IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

SPECIAL SECTION ON DATA MINING FOR INTERNET OF THINGS

Received August 12, 2020, accepted August 25, 2020, date of publication September 18, 2020, date of current version October 13, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3024891

An AIoT Based Smart Agricultural System for Pests Detection

CHING-JU CHEN[1], YA-YU HUANG[2], YUAN-SHUO LI[2], CHUAN-YU CHANG[3], (Senior Member, IEEE), AND YUEH-MIN HUANG[2], (Senior Member, IEEE)

[1]Department of Bachelor Program in Interdisciplinary Studies, National Yunlin University of Science and Technology, Douliou 64002, Taiwan
[2]Department of Engineering Science, National Cheng Kung University, Tainan 70101, Taiwan
[3]Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Douliou 64002, Taiwan

Corresponding author: Yueh-Min Huang (huang@mail.ncku.edu.tw)

This work was supported in part by the Ministry of Science and Technology, Taiwan, R.O.C., under Grant MOST 108-2321-B-067F-001-, and in part by the Intelligent Recognition Industry Service Center through the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE), Taiwan.

ABSTRACT In this study, artificial intelligence and image recognition technologies are combined with environmental sensors and the Internet of Things (IoT) for pest identification. Real-time agricultural meteorology and pest identification systems on mobile applications are evaluated based on intelligent pest identification and environmental IoT data. We combined the current mature AIoT technology and deep learning and applied it to smart agriculture. We used deep learning YOLOv3 for image recognition to obtain the location of *Tessaratoma papillosa* and analyze the environmental information from weather stations through Long Short-Term Memory (LSTM) to predict the occurrence of pests. The experimental results showed that the pest identification accuracy reached 90%. Precise positioning can effectively reduce the amount of pesticides used and reduce pesticide damage to the soil. The current research provides the location of the pest and the extent of the pests to farmers can accurately use pesticide application at a precise time and place and thus reduce the agricultural workforce required for timely pest control, thus achieving the goal of smart agriculture. The proposed system notifies farmers of the presence of different pests before they start multiplying in large numbers. It improves overall agricultural economic value by providing appropriate pest control methods that decrease crop losses and reduce the environmental damage caused by the excessive usage of pesticides.

INDEX TERMS Deep learning, YOLOv3, pests and diseases, smart agriculture, unmanned aerial vehicle (UAV), artificial intelligence (AI), Internet of Things (IoT), the artificial Intelligence of Things (AIoT).

I. INTRODUCTION

Crop production is closely related to the presence of pests and plant diseases. Pests often hide behind the leaves of plants during the day to avoid the heat and appear on the leaves in the evening or at night. Therefore, it is not easy to observe their presence on crops during the day. When farmers become aware of pest damage, pests have often multiplied and spread uncontrollably. At this stage, a large quantity of pesticides is required to spray the crops to eliminate the pests and reduce agricultural damage. However, once the crops have been sprayed with pesticides in the growing season, pesticide residues remain even after washing.

The associate editor coordinating the review of this manuscript and approving it for publication was Patrick Hung.

Crops often suffer from bacterial infections due to pests that result in large-scale crop diseases. To prevent such conditions, it is necessary to burn the infected crops to prevent the spread of the bacteria. However, this approach causes significant damages in agricultural production without effectively resolving pest problems. Therefore, the goal of this research is to apply the AIoT and deep learning technology to an environmental analysis of crop growth and the prediction of pest occurrence to improve the productivity of crops and reduce the size of the agricultural workforce.

The application of image recognition has become more extensive due to the development of artificial intelligence in image processing models. Gladence *et al.* [1] applied AIoT to home automation recommendations. In another work by Priya *et al.* [2], the common image recognition CNN model was used in medical detection and prediction.

In recent years, image recognition technology has been used to assist with pest recognition for the purpose of controlling agricultural pest damage and increasing crop production. Mary Gladence *et al.* [3] proposed the use of human-robot interaction in the field of artificial intelligence and provided a reference for pest identification through a combination of environmental factor information and deep learning to help farmers analyze crop growth trends and prevent pest damage early

II. RELATED WORKS

Martin and Moisan [4] established an automated system for pest identification. In their system, the camera is placed above a sticky insect trap and automatically captures images of pests collected by the insect trap each day. The Single Seed Descent (SSD) method was used to identify and analyze the pests in the images. The collected data has been used to develop appropriate pest control methods based on the identified pest species. Once the SSD model has been trained, the pests identification accuracy reaches 84%, and the pest species classification accuracy reaches 86%.

Wang *et al.* [5] set up a large number of environmental sensors in an apple orchard for the purpose of recording the status of the orchard. The YOLOv3 model was used to identify anthrax on the surface of the apples and analyze the health of the apples. The authors also used the YOLOv3-Dense model, which is more suitable for identifying apple anthracnose.

In the above work, the YOLOv3 and YOLOv3-Dense models were both used to identify apple anthracnose, and the SSD was successfully used to identify pests.

A. MACHINE LEARNING TO IDENTIFY AND CLASSIFY PEST IMAGES

With the advancement of science and technology, the amount of image data is increasing, along with the time required to classify the image data. Therefore, scholars are studying how to use machine learning to recognize and classify images. Machine learning is divided into supervised learning and unsupervised learning, the main difference being whether or not the machine is able to automatically extract features from the data structure.

1) SUPERVISED LEARNING

During the machine training process, it is necessary to provide machines with labeled data. For example, after a machine has seen 1,000 labeled images of apples and oranges, we can give a test image and asked whether the image contains apples or oranges.

2) UNSUPERVISED LEARNING

There is no need to label the data in advance, and the machine does not know whether or not the result of its classification is correct during learning. The machine must find the rules from all of the input examples in order to classify on its own.

In summary, supervised learning adds artificial labels to the input data and uses regression analysis methods to obtain the predicted results. Unsupervised learning finds suitable patterns from a large amount of data through algorithms and classifies the data automatically.

B. DEEP LEARNING TO IDENTIFY AND CLASSIFY PEST IMAGES

Deep learning is based on the machine learning framework, so the training process first involves unsupervised learning and clusters the training data set to learn what sort of data will be classified. Supervised learning is then performed to label the expected output value of each entry with the feature vectors in the training data set given as input and the expected classification given as the output. Finally, the loss function is used to calculate the standard deviation between the expected output and the actual output.

There are two common approaches in deep learning for pest identification and classification:

1) Miranda *et al.* [6] used the VGG19 method for image feature extraction and recognition in the detection of 24 types of pests from crop images.
2) The authors in [7] detected and classified 12 different pests using several CNN approaches and compared the classification results with machine learning methods such as SVM and Fisher. The classification accuracy of the machine learning method was 80%, and the classification accuracy rate of the CNN method was 95%.

C. USING IMAGE AUGMENTATION TO INCREASE THE PESTS TRAINING SAMPLE DATABASE

Data augmentation methods can be roughly divided into Geometric Transformation and Photometric Transformation methods. Ding and Taylor [8] studied the impact of different data augmentation methods on image recognition rate. Three geometric transformations (flipping, rotating, and cropping) and three luminosity transformations (color jittering, edge enhancement, and fancy principal component analysis). The amplified image samples and the original image samples are used as training data for training on the CNN model. The experimental results showed that each amplification method improves the accuracy of CNN, where the impact of cropping is the most obvious. Therefore, the authors speculate that the training samples generated by the cropping method will avoid the over-fitting of data and improve CNN performance.

In addition to augmenting the images by geometric and photometric transformations, it is also possible to transform the image style through the CycleGAN [9] or transfer the features of objects to other objects to increase the number of training samples. The aforementioned apple anthracnose identification research uses the CycleGAN [9] to increase the number of training samples and improve the accuracy of image recognition. Perez and Wang's research (supplementary source of literature) have also found that geometric transformation and brightness variation of image and CycleGAN are able to enhance the image recognition accuracy [10].

D. IDENTIFY THE LOCATION OF THE PEST WITH THE TARGET DETECTION MODEL

With the advancement of technology and the development of artificial intelligence, deep learning has significantly increased the efficiency of image recognition, resulting in many different artificial intelligence models that have encouraged the development in image recognition technology. These models can be divided into two types, the two-stage object detector method and the one-stage object detector method [11].

1) TWO-STAGE OBJECT DETECTOR METHOD

The two-stage method is used to generate a region proposal after receiving the input image in order to identify the type of object. Its accuracy is higher, but it requires more computing resources. The most representative model is the R-CNN [12] and the Faster R-CNN [11], [13].

1) The RNN is mainly used to solve time series problems. It typically has no temporal data in the input data. In Fig.1, it can be seen that the RNN stores the output of the hidden layer in the memory, and when the next input item has been entered, it also considers the last value stored in the memory in its calculations. For the simple stock price learning RNN, we can input the P_{t-1} data to learn the P_t stock price, and M_1 is the saved hidden output. When we modify the P_{t+1} value from training P_t, M_2 will be taken into consideration and added to the output of P_{t+1}.
2) The Faster R-CNN method is divided into four parts:
   a) Conv layers: The input image is reshaped and the rescale factor is recorded in im_info as a proposal layer for the ROIPooling alignment information. Then, the feature map is extracted after 13 Conv layers, 13 ReLU layers, and 4 Max Pooling layers.
   b) Region Proposal Network (RPN): The most significant difference from the Fast R-CNN is that the RPN is used to generate region proposals. The RPN uses softmax to distinguish positive and negative anchors, and the Bounding-box regression is used to modify anchor position information to obtain more accurate positions.
   c) RoIPooling: The results from the feature maps and the RPN are collected; the information is then integrated and sent to the FC for training.
   d) Classification: The ROIPooling input value is used for classification and the Bounding-box regression to obtain the best object position.

2) ONE-STAGE OBJECT DETECTOR METHODS

One-stage object detector methods detect and identify objects simultaneously without a prior region proposal.

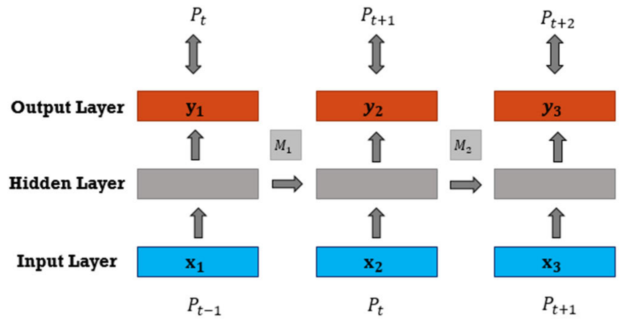

FIGURE 1. The flowchart of RNN stock price prediction.

The recognition speed is faster, but the accuracy is lower. Two popular one-stage object detector models are the Single Shot MultiBox Detector [14] and the YOLO [15] model. The YOLO model only uses a single convolutional neural network (CNN) to determine the image inputs and outputs, so the efficacy of deep learning is greatly accelerated.

After inputting the image into the YOLO model, the features of the image are first extracted. Then, the image is divided into an grid. The CNN is used to predict the number of possible bounding boxes in each grid; the detected object type is calculated, and the image recognition result is then output.

In the Titan X GPU hardware device environment, the YOLO model can process 45 images per second, where the delay time for real-time video processing is less than 25 milliseconds.

Since the YOLO model uses the entire image for training, the image detection accuracy will be higher with SSD.

The developer of the YOLO model continued to improve the model and published the YOLOv3 model [16]. The YOLOv3 model uses multi-scale fusion for predictions and increases the network architecture to 53 convolutional layers. In Fig.2, The YOLOv3 model also introduces the concept of the residual network to increase the accuracy of the model and improve the traditional YOLO model problem of having poor recognition of small objects.

The YOLO algorithm extracts the features from an input image through the feature extraction network to obtain a fixed-size feature map, such as 13×13, and divides the input image into 13×13 grid cells. If the center coordinate of the object falls in a grid cell in the ground truth, then the grid cell has predicted the object successfully.

The YOLOv3 model uses multiple-scale fusion methods to make predictions. It uses FPN-like up-sample and fusion methods (respectively 13×13, 26×26 and 52×52) to detect the multiple-scale feature maps. The detection outcome for small targets is significantly improved. It is able to find the location of the identified target in the entire image and frame it.

E. ANALYSIS MODEL OF ENVIRONMENTAL SENSING DATA

In 2016, Liu *et al.* [12] proposed integration of the number of pests, risk level, and historical environmental data

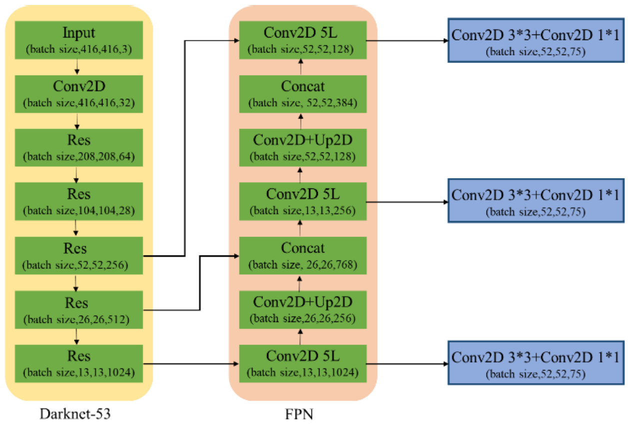

FIGURE 2. The YOLOv3 network architecture diagram.

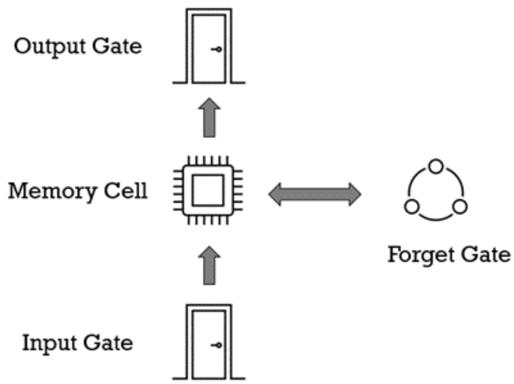

FIGURE 3. The schematic diagram of the LSTM process.

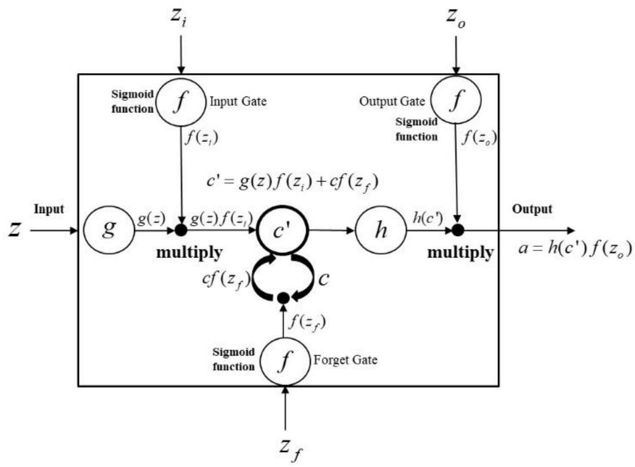

FIGURE 4. The mathematical representation of LSTM.

with the data provided by the user, where the time series data set is based on the daily maximum temperature, daily minimum temperature, daily maximum humidity, daily minimum humidity, and the number of pests caught per day. The prediction model used in the network-based decision support program is built on a mixed regression model, which is referred to as the Vector Autoregression Moving-Average with Exogenous Regressors (VARMAX) model with endogenous and exogenous variables. The prediction results are location-specific and can be used to determine the future risk level. They can provide recommendations for pest prevention and control so that farmers and pest managers can take preventative measures to avoid crop damage due to pests. These functions can be used to implement the main components of the IPM program, including the inspection, identification, and control of pests in the orchard.

F. DEEP LEARNING FOR DATA ANALYSIS

The LSTM (long short-term memory) model is currently the most commonly used model in RNNs (Recurrent Neural Networks) [17]. The flow chart of this model is shown in Fig. 3, which is composed of four main components: the input gate, the output gate, the memory cell, and the forget gate.

1) Input Gate: When a feature is given as an input, the input gate controls whether or not to input the value in this iteration.
2) Memory Cell: Stores the calculated value to be used in the next stage.
3) Output Gate: Controls whether or not to output the estimated value in this iteration.
4) Forget Gate: Controls whether or not to clear the memory, similar to a restart.

The mathematical structure of the LSTM model is shown in Fig. 4. The input is represented as $g(z)$, and the input gate is $f(z_i)$; the general activation function f uses the sigmoid function to determine the probability of turning on the gate.

The probability that the input gate is turned on can be determined by multiplying $g(z)$ and $f(z_i)$. The memory cell records the current value of the input, adds the previous input value, and multiplies the probability of the forget gate to decide whether or not to forget the previously recorded value [$c' = g(z)f(z_i) + cf(z_f)$].

The value of c_{t-1} is a vector stored in the LSTM memory, where the vector input at time t is z_t. z_t is multiplied by a linear conversion function to obtain $g(z_t)$ as an input. z_t is multiplied by another linear conversion function to obtain $I(z_t)$ to control the input gate. z_t is multiplied by another linear conversion function to obtain $F(z_t)$ to control the forget gate, and z_t is multiplied by a linear conversion to get $O(z_t)$ to control the output gate. The LSTM flowchart is shown in Fig. 5.

The value of $g(z_t)$ is multiplied by the value of $I(z_t)$ after the Sigmoid function. The result is added to the product of $F(z_t)$ after the Sigmoid function and c_{t-1} and stored in the LSTM. The value of $O(z_t)$ after the Sigmoid function is multiplied by the stored value to obtain the output value a^t. The same steps are followed to calculate the value to be stored in the LSTM and the output value from the next input data z_{t+1}.

III. METHODS

In this study, the pests are identified using artificial intelligence; the pest recognition model is trained through deep

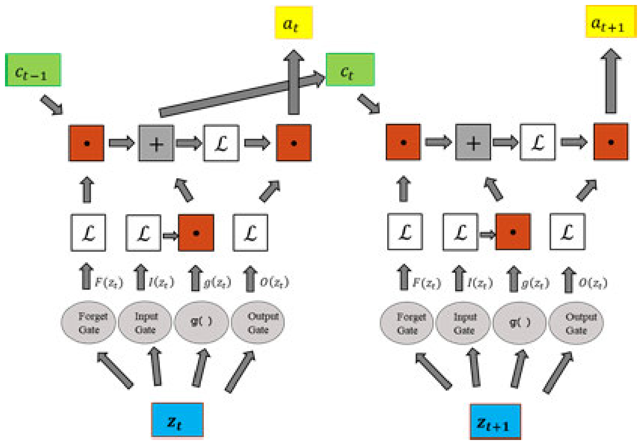

FIGURE 5. The LSTM flowchart.

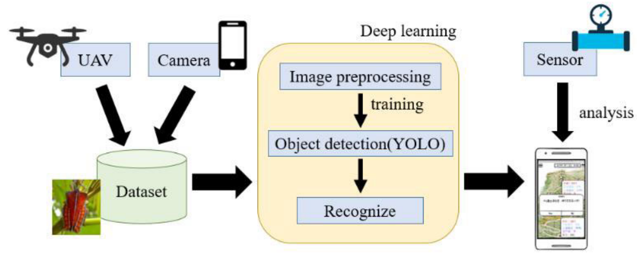

FIGURE 6. The architecture of the proposed system.

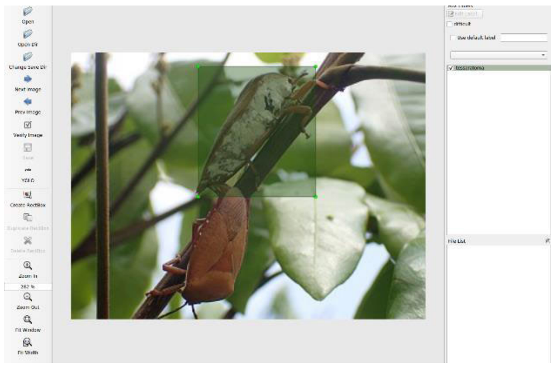

FIGURE 7. The labeled training data.

learning, and the image data for *Tessaratoma papillosa* are collected using a mobile application and a drone from a longan orchard. A real-time image recognition system for *Tessaratoma papillosa* is established. It is combined with environmental sensors to analyze the impact of environmental factors on the pests' life-cycle so that the system can promptly inform the farmers regarding pest occurrences and damage, and the farmers can apply pesticides accurately on a timely basis. The proposed system is able to provide improved control over *Tessaratoma papillosa*, reduce the labor costs in the orchard, and improve the yield and quality of the crop.

The YOLOv3 model is used in this study, which is based on CNN and has good results in terms of image detection and classification. The YOLOv3 model is used to classify and label the pests. The drone and mobile application are used to collect images of the pests in the orchards. Multiple environmental sensors are placed as meteorological observation stations for the orchard to regularly obtain environmental information including the temperature, humidity, and light intensity, for the analyses of the pest's life-cycle and population. The purpose of the system is to predict the distribution location and periods of occurrence of the pests, and to provide pest locations and damage level, as well as other relevant information to the farmers for real-time management of the orchard.

We use drones and mobile devices to collect images of *Tessaratoma papillosa*, perform image enhancement and augmentation, and use the YOLOv3 model to classify and identify the processed image samples. We use a random gradient descent as the learning algorithm and mAP to assess the accuracy. The architecture of the proposed system is shown in Fig. 6.

A. SAMPLE COLLECTION

We collaborated with the farmers to collect the training data samples. The farmers used the mobile application that we designed to take images of the pests. The mobile application sent the pest identification results to the cloud database. We also used a drone to collect images of the pest, which were sent to the cloud database as well. In this research, approximately 687 images of adult *Tessaratoma papillosa* were collected. The pest was first labeled in the image, which is then augmented through image pre-processing, as shown in Fig. 7, and finally, the GPS information from the image was used to indicate the location of the pest.

B. IMAGE PRE-PROCESSING

To improve the pest image recognition accuracy, we increased the number of training samples through image pre-processing, using feature enhancement, edge detection, and grayscale processing.

1) COLOR FEATURES ENHANCEMENT

The color information was composed of red, green and blue (RGB) channels, and the color histogram records the occurrences of each color in the image formed three 256-dimension vectors. This vector was used to represent the color characteristics of the image.

2) SOBEL EDGE DETECTION

In the proposed method, the vertical Sobel edge detector is first applied to enhance the vertical edges of the image, followed by the application of the horizontal Sobel edge detector to enhance the horizontal edges of the image. The two images are then merged to obtain the final edge detected image.

3) GRAYSCALE IMAGE PROCESSING

Since the electromagnetic radiation (reflection or emission) of various substances causes them to have different

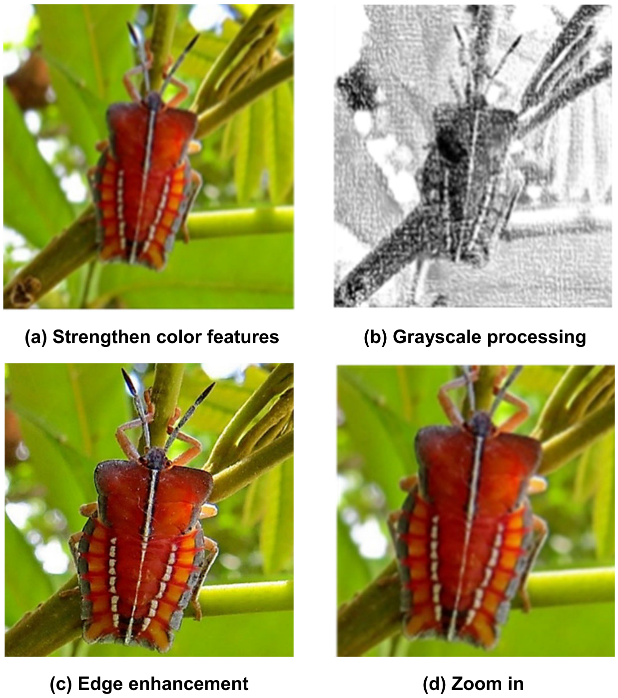

FIGURE 8. Image pre-processing of samples to enhance the characteristics of the pests.

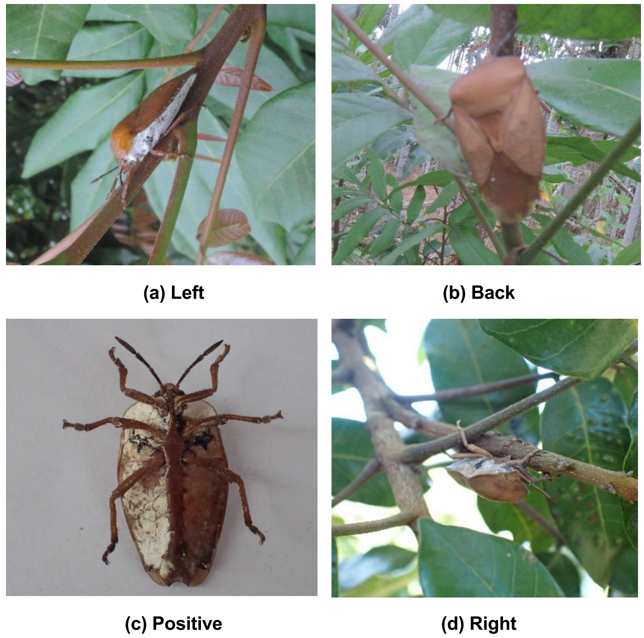

FIGURE 9. Tessaratoma papillosa images acquired from different viewing directions.

TABLE 1. 20 images per view to evaluate the recognition accuracy.

Category	Back	Left	Right	Abdomen
Number of Training Samples	433	152	98	166
Recognition Accuracy	80%	45%	30%	55%

intensities, and photo-sensitive materials also have varied sensitivities, the intensities of the detected radiation are interpreted as a grayscale image. The grayscale images are divided into seven levels: white, gray-white, light gray, gray, dark gray, light black, and black, which serve as essential markers and the basis for image interpretation.

Fig. 8 shows the different features of a captured image used as a multiple-input training data sample, where the separately obtained probability was accumulated through the deep learning model to improve identification accuracy.

We analyzed the identification accuracy of *Tessaratoma papillosa* in different directions. For each of the back, left, right, and abdominal perspectives, 20 pest images were used to determine the recognition rate, as shown in Fig. 9.

Table 1 shows that the training data had significantly fewer samples for the left, right, and abdominal views than for the back view, but the recognition rate for the back view was higher than 80%, and the recognition rates for the left, right, and abdominal views were less than 60%.

Many studies have found that data augmentation improves model recognition accuracy, where it is recommended that sufficient training samples be collected for image pre-processing, such as cropping, rotation, contrast enhancement, noise addition, and edge sharpening.

Imgaug is a library for Python with 98 types of image augmenting functions that can be used to process images and revise the image information.

After completing the labeling, the imgaug library was used for image enhancement in machine learning and the labeled information was automatically generated to increase the number of training samples and improve the recognition accuracy of the YOLOv3 model.

In this work, we augmented the image data for the left, right, and abdominal views to increase the number of images in the training dataset.

In this study, six image augmentation methods were used to increase the number of images for the YOLOv3 training samples: normalization, left-right flip, edge sharpening, gamma contrast adjustment, Gaussian noise addition, and Gaussian blur. Fig. 10 shows the image augmentation results.

We used the 1) image augmentation method and 2) sample compensation method to increase the number of left, right, and abdominal image samples before performing image augmentation for these images. Two different expanded sample methods were used to evaluate the enhanced recognition accuracy of the pest images.

4) IMAGE AUGMENTATION METHOD

Table 2 shows the pest identification accuracies for the back, left, right, and abdominal views after using six kinds of images augmentation processing: normalization, left-right flip, edge sharpening, gamma contrast adjustment, Gaussian noise addition, and Gaussian blur. It can be observed from Table 2 that the recognition accuracy of each view was improved after image augmentation. However, the recognition accuracies of the left, right and abdominal views were still lower than that for the back view.

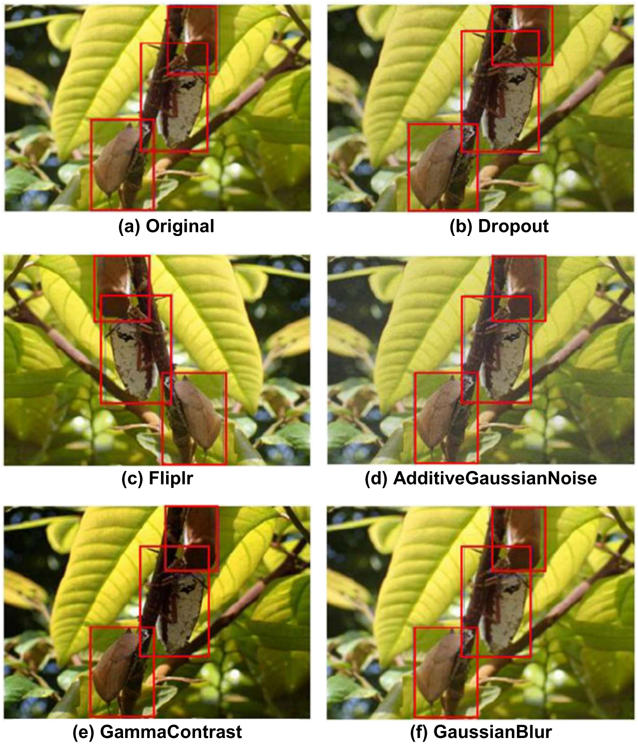

FIGURE 10. The results of image augmentation.

TABLE 2. Accuracies after the equal-scale image augmentation method.

Category	Back	Left	Right	Abdomen
Number of Original Image Samples	433	152	98	166
Sample Data of 6 Kinds of Images Augmentation	2598	912	588	996
Identification Accuracy	92%	66%	59%	73%

5) SAMPLE COMPENSATION METHOD

In this experiment, additional image data have the same amount of sample images for the left, right, and abdominal views were generated by rotating the images at different angles (90o, 180 o and 270 o), flipping the images left and right, as well as through other operations. All of the images in the training samples underwent six types of image augmentation: normalization, left-right flip, edge sharpening, gamma contrast adjustment, Gaussian noise addition, and Gaussian blur. After the training, it was found that the recognition accuracy of each view increased to more than 90%. The results of the pest recognition accuracy are shown in Table 3.

After the SSD model was trained, the pest identification accuracy reached 84%, and the pest classification accuracy reached 86%.

C. OBJECT DETECTION ON IMAGE CLASSIFICATION

Traditional fixed feature extraction methods often fail to obtain useful features for image segmentation. In recent years, the convolutional neural network (CNN) architecture was developed and has led to improved feature extraction.

TABLE 3. Accuracies after increasing the image samples followed by image augmentation.

Category	Back	Left	Right	Abdomen
Number of Original Image Samples	433	152	98	166
Rotation Cropping Method in Increase the Left, Right and Abdomen Image Samples	433	456	449	482
Sample Data of 6 Kinds of Images Augmentation	2598	2736	2694	2892
Identification Accuracy	93%	89%	90%	92%

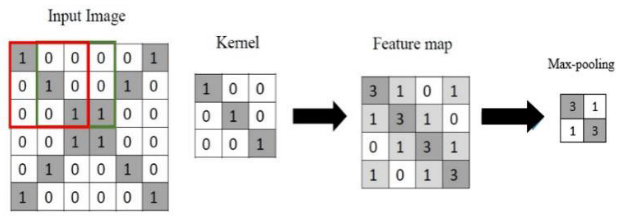

FIGURE 11. Convolutional layer and pooling layer.

Convolutional neural networks belong to one of the feed-forward neural networks, which extracts image features and completes image classification at the same time. The input of the neural network is a set of images, and each image passes through several convolutional layers and places a pooling layer in the convolution layer appropriately, as shown in Fig. 11. The features of the model are obtained first, and these features are placed into the fully connected layer for classification. Image classification is achieved through repeated learning in the deep layers.

The main differences between the CNN and the general neural network are the CNN's local perception and weight sharing. The local features of the image are extracted by the convolution kernel so that each area of the image shares this convolution kernel. This is a method widely used in the field of image recognition.

The traditional CNN model was used in this study. After inputting the original image data into the model, it only has to be classified in advance, and then it can be trained without any labeling.

In this study, the number of pests was recorded to indicate the degree of damage and the location of the pests so the drone could carry out accurate pesticide spraying. Therefore, this study used the YOLOv3 network architecture to re-train the target detection model.

The YOLOv3 model is a Darknet-53 network structure that uses full convolution and introduces a residual structure. Due to the large number of layers in the deep learning network, gradient descent problems often occur during training. With ResNet residual architecture, the YOLOv3 model reduces the difficulty of training deep networks. The Darknet-53 network system can be as deep as 53 layers and significantly improves identification accuracy.

In the experiment, we calculated the average accuracies from 10 different angle images and found that when the

TABLE 4. The research results of the number of training samples that are uneven with the YOLO V3 model, and the number of training samples is average with the YOLO V3 model.

Category	Number of Training Samples	Accuracy
The Uneven Number of Training Samples with The YOLOv3 Model	849	73%
The Average Number of Training Samples with The YOLOv3 Model	10920	92%

training samples were uneven (more images of the back, but fewer of the left, right, and abdomen), the accuracy was approximately 73%. We classified images of the back, left, right, and abdominal angle images and then generated a limited number of training samples through data preprocessing (rotation and cropping) so that all angles had the same number of training samples. The accuracy of pest recognition rose to more than 92% after training using the neural network with an average amount of training samples.

We compared the research results for an uneven number of training samples with an average number of training samples using the YOLOv3 model, as shown in Table 5. Based on the results, we inferred that a useful data pre-processing method can improve the efficiency of deep learning as well as recognition accuracy.

D. SENSOR DISTRIBUTION

Crop pests thrive in excessively dry, hot environments. In this research, we combined environmental sensors through a wireless environmental sensor transmission platform and a cloud big data computing server system to collect environmental data. These data were analyzed to determine the real-time agricultural meteorology, soil conditions, crop growth, and other relevant information necessary to assist the management of an orchard environment.

The research and development of wireless sensor transmission systems focus on multiple sensors to transmit services through a wireless network in real-time and send the data back to the system for analysis. It is one of the essential development applications of the IoT in smart agricultural technology.

In this study, the sensor modules were evenly arranged in an orchard on a hillside according to the wireless transmission sensor characteristics. The environmental data collected by the sensor modules were aggregated and transmitted to the cloud database through a wireless communication protocol for subsequent analysis and calculation.

Fig. 12. shows a schematic diagram of the transmission mechanism of the wireless sensing platform in the orchard on a slope.

We obtained 6 evenly distributed sets of environmental sensing modules in the orchard, as shown in Fig. 13, to observe the proliferation of pests in different environments.

Fig. 14 shows the designed environmental sensing module. The Arduino Nano was used in a small embedded platform. The four sensors, GY-30, soil, DHT22, and BMP180, were

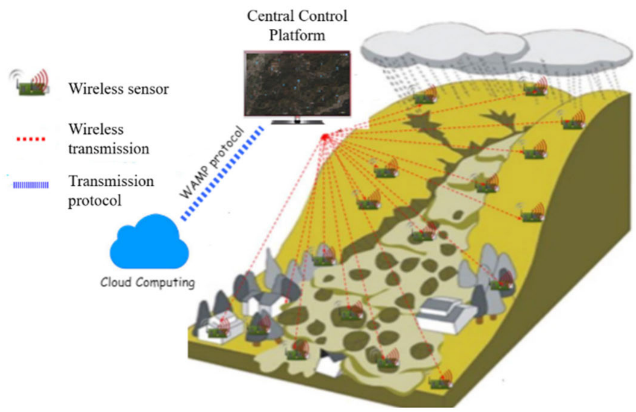

FIGURE 12. The schematic diagram of the transmission mechanism of the wireless sensing platform.

FIGURE 13. The actual configuration of the sensors in the field.

used to measure six kinds of environmental data, including temperature, humidity, light, soil humidity, atmospheric pressure, and altitude.

We use the DS3231 module to calibrate the time on the microprocessor. The Raspberry Pi receives the environmental data from the sensors every hour. It transfers the data to the Cloud via a wireless network and stores the data in the SD module on the client-side at the same time to avoid data loss through failed wireless transmission.

The environmental data are transmitted to the Raspberry Pi on the server station from the client-side through the SX1276 Lora module via a wireless network. The Raspberry Pi uploads the data to the cloud database to provide developers with real-time observations of environmental changes, as well as data retrieval and analysis.

To collect the environmental data in the field any time we wanted, we set up six sets of sensors and a LORA transmission module on the client-side. The LORA module is used to transmit the data to the Raspberry Pi on the server station, which transfers the data to the cloud database through a wireless network to provide query and analysis functions to the managers in real-time. The architecture is shown in Fig. 15.

The six sets of client sensors send data to the server station at different times. To ensure the correctness of the data received by the server station, the server station returns an

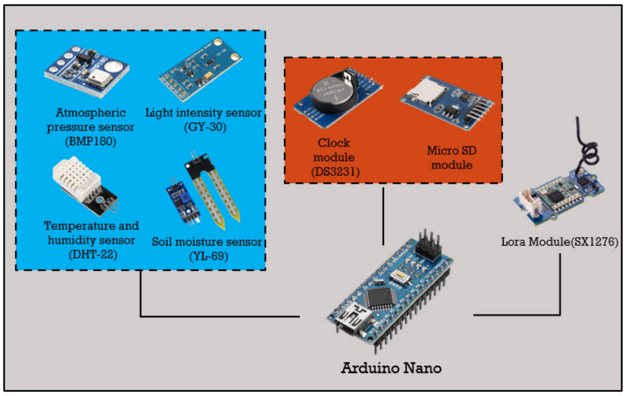

FIGURE 14. The hardware architecture of the sensor box content.

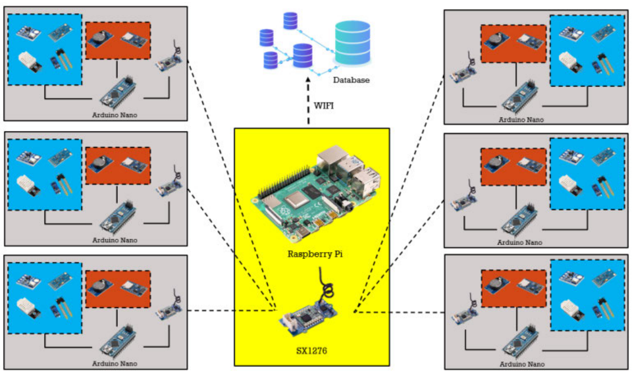

FIGURE 15. Transmission architecture of client-side sensor and server station sensor.

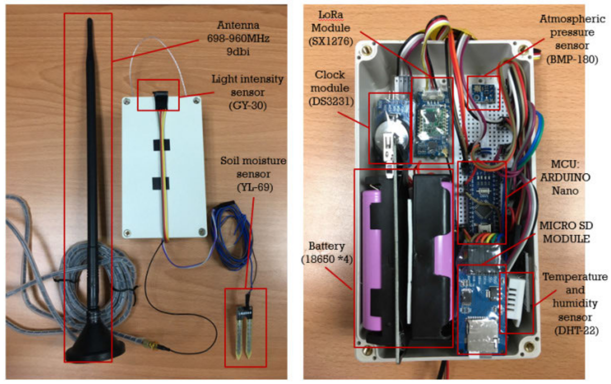

FIGURE 16. (a) The external module of the environmental sensors (b) The internal module of the environmental sensors.

acknowledgment value to the client-side after receiving the data. Otherwise, the client re-sends the data.

The environmental sensing module in this experiment included the Lora transmission module, sensors for temperature, humidity, soil humidity and atmospheric pressure, a clock module, an SD card module, a micro-controller Arduino Nano, and a lithium battery. All components are shown in Fig.16.

E. ANALYZING ENVIRONMENTAL SENSOR DATA WITH LSTM

We collected a total of 6 sets of environmental sensor information on the slope area in a longan orchard in Nanhua District in Tainan, Taiwan, from January to June 2020. Data for each environmental sensor module was collected at one hour intervals, where each set of data included real-time environmental information for the time, temperature, humidity, light, soil humidity, atmospheric pressure, and relative altitude. Up to the present time, we have collected a total of 12,960 sets of environmental data. The environmental data had slight variations due to the location of the sensor module. Hence, we used LSTM to analyze the six sensor modules to predict environmental factors at six locations and used machine learning methods to predict whether environmental factors will result in the presence of pests.

The proposed system integrated an agricultural management mobile application to provide pest image collection, real-time pest identification, and visualization of sensor data. The developed mobile application provides real-time access to environmental data and displays the historical records of the environmental sensors. In this work, the analytical results from the environmental data are displayed in the form of an App on the mobile device to provide farmers with real-time agricultural management. The mobile application is divided into two parts and is described follows:

1) BROWSING OF THE ENVIRONMENTAL SENSORS DATABASE

The environmental sensors deployed in the orchard allow users to monitor the current growing environment of the fruits trees. The historical environmental data also provide information about the environmental conditions.

2) UAV MISSION DISPATCH MODE

Users can use drones to plan precise pesticide spraying operations in real-time, either using manually-controlled flights or pre-input flight paths obtained through the mobile application.

In this work, multiple sensors are placed in an orchard to collect environmental data and provide instantaneous information on pest damage in various areas of the orchard, so that the correlation between pest occurrence and related environmental factors can be determined. We used the LSTM in the RNN neural network to build a suitable training model based on an on-site investigation of the pest severity in orchards by professional farm managers and the environmental data. The model is able to analyze the predicted data and determine the extent of pest occurrences. The prevention and control information are then provided to the farmers as quickly as possible to reduce the agricultural damage caused by the identified pests.

IV. RESULTS

In the experiment, the data was processed, and image augmentation was performed to increase the number of training

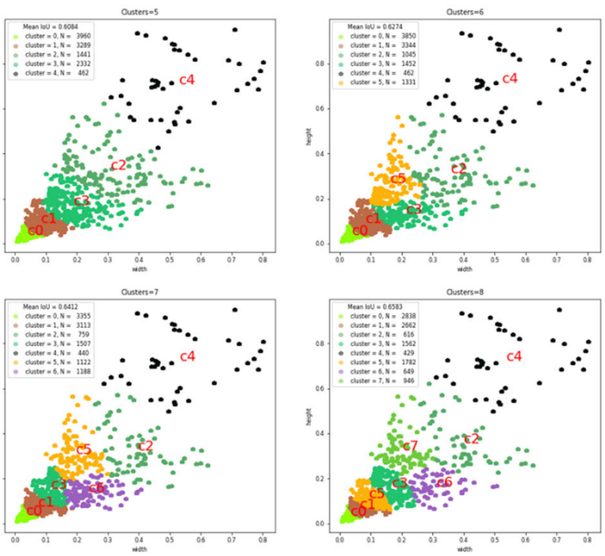

FIGURE 17. The experiment results of using K-means to find a suitable anchor box.

samples. The image augmentation methods used in this work included left and right reversal, angle conversion, Gaussian blur, and zoom in and out, among others. The number of training samples was increased from 849 to more than 7,000. The YOLOv3 model was used for deep learning training. The anchor box in the YOLOv3TensorFlow model was obtained by clustering the VOC data set. However, the 20 types in the VOC data set comprise large differences in the target object sizes, ranging from large objects such as a bicycle or a bus to smaller objects such as a bird or a cat. Therefore, the VOC data set was not suitable as a data set for target training and detection in this experiment. In the experiment, we found the anchor box with k = 9 to be more suitable for the target detection in this work, as shown in Fig. 17.

The Stochastic Gradient Descent is used in this work as the learning algorithm. The gradient descent algorithm uses each weight update and parameter deviation to advance each update to a negative gradient in order to obtain a minimum error function value. The formula is as follows:

$$\theta_{l+1} = \theta_l - a\nabla E(\theta_l) \tag{1}$$

where l is the number of iterations; a is the learning rate, which is set at 0.0001 in this experiment; θ is the parameter vector, and $E(\theta)$ is the loss function.

The gradient $\nabla E(\theta)$ of the loss function is calculated by taking a subset of the training data to measure and update the gradient parameters. The process for updating and adjusting the gradient of the entire training data set is called an epoch.

In object recognition, the score and IOU are significant figures. The score is given by the target frame when the model recognizes it as having the highest pest probability. The IOU is the intersection between the two rectangles representing the frame of the recognition result and the correct target. The amount of intersection between the rectangular frames

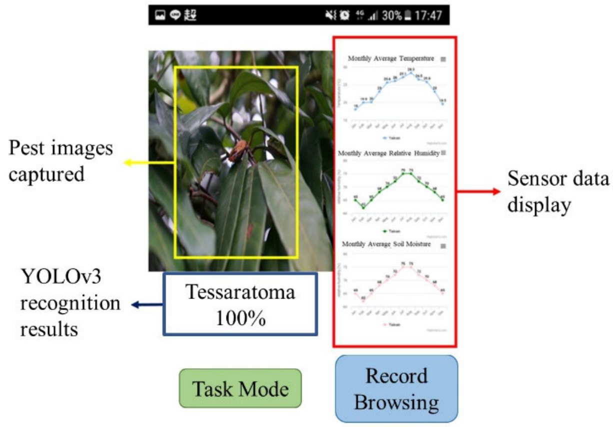

FIGURE 18. The schematic diagram of the user interface.

FIGURE 19. The spatial distribution of identified pests and the impact map.

TABLE 5. The correlation adjustment between the score/IOU parameters and mAP.

Score	IOU	mAP
0.15	0.3	92.7%
0.3	0.45	92.52%
0.4	0.5	90.25%

is assigned a value from 0 to 1. The experimental results are shown in Table. 4.

During the training process, the YOLOv3 model continuously uses different parameters in an attempt to find the most suitable model for the given dataset. The test data is given as input into the model and combined with the sensor data from the orchard to analyze the level of damage and the growth stages of the pest. It has been found that the mAP accuracy is 0.92.

To make the system more accessible and convenient to farmers, we designed a mobile application with a photography function. The mobile application is also capable of updating the environmental data at any time. The mobile application is shown in Fig. 18 and Fig. 19. The mobile application makes it possible for farmers to identify pests while patrolling the orchard. The identification results are sent to the cloud to analyze the location of the pest, so the farmers can be informed as to the best time for pesticide spraying with the goal of improving pest prevention and management.

V. CONCLUSION

It is difficult to obtain image data of pests since they prefer hiding behind leaves and in the tops of trees. Therefore, we applied image data augmentation to increase the number of pest training samples and improve the recognition accuracy for *Tessaratoma papillosa*.

In this study, the YOLO series (You Only Look Once, YOLO) neural network algorithm was used to identify the pests in the input images. The recognition rate was about 90%. The main target was *Tessaratoma papillosa*. Drones were used to collect images of these pests on fruit trees in an orchard on a slope. A mobile application was used to collect pest images from the ground at the same time. The pest images were uploaded to the Cloud for image recognition in real-time, and the pest location and distribution were determined for the purpose of pesticide application. The proposed system is designed to facilitate farmers in advanced pest control operations.

In this study, a deep-learning image recognition method for *Tessaratoma papillosa* was proposed that uses the YOLOv3 model to mark and classify the pests in images and extract the characteristics of the pests. The pest identification accuracy and model training time are used to determine the most suitable model for the application. Also, the data from the environmental sensors was used to analyze locations prone to the presence of pests. The location and distribution of the pests were instantly provided to farmers to provide accurate pesticide application, reduce agricultural pest damage, and increase crop quality and yield.

Since *Tessaratoma papillosa* mostly grows on the backs of leaves and in treetops, it is difficult to collect training samples for the purpose of identifying these pests. Further, the inconsistency of outdoor lighting results in low image recognition accuracy. Therefore, here, an APP installed on a mobile phone was used to collect images of *Tessaratoma papillosa* that appeared on the back of the leaves in order to locate these pests. We plan to apply the *Tessaratoma papillosa* identification model to small drones in the future, hoping to reduce the human effort for collecting pest image samples by the maneuverability of drones. It is hope that this will increase the number of images of *Tessaratoma papillosa* that appear in treetops and improve identification accuracy.

Some difficulties were encountered in the research institute: The first was the need to continuously increase the different angles in the images to solve the issue of insufficient training samples; the second was to stabilize the drone flying between the trees to get good quality images of pests and resolve the disturbances of the leaves caused by the wind from the UAV propeller; the third was speeding up the image recognition process to improve pest positioning efficiency and reduce the time required for farmers to inspect the area for pests. We plan to overcome the problems mentioned above and continue to experiment to find solutions, hoping to achieve immediate recognition of pest by drones as soon as possible and establish intelligent agriculture.

REFERENCES

[1] L. M. Gladence *et al.*, "Recommender system for home automation using IoT and artificial intelligence," *J. Ambient Intell. Hum. Comput.*, 2020, doi: 10.1007/s12652-020-01968-2.

[2] S. S. Priya, L. M. Gladance, A. Valarmathi, and J. M. S. Joy, "Brain tumor detection and disease prediction using CNN algorithm," in *Test Engineering and Management*, vol. 82. Oakland, CA, USA: The Mattingley Publishing Co., Inc., Jan./Feb. 2020, pp. 13376–13381.

[3] G. L. Mary, C. K. Vakula, M. P. Selvan, and T. Y. S. Sam-Hita, "A research on application of human-robot interaction using artificial intelligence," *Int. J. Innov. Technol. Exploring Eng.*, vol. 8, no. 9, pp. 784–787, 2019.

[4] V. Martin and S. Moisan, "Early pest detection in Greenhouses," in *Proc. Int. Conf. Pattern Recognit.*, 2008, pp. 1–4.

[5] K. Wang, S. Zhang, Z. Wang, Z. Liu, and F. Yang, "Mobile smart device-based vegetable disease and insect pest recognition method," *Intell. Autom. Soft Comput.*, vol. 19, no. 3, pp. 263–273, Aug. 2013.

[6] J. L. Miranda, B. D. Gerardo, and B. T. Tanguilig, III, "Pest detection and extraction using image processing techniques," *Int. J. Comput. Commun. Eng.*, vol. 3, no. 3, pp. 189–192, 2014.

[7] M. D. Gondal and Y. N. Khan, "Early pest detection from crop using image processing and computational intelligence," *FAST-NU Res. J.*, vol. 1, no. 1, pp. 59–68, Jan. 2015.

[8] W. Ding and G. Taylor, "Automatic moth detection from trap images for pest management," *Comput. Electron. Agricult.*, vol. 123, pp. 17–28, Apr. 2016.

[9] A. Almahairi, S. Rajeswar, A. Sordoni, P. Bachman, and A. Courville, "Augmented CycleGAN: Learning Many-to-Many mappings from unpaired data," 2018, *arXiv:1802.10151*. [Online]. Available: http://arxiv.org/abs/1802.10151

[10] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[11] F. Faith praise, P. Birch, R. Young, J. Obu, B. Faith praise, and C. Chatwin, "Automatic plant pest detection and recognition using k-means clustering algorithm and correspondence filters," *Int. J. Adv. Biotechnol. Res.*, vol. 4, no. 2, pp. 189–199, 2013.

[12] Z. Liu, J. Gao, G. Yang, H. Zhang, and Y. He, "Localization and classification of paddy field pests using a saliency map and deep convolutional neural network," *Sci. Rep.*, vol. 6, no. 1, Apr. 2016.

[13] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[16] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[17] D.-R. Liu, S.-P. Chuang, and H.-Y. Lee, "Attention-based memory selection recurrent network for language modeling," 2016, *arXiv:1611.08656*. [Online]. Available: http://arxiv.org/abs/1611.08656

CHING-JU CHEN received the Ph.D. degree in engineering science from National Cheng Kung University, Taiwan, in 2011. She is currently an Assistant Professor with the Department of Bachelor Program in Interdisciplinary Studies, National Yunlin University of Science and Technology, Taiwan. Her research interests include artificial intelligence, UAV system integration, embedded systems, environmental resource applications, computational intelligence applications to UAV telemetry image processing, geographical information systems, global positioning systems, remote sensing, computer vision, and image recognition.

YA-YU HUANG was born in Changhua, Taiwan, in 1997. She received the B.S. degree in computer science and information engineering from the National University of Tainan, Tainan, in 2019. She is currently pursuing the master's degree with the Department of Engineering Science, National Cheng Kung University. From 2018 to 2019, she was a Research Assistant with the Knowledge, Information and Database System Laboratory. She is also working on machine learning applied to computer vision and image processing in robotics applications. Her research interests include remote sensing applications in agriculture, applications of machine learning in satellite imagery, optical satellite and aerial imagery, and image processing.

YUAN-SHUO LI was born in Tainan, Taiwan, in 1996. He received the B.S. degree in computer science from National Pingtung University, Pingtung, in 2018. He is currently pursuing the master's degree with the Department of Engineering Science, National Cheng Kung University, Tainan. His research interests include remote sensing imagery in multispectral, application of embedded systems, and LoRa wireless transmission technology.

CHUAN-YU CHANG (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from National Cheng Kung University, Taiwan, in 2000.

He was the Chair with the Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology (YunTech), Taiwan, from 2009 to 2011. From 2011 to 2019, he served as the Dean of Research and Development and the Director of the Incubation Center for Academia-Industry Collaboration and Intellectual Property, YunTech. He is currently the Deputy General Director of the Service Systems Technology Center, Industrial Technology Research Institute (ITRI), Taiwan. He is also a Distinguished Professor with the Department of Computer Science and Information Engineering, YunTech. His current research interests include computational intelligence and their applications to medical image processing, automated optical inspection, emotion recognition, and pattern recognition. In the above areas, he has more than 200 publications in journals and conference proceedings.

Dr. Chang is a Fellow of IET and a Life Member of IPPR and TAAI. He served as the Program Co-Chair of TAAI 2007, CVGIP 2009, 2010–2019 International Workshop on Intelligent Sensors and Smart Environments, and the Third International Conference on Robot, Vision and Signal Processing (RVSP 2015). He served as the General Co-Chair of 2012 International Conference on Information Security and Intelligent Control, 2011–2013 Workshop on Digital Life Technologies, CVGIP 2017, WIC2018, ICS 2018, and WIC2019. From 2015 to 2017, he was the Chair of the IEEE Signal Processing Society Tainan Chapter and the Representative of the Region 10 of IEEE SPS Chapters Committee. He is also the President of the Taiwan Association for Web Intelligence Consortium.

YUEH-MIN HUANG (Senior Member, IEEE) received the M.S. and Ph.D. degrees in electrical engineering from The University of Arizona, in 1988 and 1991, respectively.

He is currently the Chair Professor with the Department of Engineering Science and the Institute of Education, National Cheng Kung University, Taiwan. He has trained over 60 Ph.D. and 300 M.S. thesis students. He has coauthored three books and has published more than 280 refereed journal research articles. His research interests include e-Learning, multimedia communications, and artificial intelligence. He became a Fellow of the British Computer Society, in 2011. He received many research awards, such as the Taiwan's National Outstanding Research Award, in 2011 and 2014, and the 2017 Taiwan Outstanding IT Elite Award. He is in the editorial board of several international journals in the area of educational technology, computer communications, and web intelligence. He is also the Founding Chair of International Symposium of Emerging Technologies for Education (SETE) and International Conference of Innovative Technologies and Learning (ICITL).