal downtime and without data corruption.

The website application consists of two separated web pages linked to the sensed data and the machine learning results gathered in one website to be viewed by the end user as shown in Fig. 4. The website is designed as a secure system by having a username and password for every client, linked to his data without accessing any other data.

III. Machine Learning Algorithms

A. Dataset and Proposed Models

The used dataset is real data, gathered from CLAC and is called a Day Degree data, it is shown in Fig. 5. GDD is the growing daily degree its calculation will be given in eqn. (1), Accu. GDD is the accumulative growing daily degree, DS is disease severity.

In this prototype, the Day Degree Data was used to build two different models, the first model classifies whether the first disease infection occurred or not for any general day, but that model was not generalized in this prototype as it cannot be used to apply protective procedures and the losses already occurred.

While the second model is built to predict the first occurrence of the disease eight days earlier than the first actual infection occurs, then a warning message is sent to the farmer.

B. Choosing the Correct Algorithm

To achieve the required output in the two previously discussed models, two steps must be done, the first step is to study the data to remove any redundant features for example (IPS) which is in all the dataset is zero.

The second step is to choose the correct algorithm, some different algorithms were tested such as support vector machine (SVM) and logistic regression (LR) for the classification model, linear regression and neural networks for the prediction model, the comparison between the different algorithms was with respect to the achieved accuracy versus the training time, the number of samples and the number of features.

For the prediction model, theoretically neural networks are preferred to be used when the number of the training samples exceeds 5000 or when the dataset has an enormous number of features, but in order to achieve a good accuracy, the trade-off is the time and the memory consumed in training the model.

In general, neural networks consumes more time to be trained than the regression. So, the algorithm chosen for the prediction model was linear regression with different orders, so as to save both time and memory as all the processing is performed on the gateway [13].

For the classification model, the problem is a simple two classes' model to decide if an infection occurred or not; so, both algorithms can be applied. Theoretically, the difference between the two algorithms is in the output calculation.

In the LR algorithm, the output expresses the probability of one of the two cases to occur which takes a value from 0 to 1, on the other hand, the SVM algorithm's decides immediately which case occurred; so the output is 0 or 1. As a conclusion if the output is required to be determined, choose the SVM, but if the output is needed to be slightly smooth, choose the LR, finally after testing both algorithms, the SVM algorithm reached higher accuracy than the LR by difference around 2%, so SVM was chosen for the classification model [13].

C. The Prediction Model

In the training model, the input is the Accu. GDD, but in the on-field model, the data is based on the weather conditions measured from the field; so, it needs to be adapted

Fig. 4. The home page of the website

Accu. GDD	IPS	DS
7.25	0	0
15.35	0	0
24.6	0	0
32.8	0	0
41.4	0	0
49.4	0	0
57.4	0	0
64.3	0	0
73.4	0	0
82.5	0	0
91.85	0	2
100.75	0	3.6
109.7	0	5.2

Fig. 5. A Sample of the dataset from CLAC

to match training algorithm by calculating GDD using eqn. (1) [2].

$$GDD = \frac{T_{Max} + T_{Min}}{2} - 10^o\ C \quad (1)$$

Where T_{Max} is the highest temperature measured daily.
T_{Min} is the lowest temperature measured daily.
The output from the model is the DS which predicts the percentage of crop infection and based on its value, the warning message is prepared.

IV. Results and Discussion

A. Prototype Discussion

A prototype of the proposed architecture for IoT precision agriculture applications has been implemented. Sensor nodes mentioned in the perception layer are deployed and tested. The nodes send data to a single Gateway using the MQTT protocol. Nodes act as a publisher MQTT client and initiate the connection by sending CONNECT message to the MQTT broker. The broker then responds with CONNACK to the client. Once the connection is established, the client keeps sending the sensor readings in the specified topics to the broker. The broker keeps the connection open as long as the client doesn't send a disconnect command or lose the connection. The Subscriber MQTT client which is the gateway subscribe the data which sent from the same topics. Then data are analyzed and stored on the gateway device to apply the machine learning algorithm code and predict the suitable action. The data and the warning action message are visualized and displayed on the website as a GUI which facilitate the user interfacing to keep the farmer monitoring the field.

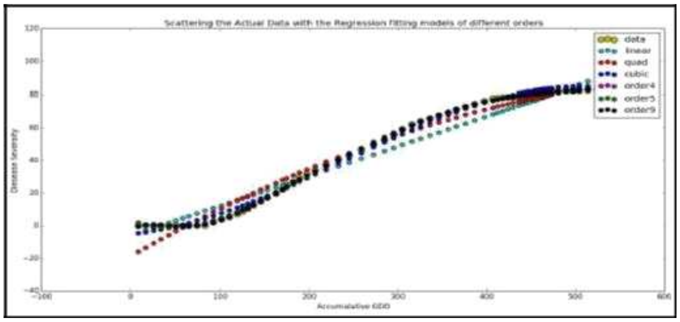

Fig. 6. Potato DS vs Accu. GDD Fitting curves

```
In [24]: runfile('C:/Users/loloa/Desktop/finalnada/ml1.py',
wdir='C:/Users/loloa/Desktop/finalnada')
('Accuarcy of Regression of order 1-', 95.21616681084241)
('Accuarcy of Regression of order 2-', 97.635842621167114)
('Accuarcy of Regression of order 3-', 99.469916822607985)
('Accuarcy of Regression of order 4-', 99.932565796407943)
('Accuarcy of Regression of order 5=', 99.93415900286206)
('Accuarcy of Regression of order 9=', 99.975168452128457)
```

Fig. 7. Potato DS Different order of regressions accuracy

```
In [5]: runfile('E:/FinalWork_new/Regression_potato_tomato_Training_Models/advicetry.py',
wdir='E:/FinalWork_new/Regression_potato_tomato_Training_Models')
The Accumalative GDD is = 12 so all is clear for now
This is an advisory message: The Current Accumalative GDD = 137 approched the thershold
value for a regular unepidemic session therefor you have an approximatly from 10 to 14 days
to apply the protective fungisides to prevent the spreading of the disease
This is an advisory message: note the current Accumalative GDD = 254 so starting from
approximately 5 days of today the disease spreading will become linear with the accumaltive
GDD which will cause sevier losses in the yield if no action is taken
Warning !! The Accumalative GDD is = 397 so if no action was taken during the previous days
THERE WILL BE an EPIDEMIC SPREAD
```

Fig. 8. Warning Messages for different days in the infection season

B. Results

In Fig. 6, the model fitting curves which are obtained for several orders after the training process are shown.

The accuracy of each curve is calculated using the mean square error. The accuracy is based on comparing the predicted values obtained from the model to the mean values expected. It was found that as the order increases the accuracy increases to obtain 99.97% at the ninth order as shown in Fig. 7.

The results are then given to the user based on the accumulative GDD thresholds. In Fig. 8, there are four messages which can be sent to the farmer in order to advise him about the proper action that could be taken. If the accumulative GDD<125, it means the land is clear from any diseases. For the GDD range between 125 and 250, it approaches the regular un-epidemic season, therefore, the farmer should apply fungicides within 10-14 days. While in the GDD range between 250 and 390, a linear spread of the disease is expected to happen within 5 days and farmer must take an action to protect his crop. Finally, if the accumulative GDD exceeded 390, it means an epidemic spread occurs and the crop was already lost [2].

V. Conclusion

This paper proposes, a smart system based on the integration between IoT and machine learning to predict the late blight disease in potatoes and tomatoes before the first occurrence which reduces the costs by giving the farmer an exact warning message on the specific time to apply the protective pesticides which help to save the yield production in the infection seasons and reduce the usage of the unnecessary pesticides.

Each sensing node contains some low-cost sensors to collect the factors needed for the machine learning model which are: temperature, relative humidity, leaf wetness, and the soil moisture. These data are collected by nodes deployed in the field to be sent using the Wi-Fi module and MQTT protocol to the gateway on which the machine learning algorithm runs to output the predicted warning message. The predicted action or warning message is displayed on the website for the farmer.

In the future work the system can be expanded to be fully automatically actuating as the proposed system in this paper needed human interference to apply the action but, in the future, the gateway can send actions by itself and spray the field when needed.

Acknowledgment

This research was partially funded by the Information Technology Industry Development Agency (ITIDA), National Telecommunications Regulatory Authority (NTRA), and ONE Lab at Cairo University, Zewail City of Science and Technology.

References

[1] B. S, "Precision Agriculture: Tomorrow's Technology for Today's Farmer," in *Journal of Food Processing & Technology*, vol. 06, no. 08, p. 1, 2015.

[2] Fahim, M. A., M. K. Hassanien, and M. H. Mostafa. "Relationships between climatic conditions and Potato Late Blight epidemic in Egypt during winter seasons 1999–2001," in *Applied Ecology and Environmental Research*, vol. 1, no. 1, pp. 159-172, 2003.

[3] A. Chandgude, N. Harpale, D. Jadhav, P. Pawar and S. M. Pati, "A Review on Machine Learning Algorithm Used for Crop Monitoring System in Agriculture," in *International Research Journal of Engineering and Technology (IRJET)*, vol. 05, no. 04, pp. 1470, 2018.

[4] K. Yelmarthi, A. Abdelgawad and A. Khattab, "An architectural framework for low-power IoT applications," in *28th International Conference on Microelectronics (ICM)*, pp. 373-376, 2016.

[5] A. Khattab, A. Abdelgawad and K. Yelmarthi, "Design and implementation of a cloud-based IoT scheme for precision agriculture," in *28th International Conference on Microelectronics (ICM)*, pp. 201-204, 2016.

[6] M. Yuan, "Getting to know NodeMCU and its DEVKIT board," Ibm.com. [Online]. Available: https://www.ibm.com/developerworks/library/iot-nodemcu-open-why-use/index.html, 2018.

[7] K. O. Flores, I. M. Butaslac, J. E. M. Gonzales, S. M. G. Dumlao and R. S. J. Reyes, "Precision agriculture monitoring system using wireless sensor network and Raspberry Pi local server," in *IEEE Region 10 Conference (TENCON)*, Singapore, pp. 3021, 2016.

[8] A. Glória, F. Cercas, and N. Souto, "Design and implementation of an IoT gateway to create smart environments," in *Procedia Computer Science*, vol. 109, pp. 568–575, 2017.

[9] K. Grgic, I. Špeh and I. Hedi, "A web-based IoT solution for monitoring data using MQTT protocol," in *International Conference on Smart Systems and Technologies (SST)*, Osijek, pp. 249-253, 2016.

[10] R. P. Foundation, "Teach, Learn, and Make with Raspberry Pi," Raspberry Pi. [Online]. Available: https://www.raspberrypi.org/, 2018.

[11] Eclipse Mosquitto. [online] Available at: http://mosquitto.org/, 2018.

[12] I. M. Tolentino, R. S. Juson, B. U. Tan, and M. R. Talampas, "Design, development, and evaluation of a simple wireless sensor network for indoor microclimate monitoring," in *IEEE Region 10 Conference TENCON 2010*, 2010.

[13] "How to choose machine learning algorithms," Docs.microsoft.com, [Online]. Available: https://docs.microsoft.com/en-us/azure/machine-learning/studio/algorithm-choice, 2018.