IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 24, 2019, accepted August 6, 2019, date of publication August 15, 2019, date of current version September 3, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2935564

Applying Big Data for Intelligent Agriculture-Based Crop Selection Analysis

FAN-HSUN TSENG[1], (Member, IEEE), HSIN-HUNG CHO[2], (Member, IEEE), AND HSIN-TE WU[3]

[1]Department of Technology Application and Human Resource Development, National Taiwan Normal University, Taipei 106, Taiwan
[2]Department of Computer Science and Information Engineering, National Ilan University, Yilan 260, Taiwan
[3]Department of Computer Science and Information Engineering, National Penghu University of Science and Technology, Penghu 880, Taiwan

Corresponding author: Hsin-Te Wu (wuhsinte@gms.npu.edu.tw)

This work was financially supported from the Young Scholar Fellowship Program by Ministry of Science and Technology (MOST) in Taiwan, under Grant MOST108-2636-E-003-001, and was partly funded by the MOST in Taiwan, under grant MOST108-2221-E-197-012-MY3.

ABSTRACT With the growth of the human population comes the constantly rising demand for agricultural products. Nevertheless, as the world experiences climate change, many crops are often damaged by weather conditions.This study utilizes Intelligent Agriculture IoT equipment to monitor the environmental factors on a farm. The collected data underwent 3D cluster analysis to yield analysis of the environmental factors of that farm. The proposed scheme bears the following features: (1) data normalization is achieved via the combination of moving average and average variance; (2) we applied 3D cluster analysis to analyze the relation between environmental factors and subsequently examine the rules of thumb held by the farmers; (3) the system determines whether a selected crop has been placed in the appropriate cluster; and (4) the system sets a critical value in the cluster based on future environments and provides advice on whether a crop is suitable for the farm. We placed Intelligent Agriculture IoT equipment in the farm for monitoring purposes and ran an actual-scenario analysis using the algorithm in our study; results confirm that our proposed scheme is indeed feasible.

INDEX TERMS Big data, intelligent agriculture, internet of things, agricultural engineering, data mining.

I. INTRODUCTION

With the growth of the human population comes the constantly rising demand for agricultural products. According to study [1], the global population is expected to rise from 1.8 billion in 2009 to 4.9 billion in 2030, leading to drastic rise in demand for dairy products. Study [2] suggests that, in the future, human beings will have a growing demand for agricultural products, which will require expansion of farm lands and growth in yield of agricultural products. Meanwhile, due to global warming, crops are often damaged by extreme weather conditions. Study [3] points out that countries around the globe are investing in the development of Intelligent Agriculture out of concern for food crisis. For instance, as mentioned in study [4], many farms have begun to rely more heavily on natural resources – such as utilizing hydropower, geothermal energy, or solar power – in order to reduce cultivation costs, especially in water resources. Study [5] argues that lack of labor force will become a serious issue in the next few decades. Currently, the agricultural labor force constitutes mainly of elder people who operate on experience and rules of thumb; moreover, they lack the adequate equipment to monitor environmental factors on their farm. Consequently, farmers have no way of knowing which crops can grow on their soil, especially since the global climate change has deemed many crops to be unsuitable for long-term cultivation. Given the above, we must utilize the Internet of Things and big data to analyze farm soil and environment in order to provide farmers with suitable crop options and keep track of their cultivation techniques as well as environmental factors at their farms.

Study [6] argues that we must establish an agricultural ecosystem if we want to combat food crisis. The system should be used to monitor farm data that can be used to improve production issues. On the other hand, study [7] suggests that our future world will acquire 2 more billion people in population, which will not only render our residential spaces smaller, but also prohibit farm lands from expanding. Therefore, we must rely on Intelligent Agriculture to elevate the productivity of farms. In addition to monitoring

The associate editor coordinating the review of this article and approving it for publication was Mu-Yen Chen.

environmental factors on a farm, Intelligent Agriculture must analyze how weather conditions cause environmental changes at the farm and how long-term crop cultivation brings about soil erosion or changes in the soil structure. Monitoring a farm using big data and IoT can prevent low productivity of crops. More often than not, when it comes to cultivation techniques, farmers rely on rules of thumb; however, as we experience climate change, crops that grew well in the past might not be able to adapt to present farm conditions. Meanwhile, farmers have no way of learning about the condition of their farm soil, which is why they should adopt Intelligent Agriculture systems to monitor their farm environment. Since many farms are small, outdoor farms, the Intelligent Agriculture system must be financially affordable and offer green power storage function. Monitoring farm data via the Intelligent Agriculture system facilitates big data analysis and helps us understand the environmental factors and soil structure of a farm.

This study proposes big data analysis of farms based on Intelligent Agriculture. Our Intelligent Agriculture platform provides the following features: (1) IoT sensors for temperature, humidity, atmospheric pressure, illumination, electrical conductivity of soil, and irrigation, and the system will send the data back to the platform regularly every 10 minutes; (2) a solar power storage system that helps reduce electricity costs; and (3) an XMPP web platform that helps collect information from the IoT sensors. The proposed Intelligent Agriculture system employs affordable sensors, which facilitates the popularization of Intelligent Agriculture systems; furthermore, the solar power storage system enables the mobility of the system and makes it suitable for outdoor farms. This study conducts big data analysis on the information collected from the Intelligent Agriculture system. The analysis goals include: (1) analyzing cultivation techniques practiced by the farmers, understanding how the crops are growing, and examining environmental changes; (2) analyzing environmental factors of a farm and choosing suitable crops for cultivation. This study investigates experiences of farmers and their cultivation techniques through behavior analysis to examine information such as irrigation pattern and how much water the soil needs. Additionally, the study analyzes environmental factors of the farm to further understand how weather conditions bring about environmental changes and determines which crops are suitable for the farm based on time sequence. Our proposed big data analysis approach includes the following concepts: (1) data normalization via the combination of moving average and average variance; (2) application of 3D cluster analysis to analyze the relation between different environmental factors and subsequently examine the rules of thumb held by the farmers; (3) determination of whether a selected crop has been placed in the appropriate cluster; and (4) setting a critical value in the cluster based on future environments and providing advice on whether a crop is suitable for the farm. Our proposed big data analysis method has undergone simulation and helped analyze whether the locally grown crop is a suitable choice; experiment results show that the proposed scheme is not only feasible but also helps farmers understand their farm environmental through the environmental indices of their farms.

II. RELATED WORKS

Study [8] proposes a system framework that combines Intelligent Agriculture and big data. It utilizes cloud technology to store IoT sensor data, which are then used in big data analysis for farm management. Prior to analysis, the big data must undergo data normalization through procedures such as data cleaning and normalization; the farm may utilize big data analysis for prediction of crop growth and environmental changes and even provide the information to food processing companies for authentication or product traceability. Study [6] provides a literature review that illustrates how past collaborations between big data and IoT networks have brought immense impact to agriculture and farming; many of the researches mentioned in the study targeted analysis of farm water resources and helped reduce the waste of water resources. In study [9], it is illustrated that because large-scale farms lack adequate power equipment, network transmission of sensors become a huge issue, and drones are hence used for collection of sensor data. However, data collection performed in the air faces signal interference problems that require multi-level data cleaning procedure before obtaining comprehensive sensor data. On a different approach, study [10] uses big data to analyze irrigation locations in agricultural practices. The study proposes an agent method for determining the optimal irrigation location in a cropland. Given that many IoT networks need to transmit data back to a remote server and yet the remote server may be unable to respond instantaneously due to network issues, the study proposes an agent mechanism that allows for communication between IoT networks for monitoring irrigation conditions. Meanwhile, study [11] applies Kalman filtering to evaluate the data of a given area and provide further estimate of future data. The study makes use of fuzzy neural networks to elevate the rate of convergence and forecast accuracy. It relies on the collected data on environmental factors to predict future environmental changes on the farm and facilitate temperature control for indoor farms.

Study [12] investigates pig diseases in the livestock industry by analyzing mean values and similarities in the context of seasonal changes. The study applies big data analysis to examine the health condition of pigs in 44 different farms to provide insight into disease prevention and trends. On a different topic, study [9] suggests setting up not only IoT equipment on the farm but also small weather stations to monitor local weather and atmospheric changes. Since farm areas often suffer from poor wireless network connection, transmission of vast image data often results in packet loss; in consideration of this, the study employs all kinds of image compression techniques in their analysis to effectively improve data compression volume and minimize image distortion. On the topic of data preservation, study [13] believes that data preservation can be optimized by combining Intelligent Agriculture and big data and using cloud technology for data storage. Additionally, it employs General Packet Radio Service and Global Positioning System for data transmission and sensor positioning. Also discussing the adoption of cloud service, study [14] suggests that applying cloud technology in IoT data storage can effectively enhance data integrity and security. In the intelligent healthcare system mentioned in study [15], aside from monitoring the physical conditions of a patient and running other related applications, the system utilizes big data to assess relevant data for prevention of certain diseases that in turn reduce medical burden. Study [16] uses big data to help optimize power distribution in Volt-VAR Control (VVC) processes in wind power generation. The VVC centers can be either a centralized VVC, a distributed VVC, or a hierarchical VVC; the study focuses on how to combine the main power source with wind power generation to achieve the best outcome.

Study [17] advocates a transmission structure involving IoT and big data. An IoT network shares its data across multiple systems; for instance, the monitoring platform needs instantaneous warnings while each mechanical platform reads data, performs operations, and runs big data analysis. Meanwhile, study [18] proposes using a hybrid weighted geometric averaging operator to compute group decision-making problems. Multiple group decision-making problems have been an emphasized issue in the field of decision-making studies; cluster mechanisms are also commonly used in different fields. The swift convergence and critical value issues related to clusters can be improved by adopting the proposed approach in study [18]. Study [19] offers a clustering decision-making system with a feedback mechanism; the study utilizes the feedback mechanism and threshold to rank, by order, the optimal investor. Study [20] provides a probabilistic hesitant multiplicative preference relations model in which the decision-maker will assess and grade alternative product options before utilizing the group decision-making model to identify the optimal alternative product. The proposed Intelligent Agriculture platform in our study uses 4G network for data transmission to ensure the quality of network transmission. We employ Extensible Markup Language (XML) for data transmission, which allows for data transmission across different platforms or company labels. Our study adopts cluster technology to run farm data analysis and combines critical values using the threshold approach to determine whether a certain crop is suitable for cultivation at a specific farm. Our proposed mechanism can help farmers obtain a firm grasp of their farm environment and further improve the quantity and quality of their farm products.

III. THE PROPOSED SCHEME

In this paper, Section 3.1 illustrates the system model, Section 3.2 discusses data cleaning and normalization, Section 3.3 involves 3D cluster correlation analysis, while Section 3.4 elaborates on crop selection analysis and decision-making.

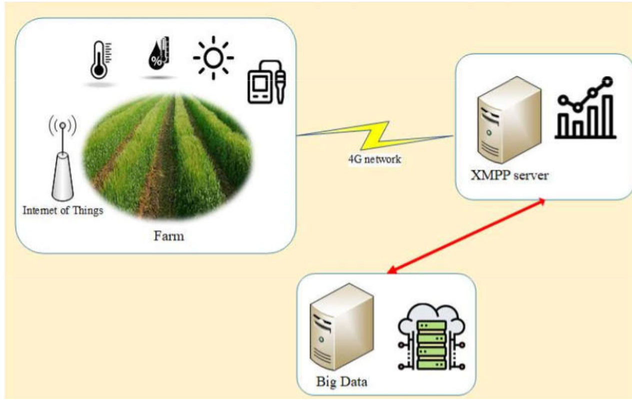


FIGURE 1. System illustration.

A. SYSTEM MODEL

Our proposed scheme requires establishing an Intelligent Agriculture platform as shown in Figure 1. We established IoT sensors in farms to monitor the farm environment; the sensor equipment can help detect temperature, humidity, illumination, atmospheric pressure, soil electrical conductivity (EC), soil moisture content, and soil salinity. We can learn about soil fertility conditions through the EC and can utilize the sensors to determine irrigation time and quantity. Our system calls for IoT sensors that are financially affordable, which should help popularize its adoption. As for IoT power consumption, our system employs solar power storage, which not only frees the system from the need for external power support but also gives it physical mobility. The data of our IoT sensors are transmitted to the server via 4G networks in XML format; this helps achieve information integration across different platforms and formats. We have also included charts and graphs in our XMPP platform to aid data presentation of each sensor; among the features, the charts and graphs can be exported into reports. Data from all sensors can be exported from the database to undergo big data analysis. The goals of our big data analysis include: (1) analyzing cultivation techniques practiced by the farmers, understanding how the crops are growing, and observing environmental changes, and (2) analyzing the environmental factors of the farm and selecting suitable crops. Big data defines the 5V features as [22]: (1) volume, (2) velocity, (3) variety, (4) veracity, and (5) value. Our proposed big data approach satisfies the 5V definitions. The data in our analysis comes from IoT sensor data from the farm. In our scheme, the IoT sensors acquire data and send it back to the platform every 10 minutes. Our analysis offers data variety by obtaining relevant information from weather stations. Before analyzing any data, our system performs data cleaning and normalization to ensure that both the analyzed data and the analysis result of predetermined targets are accurate. Our proposed approach and goals can effectively increase crop production and help analyze cultivation techniques of farmers.

B. DATA CLEANING AND NORMALIZATION

The first step in big data analysis is performing data cleaning and normalization. Our system uses 4G networks for IoT data transmission; while this ensures stable network quality, the system might still run into issues such as packet loss or poor signal might. The big data first undergoes data cleaning; since farm environments usually fluctuate in the pattern of linear increase or decrease, it rarely varies drastically. The symbols used in our study are illustrated in Table 1. We first compute the moving average with the Algorithm 1.

Algorithm 1 Moving Average Algorithm

Compute the data of each sensor
for each $i = [1, m]$ **do**
  Compute the moving average of the sensor data
  for each $j = [1, r]$ **do**
    $\mathcal{MA}_{D_{i,j}} = \frac{\sum_{j}^{j+n} \mathbb{D}_{i,j}}{n}$
  end for
end for

Once obtaining the moving average, the system proceeds to compute the half-hard threshold. We use the variance to compute the upper/lower limit of the moving average. The algorithm is as Algorithm 2.

Algorithm 2 The Upper/Lower Limit of the Moving Average

Compute the data of each sensor
for each $i = [1, m]$ **do**
  Compute the moving average of the sensor data
  for each $j = [1, r]$ **do**
    $\mathcal{MA}_{D_{i,j}} = \frac{\sum_{j}^{j+n} (\mathbb{D}_{i,j} - u)^2}{n}$
  end for
end for

When the system has identified the moving average and variance of each sensor, any data that falls outside the upper/lower limit indicates variation that is too drastic and is thus removed as an anomaly; any data that falls within the upper/lower limit is preserved and undergoes the following computation:

$$\mathbb{D}'_{i,j} = \left\{ \begin{array}{l} \mathbb{D}_{i,j}, \mathbb{D}_{i,j} <= \mathcal{MA}_{D_{i,j}} + \sigma^2 \left(\mathbb{D}_{i,j}\right) \\ \text{ and } \mathbb{D}_{i,j} >= \mathcal{MA}_{D_{i,j}} - \sigma^2 \left(\mathbb{D}_{i,j}\right) \\ none, otherwise \end{array} \right\} \tag{1}$$

When the data cleaning procedure is completed, the data then undergoes normalization. For instance, given a temperature of 33.6 degrees, the system rounds off the decimal point to the integer place unconditionally because the decimal points in a temperature value will not affect the results in big data analysis. Our system also conducts normalization computation based on the accuracy of each sensor using the equation: $\hat{\mathbb{D}}_{i,j} = \text{Norm}\left(\mathbb{D}'_{i,j}\right)$

TABLE 1. Symbols used in this study.

Symbol	Meaning of the symbol
$\mathbb{D}_{i,j}$	Sensor data.
$\mathcal{MA}_{D_{i,j}}$	The moving average of the sensor data.
$\mathbb{D}'_{i,j}$	Valid sensor data.
σ^2	Variance calculation.
σ	Standard deviation calculation.
u	Mean value calculation.
Norm	Normalization calculation.
$\hat{\mathbb{D}}_{i,j}$	Value of post-normalization sensor data.
$\mathcal{CORR}()$	Autocorrelation function calculation.
Max	Maximum value calculation.
$\mathcal{CY}_i$	Maximum value of each day/week.
$\mathcal{RE}$	Periodical determination.
$\overline{\hat{\mathbb{D}}_{i,l,t}}$	Average cycle time duration.
$\overline{\hat{\mathbb{D}}_{i,j}}$	Average soil moisture content.
$\overline{\hat{\mathbb{D}}_{i,j,x}}, \hat{\mathbb{D}}_{i,j,y}$	The x and y axis values of the sensor data.
$\mathcal{SL}_i$	Determining the slope of the EC value of the soil.
$\mathcal{RSL}$	Determining whether the soil has a rising EC value.
$\mathbb{G}'_{i,j}$	Sensor cluster data.
$\mathcal{REC}$	Analysis results of determining whether it is an irrigation behavior.
$\mathcal{CR}_{i,min}$	Minimum value of cultivation environmental factors of the crops.
$\mathcal{CR}_{i,max}$	Maximum value of cultivation environmental factors of the crops.
$f()$	Normal distribution calculation.
$\mathbb{CRE}_i$	Determine whether it suits the farm environment.

C. 3D CLUSTER CORRELATION ANALYSIS

Our proposed system utilizes 3D correlation analysis to examine cultivation techniques of farmers. First, we calculate the irrigation cycle practiced by the farmers by applying autocorrelation to obtain the periodicity. The algorithm is as follows:

$$\mathcal{CORR}\left(i, d_k\right) = \prod_{j}^{n} \hat{\mathbb{D}}_{i,j} \tag{2}$$

In which d_k stands for the autocorrelation results of each day or each week. Following that, the system identifies the maximum value d_k and uses it as the cycle value. The computation is as follows:

$$\mathcal{CY}_i = \text{Max}\left(d_k\right) \tag{3}$$

In which $\mathcal{CY}_i$ stands for the maximum value taken to be the cycle for each day or week. Then, the system identifies whether the farmer irrigates daily or weekly. Our system begins by calculating the daily soil moisture content for each day. It examines whether the variance of $\mathcal{CY}_i$ is greater than the threshold value; if yes, then it is an indication that the periodicity is unstable, so the system reruns the calculation based on a weekly basis. The calculation for the variance of $\mathcal{CY}_i$ is shown in Equation 4; the determination of the

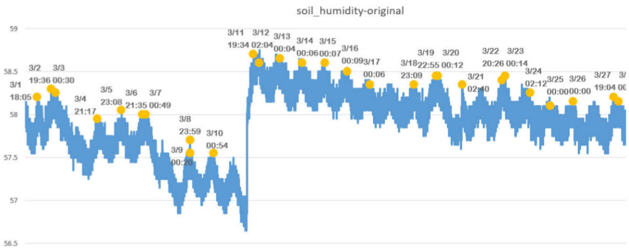


FIGURE 2. Autocorrelation results.

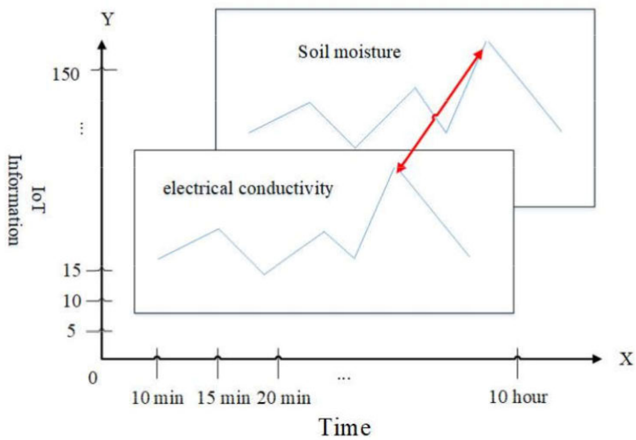


FIGURE 3. An illustration of 3D correlation analysis.

periodicity of a cycle relies on Equation 5.

$$\sigma^2(\mathcal{CY}_i) = \frac{\sum_{n=1}^{l} (\mathcal{CY}_n - u)^2}{l} \tag{4}$$

$$\mathcal{RE} = \begin{cases} True, & \sigma^2(\mathcal{CY}_l) < threshold \\ flase, & otherwise \end{cases} \tag{5}$$

When $\mathcal{RE}$ is flase, the system returns to Equation 1 and the data range is readjusted to a weekly basis. When RE is true, the system calculates the average cycle time $\hat{\mathbb{D}}_{i,l,t}$, in which t stands for the cycle time while $\hat{\mathbb{D}}_{i,j}$ stands for the average soil moisture content. Figure 2 is an illustration of our system's autocorrelation computation results.

This study analyzes and determines whether the farmer has performed tasks such as applying fertilizers or pesticides. As shown in Figure 3, after identifying the biggest ascending curve of the soil, our proposed system maps the result onto the monitoring of soil moisture content to determine whether it is the maximum cycle value; if not, it indicates that the farmer has performed application of fertilizer or pesticide. We first apply the slope to calculate how the electrical conductivity (EC) of the soil is ascending using the following equation:

$$\mathcal{SL}_l = \frac{\hat{\mathbb{D}}_{i,1,y} - \hat{\mathbb{D}}'_{i,j,y}}{\hat{\mathbb{D}}'_{i,1,x} - \hat{\mathbb{D}}'_{i,j,x}} j = 2, \ldots, m \tag{6}$$

Next, we determine whether $\mathcal{SL}_i$ is greater than the threshold value, which would signify changes in soil environment and significant rise in EC value. The following equation is applied:

$$\mathcal{RSL} = \begin{cases} true, & \mathcal{SL}_l > threshold \\ flase, & otherwise \end{cases} \tag{7}$$

When the resulting $\mathcal{RSL}$ is rendered true, it indicates that the soil EC has increased significantly. Following that, the system computes whether the results maps onto the maximum moisture sensor value; if yes, then it means that the soil EC value was raised by irrigation; if not, then it indicates no irrigation. The equation is as follows:

$$\mathcal{REC} = \begin{cases} true, & \hat{\mathbb{D}}'_{EC,j} \rightarrow \mathcal{CY}_l \exists Max\left(\hat{\mathbb{D}}'_{Moisture,j}\right) \\ flase, & otherwise \end{cases} \tag{8}$$

When $\mathcal{REC}$ is flasefalse, it means that the farmer has applied fertilizer or pesticide. Our proposed scheme utilizes EC and water sensors to conduct analysis on the cultivation techniques of farmers and gain insight into growth factors in the soil, all in the hopes of aiding newcomer future farmers in the future taker over the farm and carry on the cultivation.

D. CROP SELECTION ANALYSIS AND DECISION-MAKING

This study divided sensor data into clusters based on their numerical size, which resulted in 7 clusters: temperature, air humidity, atmospheric pressure, soil moisture content, soil electrical conductivity, illumination, and soil salinity. The sensor clustering data is $\mathbb{G}'_{i,j}$. Figure 4 is an illustration of the clustering; the x axis illustrates a far-to-near time order while the y axis shows a small-to-big numerical order of the values. Next, the system places into in order the necessary conditions as well as other conditions to select feasible suitable crops. The necessary conditions include temperature, soil electrical conductivity, and soil salinity; meanwhile, other conditions may include atmospheric pressure, soil moisture content, air humidity, and illumination. For instance, soil moisture content and air humidity can be improved using the irrigation system and; illumination can also be improved using lighting equipment. After completing all the above, the system determines whether a crop is suitable for the farm using the following equation:

$$\mathbb{CRE}_i = \begin{cases} true, & \mathbb{G}'_{i,\max} < \mathcal{CR}_{i,\max} \text{ and } \mathbb{G}'_{i,\min} > \mathcal{CR}_{i,\min} \\ flase, & otherwise \end{cases} \tag{9}$$

Equation 9 determines whether each value of the crop renders it suitable for the farm environment, in which $\mathbb{G}'_{i,\max}$ represents the maximum sensor data value, $\mathbb{G}'_{i,\min}$ represents the minimum sensor data value, $\mathcal{CR}_{i,min}$ represents the minimum crop environment value, and $\mathcal{CR}_{i,max}$ represents the maximum crop environment value. If the system deems other conditions to be unsuitable, it provides the assessment result to the farmer so that they may decide for themselves

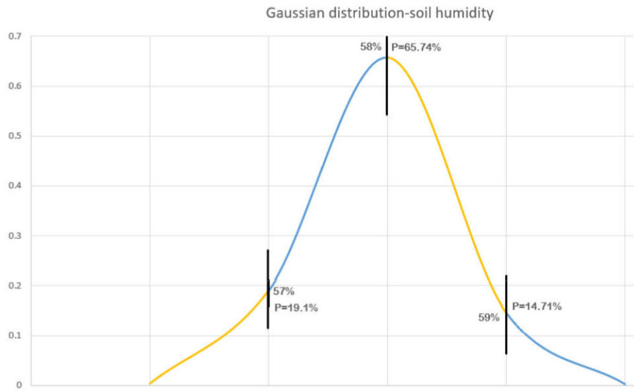


FIGURE 4. An illustration of the clusters.

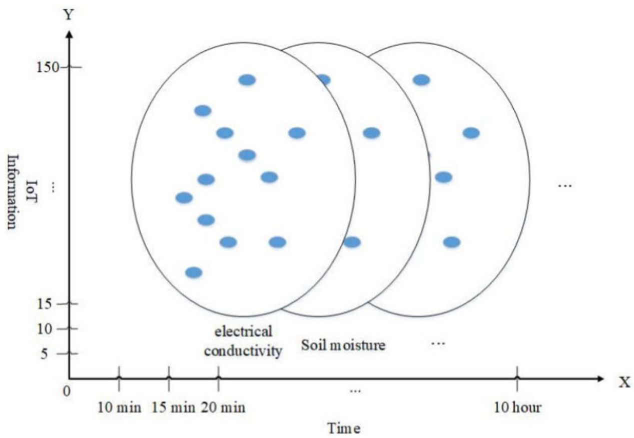


FIGURE 5. Risk assessment of crops.

whether they wish to proceed with the cultivation. If the crop is suitable for the farm environment, the system must take into consideration weather condition impacts; hence, our study utilizes normal distribution calculation to determine which area the crop falls under. As shown in Figure 5, if it falls within 2 standard deviation (2σ), then the system will advise the farmer to take cultivation risks into account. The computation is as follows:

$$f(\mathcal{CR}_{i,\max ormin}) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(\mathcal{CR}_{i,\max ormin}-u)^2}{2\sigma^2}} \tag{10}$$

Upon obtaining $f(\mathcal{CR}_{i,\max ormin})$, the system applies the lookup table method to determine which area it falls under and provides the result to farmer so they may assess the risks of cultivating that crop.

IV. PERFORMANCE

This chapter elaborates on our Intelligent Agriculture experiment results (Section 4.1) and big data experiment results (Section 4.2).

A. INTELLIGENT AGRICULTURE EXPERIMENT RESULTS

The hardware and software used in our Intelligent Agriculture platform are shown in Table 2. Figure 6, on the other hand, demonstrates our experimental farm and experiment equipment. The farm is 0.19835 hectares in size; the IoT equipment can detect air temperature, humidity, atmospheric pressure, soil moisture content, soil electrical conductivity, illumination, and soil salinity. Figure 7 shows data acquired by the IoT sensors; under our proposed scheme, the IoT collects farm environment data every 10 minutes. Meanwhile, our proposed system runs on power in the form of solar power storage, enabling the equipment from being moved to any corner of the farm at any time. The IoT equipment uses 4G networks for packet transmission; the XMPP platform receives the IoT data and then presents the data through charts and graphs.

(a) An Actual Image of the Farm

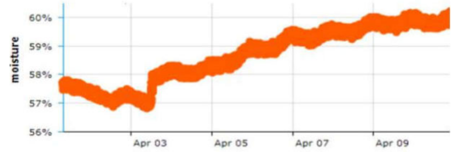
(b) IoT Equipment

FIGURE 6. The experimental farm and equipment.

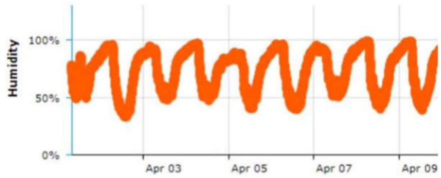
TABLE 2. The Hardware and Software of the Intelligent Agriculture Platform.

Hardware	Software
SIM5320E Iot Development Board	XMPP Web Platform
BME280 Temperature and Moisture Sensor	Python
BH1750 Illumination Sensor	MySQL Server
3-in-1 Soil Sensor	
Solar Power System	

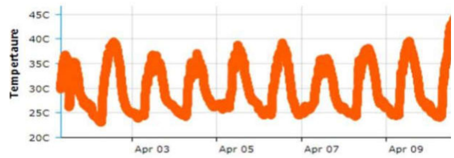
(a) Soil electrical conductivity



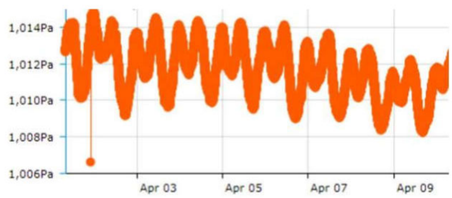
(b) Soil moisture



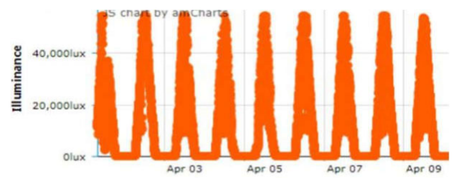
(c) Air humidity



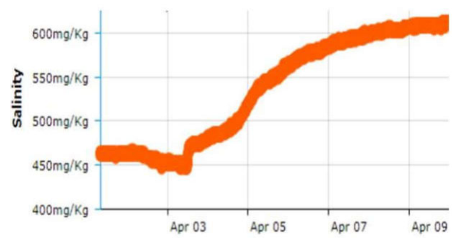
(d) Air temperature



(e) Air Pressure

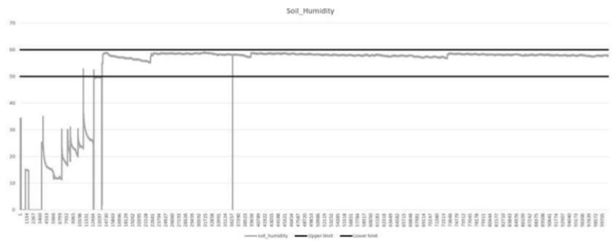


(f) Illumination

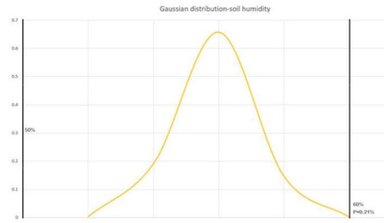


(g) Salinity

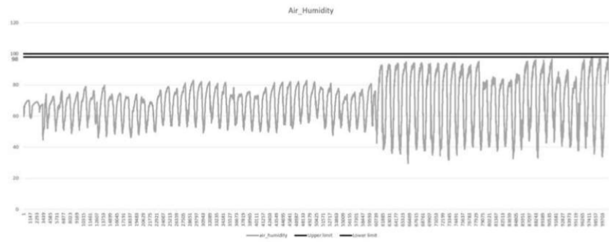
FIGURE 7. **Environmental factors of the farm.**



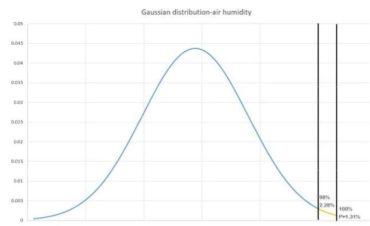
(a) Soil moisture detection



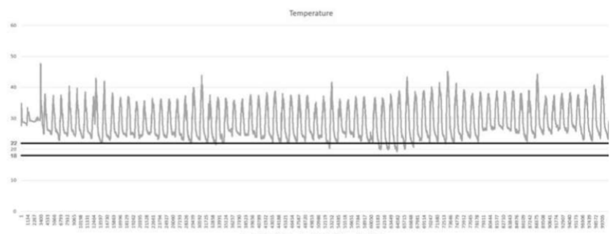
(b) Soil moisture Gaussian distribution



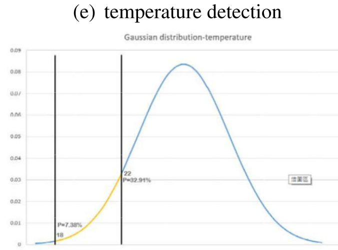
(c) Air humidity detection



(d) Air humidity Gaussian distribution



(e) temperature detection

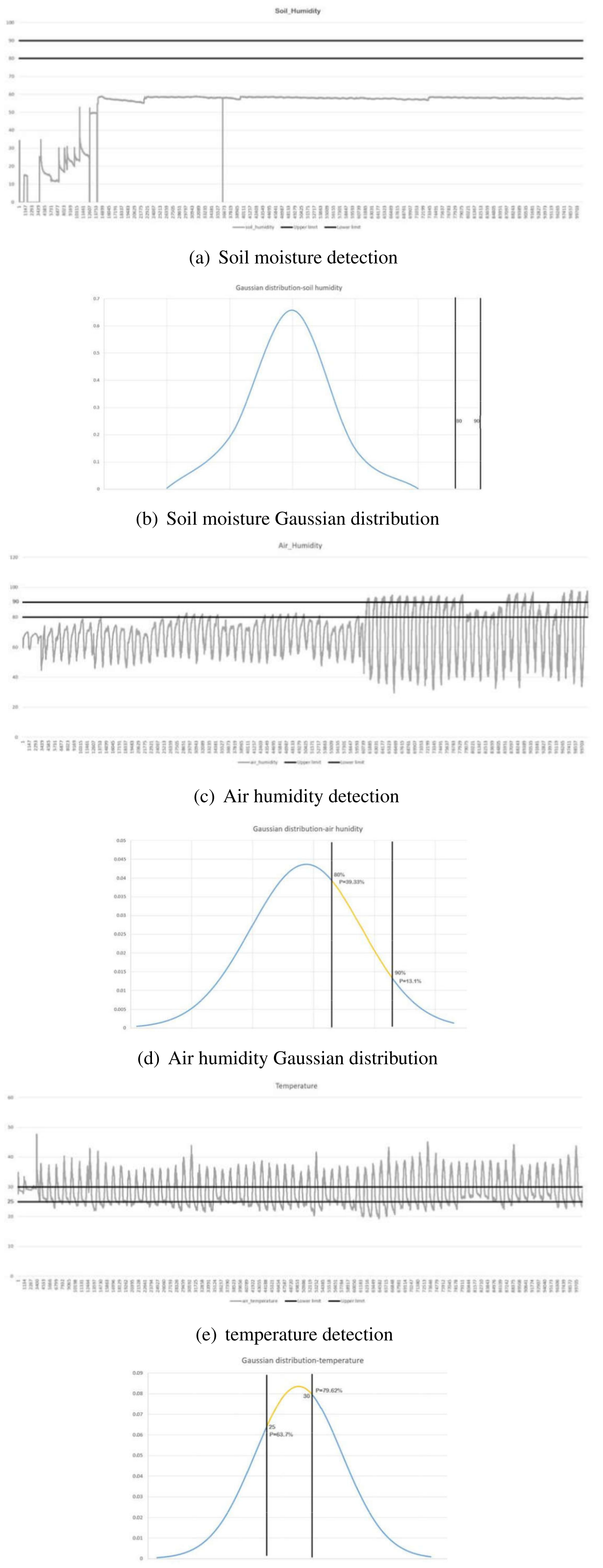


(f) temperature Gaussian distribution

FIGURE 8. **Crop analysis on celery.**

B. BIG DATA EXPERIMENT RESULTS

Our study conducted big data analysis. Each sensor contains 100,000 units of information, which is 700,000 units in total for all sensors. For our study, we chose celery, water spinach, green beans, and daikon for crop analysis. Our system first



(a) Soil moisture detection

(b) Soil moisture Gaussian distribution

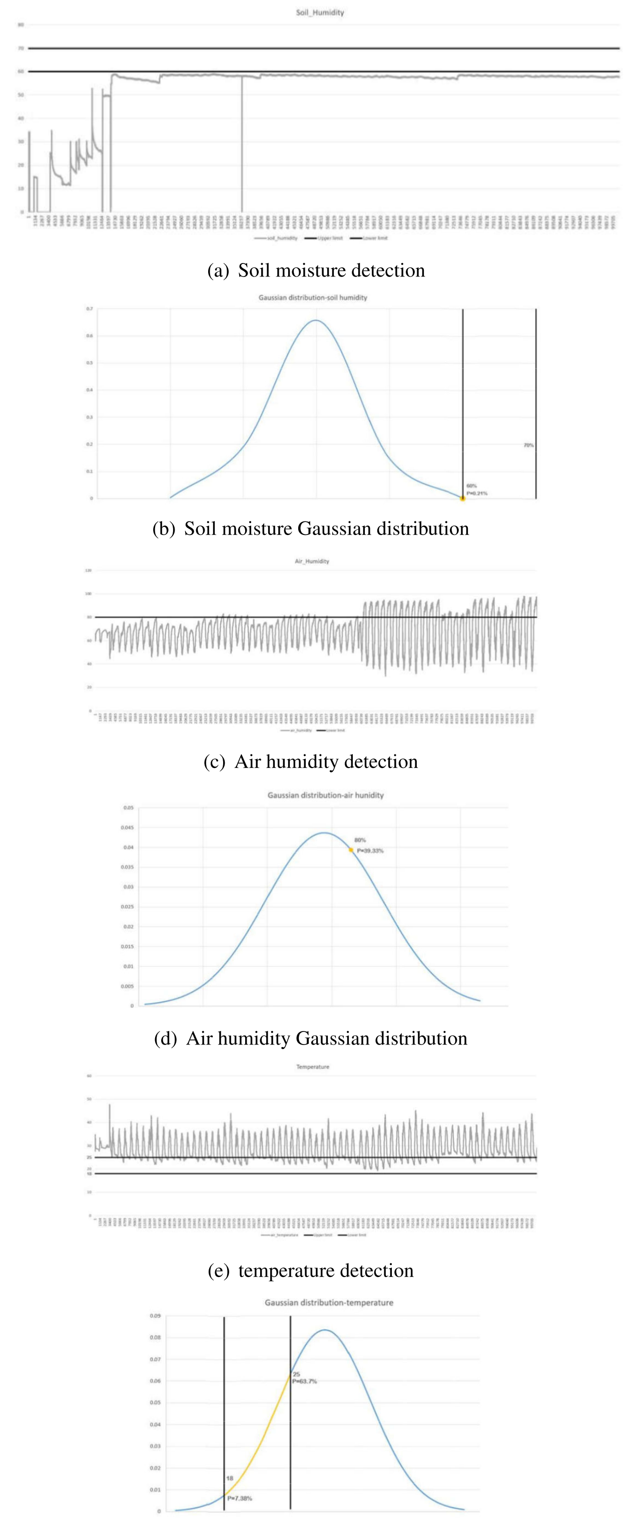
(c) Air humidity detection

(d) Air humidity Gaussian distribution

(e) temperature detection

(f) temperature Gaussian distribution

FIGURE 9. **Crop analysis on water spinach.**



(a) Soil moisture detection

(b) Soil moisture Gaussian distribution

(c) Air humidity detection

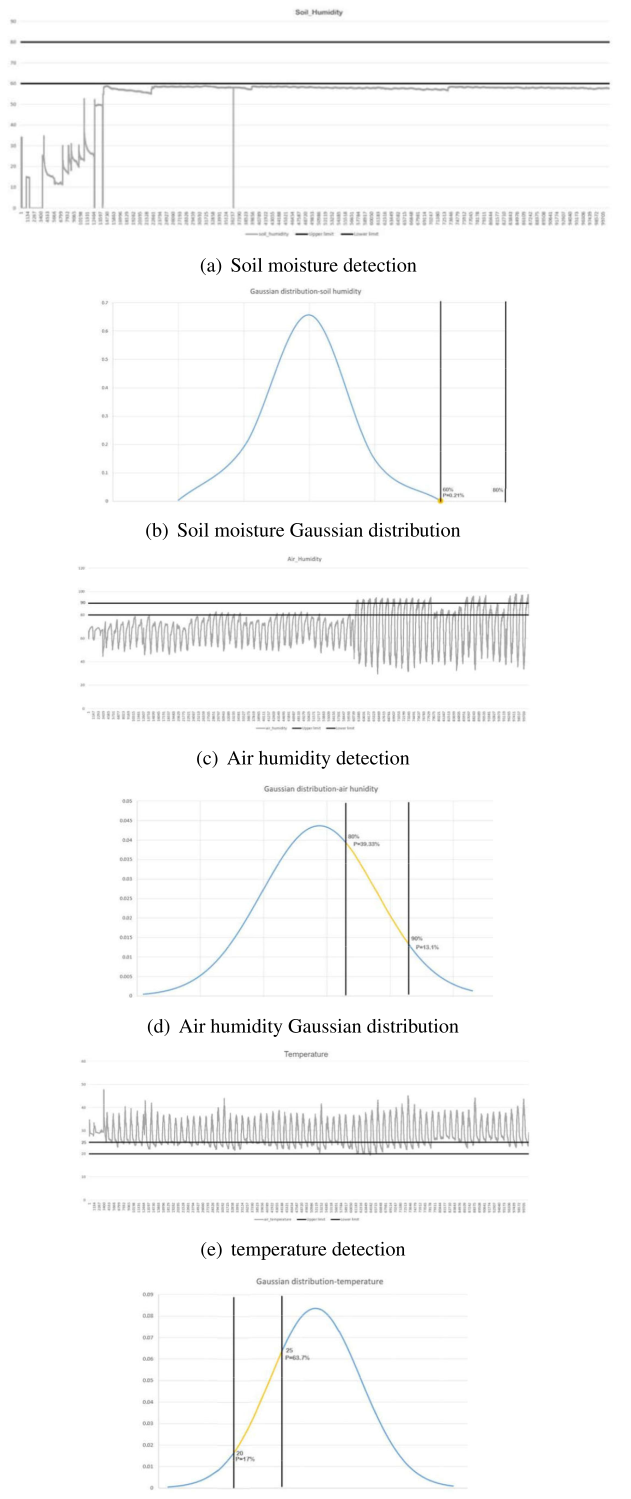
(d) Air humidity Gaussian distribution

(e) temperature detection

(f) temperature Gaussian distribution

FIGURE 10. **Crop analysis on green beans.**

determined whether the environmental factors of the crop were a match for the farm; if yes, then the system applied normal distribution to conduct risk assessment. As shown in Figure 8, temperature issues make it hard for the celery to adapt to the farm environment; normal distribution analysis also shows potential risks in soil moisture content. Moreover, it is hard to control temperature in an outdoor environment;



(a) Soil moisture detection

(b) Soil moisture Gaussian distribution

(c) Air humidity detection

(d) Air humidity Gaussian distribution

(e) temperature detection

(f) temperature Gaussian distribution

FIGURE 11. Crop analysis on daikon.

hence, the celery does not meet the necessary conditions of the farm. By contrast, Figure 9 shows that temperature conditions make the water spinach is indeed a good fit for the farm, but the normal distribution analysis reveals risks in soil moisture content and air humidity. Nonetheless, soil moisture content and air humidity are issues that can be combatted using the irrigation system. We can see from Figure 10 that green beans seem to be a good cultivation match for the farm in terms of environmental factors such as temperature and soil moisture content. In Figure 11, analysis suggests that daikon are a suitable crop for the farm; however, further normal distribution analysis reveal the potential risks in temperature, soil moisture content, and air humidity. Our proposed scheme is capable of utilizing big data analysis to help a farm select a suitable crop for cultivation and increase crop productivity.

V. CONCLUSION

Faced with extreme climate changes and increased global population, we are forced to emphasizemust address food issues including such as crops and agriculture. Our study proposes using an Intelligent Agriculture platform to monitor the environmental factors at on a farm and applying these environmental factors in analyzing cultivation techniques held by farmers. Our proposed scheme employs moving average and variance in data cleaning, which cleans out data with more drastic variation. Our study applies autocorrelation to compute periodicity while using 3D cluster correlation to conduct behavior analysis of farmer actions such as application of fertilizer or pesticide. Our study takes looks into environmental factors to assess whether a crop is suitable for a farm; it also takes global warming into consideration. The eExperiment results demonstrate show our analysis of four crops using our proposed approach; the results prove that farmers can gain a better understanding of whether a crop is suitable for their farm through by looking into factors such as temperature and soil moisture content. TThrough the environmental factor analysis proposed in our study help, farmers can gain insight into which crops they can grow; meanwhile, while the system keeps track of and analyzes crop cultivation behavior. In the future, our proposed scheme can incorporate artificial intelligence and apply the analysis results to help farmers achieve automatic cultivation and environment control.

ACKNOWLEDGMENT

This work was financially supported from the Young Scholar Fellowship Program by Ministry of Science and Technology (MOST) in Taiwan, under Grant MOST108-2636-E-003-001, and was partly funded by the MOST in Taiwan, under grant MOST108-2221-E-197-012-MY3.

REFERENCES

[1] C. Kulatunga, L. Shalloo, W. Donnelly, E. Robson, and S. Ivanov, "Opportunistic wireless networking for smart dairy farming," *IT Prof.*, vol. 19, no. 2, pp. 16–23, Mar./Apr. 2017.

[2] F. Y. Narvaez, G. Reina, M. Torres-Torriti, G. Kantor, and F. A. Cheein, "A survey of ranging and imaging techniques for precision agriculture phenotyping," *IEEE/ASME Trans. Mechatronics*, vol. 22, no. 6, pp. 2428–2439, Dec. 2017.

[3] R. Gebbers and V. I. Adamchuk, "Precision agriculture and food security," *Science*, vol. 327, no. 5967, pp. 828–831, 2010.

[4] M. Taniguchi, N. Masuhara, and K. Burnett, "Water, energy, and food security in the Asia Pacific region," *J. Hydrol., Regional Stud.*, vol. 11, pp. 9–19, Jun. 2017.

[5] H. Navarro-Hellín, R. Torres-Sánchez, F. Soto-Valles, C. Albaladejo-Pérez, J. A. López-Riquelme, and R. Domingo-Miguel, "A wireless sensors architecture for efficient irrigation water management," *Agricult. Water Manage*, vol. 151, pp. 64–74, Mar. 2015.
[6] A. Kamilaris, A. Kartakoullis, and F. X. Prenafeta-Boldú, "A review on the practice of big data analysis in agriculture," *Comput. Electron. Agricult.*, vol. 143, pp. 23–27, Dec. 2017.
[7] R. Peters, "Nine billion and beyond: From farm to fork [agriculture big data]," *Eng. Technol.*, vol. 11, no. 4, p. 74, May 2016.
[8] S. Wolfert, L. Ge, C. Verdouw, and M.-J. Bogaardt, "Big data in smart farming—A review," *Agricult. Syst.*, vol. 153, pp. 69–80, May 2017.
[9] Y. Huang, Z.-X. Chen, T. Yu, X.-Z. Huang, and X.-F. Gu, "Agricultural remote sensing big data: Management and applications," *J. Integrative Agricult.*, vol. 17, no. 9, pp. 1915–1931, 2018.
[10] X. Tan, L. Di, M. Deng, A. Chen, F. Huang, C. Peng, M. Gao, Y. Yao, and Z. Sha, "Cloud- and agent-based geospatial service chain: A case study of submerged crops analysis during flooding of the Yangtze River Basin," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 3, pp. 1359–1370, Mar. 2015.
[11] X. Bai, L. Liu, M. Cao, J. Panneerselvam, Q. Sun, and H. Wang, "Collaborative actuation of wireless sensor and actuator networks for the agriculture industry," *IEEE Access*, vol. 5, pp. 13286–13296, Jul. 2017.
[12] B. Hulsegge and K. H. de Greef, "A time-series approach for clustering farms based on slaughterhouse health aberration data," *Preventive Veterinary Med.*, vol. 153, pp. 64–70, May 2018.
[13] J. Ruan, Y. Wang ; F. T. S. Chan, X. Hu, M. Zhao, F. Zhu, B. Shi, Y. Shi, and F. Lin, "A life cycle framework of green IoT-based agriculture and its finance, operation, and management issues," *IEEE Commun. Mag.*, vol. 57, no. 3, pp. 90–96, Mar. 2019.
[14] R. Ranjan, L. Wang, A. Y. Zomaya, J. Tao, P. P. Jayaraman, and D. Georgakopoulos, "Advances in methods and techniques for processing streaming big data in datacentre clouds," *IEEE Trans. Emerg. Topics Comput.*, vol. 4, no. 2, pp. 262–265, Apr. 2016.
[15] P. Sundaravadivel, E. Kougianos, S. P. Mohanty, and M. K. Ganapathiraju, "Everything you wanted to know about smart health care: Evaluating the different technologies and components of the Internet of Things for better health," *IEEE Consum. Electron. Mag.*, vol. 7, no. 1, pp. 18–28, Jan. 2018.
[16] Q. Li, Y. Zhang, T. Ji, X. Lin, and Z. Cai, "Volt/var control for power grids with connections of large-scale wind farms: A review," *IEEE Access*, vol. 6, pp. 26675–26692, 2018.
[17] J. Wan, S. Tang, Z. Shu, D. Li, S. Wang, M. Imran, and A. Vasilakos, "Software-defined industrial Internet of Things in the context of industry 4.0," *IEEE Sensors J.*, vol. 16, no. 20, pp. 7373–7380, Oct. 2016.
[18] J. Lin and R. Chen, "A novel group decision making method under uncertain multiplicative linguistic environment for information system selection," *IEEE Access*, vol. 7, pp. 19848–19855, 2019.
[19] M. Lin, Q. Zhan, R. Chen, and Z. Xu, "Group decision-making model with hesitant multiplicative preference relations based on regression method and feedback mechanism," *IEEE Access*, vol. 6, pp. 61130–61150, 2018.
[20] M. Lin, Q. Zhan, Z. Xu, and R. Chen, "Group decision making with probabilistic hesitant multiplicative preference relations based on consistency and consensus," *IEEE Access*, vol. 6, pp. 63329–63344, 2018.

FAN-HSUN TSENG (S'12–M'18) received the Ph.D. degree in computer science and information engineering from National Central University, Taiwan, in 2016. He is currently an Assistant Professor with the Department of Technology Application and Human Resource Development, National Taiwan Normal University, Taipei, Taiwan. He received the 2018 Young Scholar Fellowship Program by the Ministry of Science and Technology, Taiwan, for his dedication to research of engineering and technologies. His research interests include mobile edge computing, 5G mobile networks, and artificial intelligence. He has served as an Associate Editor-in-Chief for the *Journal of Computers* and an Associate Editor for *Human-Centric Computing and Information Sciences*.

HSIN-HUNG CHO (M'18) received the B.S. degree from the Department of Applied Mathematics, Hsuan Chuang University, Taiwan, in 2010, the M.S. degree from the Institute of Computer Science and Information Engineering, National Ilan University, in 2011, and the Ph.D. degree from the Department of Computer Science and Information Engineering, National Central University. He has been an Assistant Professor with the Department of Computer Science and Information Engineering, National Ilan University, since 2017. His research interests include wireless sensor networks, mobile networks, cloud computing, artificial intelligence, and metaheuristics. He has served as the Managing Editor for *Communications of the CCISA* (CCCISA).

HSIN-TE WU received the Ph.D. degree from the Department of Computer Science and Engineering, National Sun Yat-sen University, Taiwan, in 2013. He is currently an Assistant Professor with the Department of Computer Science and Information Engineering, National Penghu University of Science and Technology, Taiwan. His research interests include computer networks, wireless networks, speech compression, network security, and the Internet of Things.